Are Your Agents Too Predictable?
Enhancing Human Likeness through Action Uncertainty

Anonymous EMNLP submission

Abstract

Authentic agent behavior is a critical component of credible social simulations. Social behavioral theories suggest that human behavior is characterized by an unpredictable uncertainty, which is the critical driver behind the emergence of genuine social phenomena in the real world. However, existing simulation methods tend to neglect modeling such uncertainty, resulting in highly predictable agent behavior and thereby reducing their human-likeness. To address this limitation, we propose the Agent framework with Endogenous Uncertainty (EUA) to explicitly model the action uncertainty. Specifically, inspired by the *Extended Somatic Marker Theory* and *Dual process Theory*, EUA deconstructs action uncertainty along physiological and psychological dimensions. Physiologically, a Hormone module is proposed to serve as a latent variable for physiological sensations, thereby influencing the decision-making process. Psychologically, the Subconscious Discrimination module quantifies agent's latent disposition to generate behavioral diversity. Finally, these two sources of uncertainty are integrated with intrinsic desires to drive action generation. The experimental results indicate that daily activities produced by EUA more closely resemble actual human behavior.

1 Introduction

The development of autonomous, human-like agents is a long-standing objective in social simulations (Wang et al., 2016, 2024b; Fetzer, 1990). Numerous sociological theories (Latour, 1996; Blumer, 1986; Martin, 2003) suggest that human behavior is characterized by an unpredictable uncertainty, which is the foundation to reliably model and predict realistic social phenomena. For instance, sentencing decisions made by the same judge for similar cases can differ significantly depending on the time of day (Kahneman et al., 2021; Tversky and Kahneman, 1974). Therefore, modeling such action uncertainty (Knight, 1921) is paramount for creating agents that exhibit realistic human behavior.

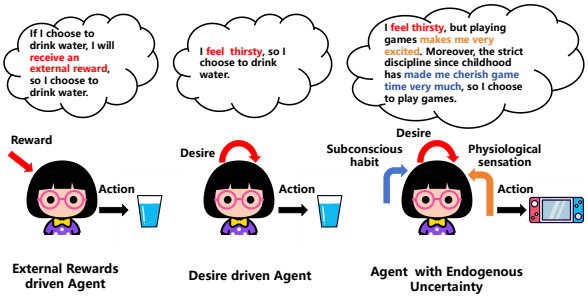

Figure 1: Agent Behavior Generation Mechanism. Our framework generates more human-like behaviors by fusing physiological and psychological uncertainty with the agent's current desires (right). In contrast, the highly deterministic nature of other methods produces behaviors that are highly predictable, and consequently fail to align with authentic human behavioral patterns.

However, most existing approaches employ external rewards to facilitate human-like behavior (Shen et al., 2023; Zhang et al., 2025; Song et al., 2023; Sivamayil et al., 2023). These reward mechanisms are highly task-specific, which leads to stereotyped behavioral patterns. To address this rigidity, subsequent methods (Wang et al., 2025b; Piao et al., 2025) have proposed desire-driven autonomous frameworks for adaptive activity generation. Nevertheless, the behaviors generated by these agents are purely rational actions. This mechanism inherently orients them toward an optimal solution, resulting in highly predictable behavioral trajectories. Ultimately, as shown in Fig 1, the determinism inherent in prevailing agent paradigms presents a fundamental barrier to achieving genuine alignment with human conduct. The full Related Works are in Appendix C.

To address this limitation, our work draws on *Psychoanalytic Theory* (Freud, 2012, 2016), which argues that a substantial portion of human action

is shaped by non-conscious physiological and psychological processes. Building on this premise, we posit that human action is, in large part, inherently uncertain. Guided by this view, we explicitly model action uncertainty in intelligent agents. To make this modeling principled and structured, we further rely on two seminal frameworks: the *Extended Somatic Marker Theory* (Damasio, 1996), which characterizes how physiological sensations covertly bias decision-making, and *Dual Process Theory* (Evans, 2008), which formalizes the notion that subconscious often precede and can bias subsequent deliberative reasoning. Accordingly, we decompose action uncertainty into its physiological and psychological dimensions.

Firstly, we incorporate **Physiological Sensation Fluctuations** to model physiological uncertainty. Inspired by *Hormonal Determinism* (Sapolsky, 2017, 2005, 2023), a Hormone module is introduced which uses names of neurotransmitters as interpretable channels to parameterize physiological-like sensations (*e.g.*, dopamine, endorphins, oxytocin, and serotonin) (Schultz, 2006; Hornykiewicz, 1966; Ferrari et al., 2005). These sensations modulate an agent's behavioral choices, and subsequent actions trigger updates to corresponding hormone levels, which creates a feedback loop: hormone levels → physiological sensation → behavioral choice → updated hormone levels. Furthermore, we introduce a **Subconscious Discrimination module** to model psychological uncertainty. This module classifies agents based on their backstories into distinct behavioral archetypes: for example, cognitively rigid agents with a singular perspective versus those capable of critical thinking. The underlying subconscious disposition then shapes the agent's behavioral choices, creating a second feedback loop: subconscious → behavioral choice → subconscious update. Finally, these distinct sources of uncertainty are synthesized with the agent's intrinsic desires to drive action generation, yielding behaviors that more faithfully approximate authentic human conduct.

Our main contributions can be summarized as follows: (1) We are the first to systematically model the inherent action uncertainty in human-like agents, which establishes a new perspective for agent design and simulation. (2) we propose a novel EUA framework which generates more human-like agent behavior by constructing two dimensions of action uncertainty: physiological uncertainty driven by hormonal fluctuations, and psychological uncertainty driven by subconscious differences. (3) Comprehensive experiments demonstrate the effectiveness and robustness of our proposed framework.

2 Methodology

Figure 2 illustrates the overall flowchart of the EUA method. First, we introduce an Action Uncertainty module, which is composed of two components: hormone-induced physiological sensation fluctuations and profile-driven subconscious differences. Next, the Behavioral Choice module synthesizes the outputs from the Desire module and the Action Uncertainty module, driving the agent to generate an action that best aligns with its motivation under the given perturbations. This action, in turn, completes the feedback loop by triggering updates to the agent's hormonal, subconscious, and desire states, thereby influencing all subsequent decisions.

2.1 Action Uncertainty

Inspired by the *Extended Somatic Marker Theory* (Damasio, 1996), which posits that physiological sensations can covertly guide decision-making, and *Dual Process Theory* (Evans, 2008), which demonstrates that subconscious processes precede and influence an individual's conscious decision-making, we deconstruct the uncertainty of human behavior into two layers: a physiological one, manifested as the physiological sensation fluctuation, and a psychological one, represented by subconscious difference. Subsequently, we introduce a Hormone module and a Subconscious Discrimination module to model these sources of uncertainty, respectively.

2.1.1 Physiological Sensation Fluctuation

Inspired by the *Hormonal Determinism* (Sapolsky, 2017, 2005, 2023), EUA models the physiological source of action uncertainty by treating hormone signals as latent variables. This mechanism furnishes the agent with a quantifiable physiological sensation, which in turn aligns its behavior generation pattern closely with human decision-making processes. We emphasize that these signals are computational abstractions inspired by, but not equivalent to, biological hormone concentrations.

Specifically, we introduce a Hormone module that uses four neurotransmitters (dopamine, serotonin, endorphin, and oxytocin) as interpretable channels to parameterize physiological-like sensations through the use of dynamic values and thresholds (Schultz et al., 1997; Insel and Young, 2001;

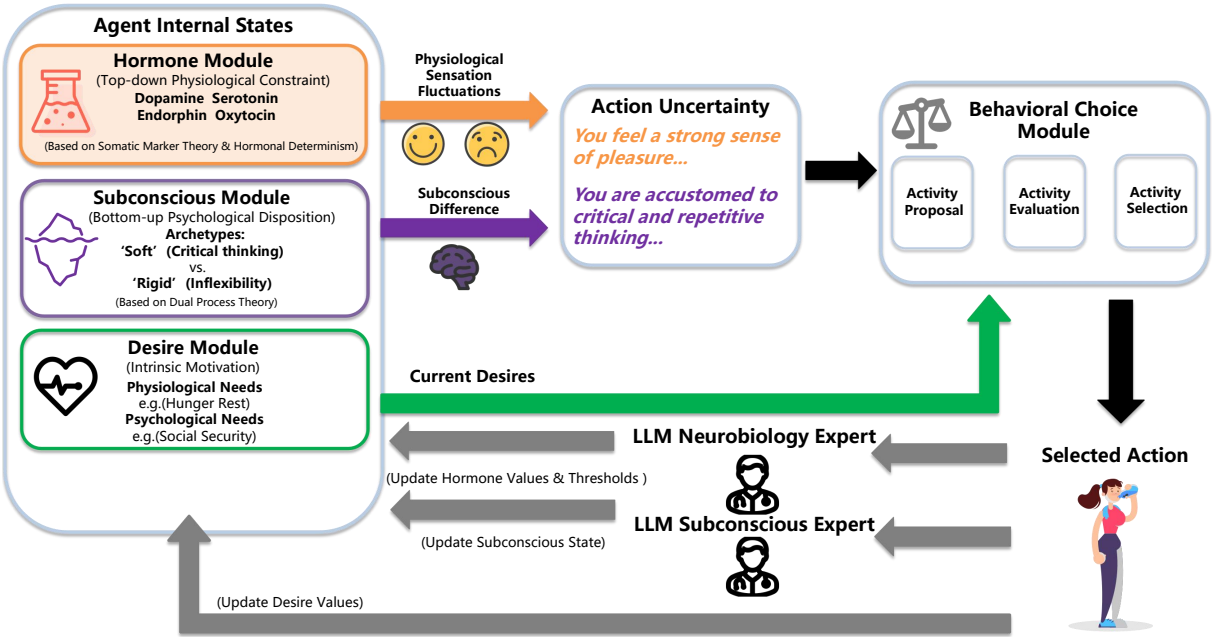

Figure 2: The Agent framework with Endogenous Uncertainty (EUA). The framework consists of three core modules: the **Hormone Module** (modeling physiological sensations via neurotransmitters), the **Subconscious Module** (defining psychological dispositions and archetypes), and the **Desire Module** (representing intrinsic motivations). The Hormone and Subconscious modules generate **Action Uncertainty** through physiological sensation fluctuations and subconscious differences, respectively. These factors, along with current desires, inform the **Behavioral Choice Module**, which proceeds through Activity Proposal, Evaluation, and Selection. Finally, **LLM Experts** facilitate feedback loops to update hormone values, subconscious states, and desire values based on the **Selected Action**.

Canli and Lesch, 2007). Following each action, the model consults an LLM-based "neurobiology expert" to infer the resultant change for each hormone. Notably, this step serves as a text-to-signal mapping rather than a biological simulator. Existing literature (Xu et al., 2025) indicates that current large models possess sufficient biological knowledge to support this kind of determination with high accuracy (Experiments in Appendix B.4 also validates the conclusion).

Drawing inspiration from *Sensory Adaptation Theory* (Thompson and Spencer, 1966)—which posits that persistent stimulation diminishes receptor sensitivity—we formulate the Hormone module's update rules as a computational abstraction. We emphasize that these hormone values are internal signals governed by heuristic dynamics, designed to simulate perceptual adaptation rather than to represent measurable biological concentrations. Initially, the value v_h and threshold v_{thre} for each hormone dimension h are set to $v_h^0 \in [1, 10]$ and $v_{\text{thre}}^0 = \max(1, v_h^0 - 1)$, respectively. Their dynamics are then governed by the following equations:

$$v_h^t = \begin{cases} v_h^{t-1} + k & \text{if} \quad \Pi_h\left(a_{t-1}\right) = 1 \\ v_h^{t-1} - p_h \times k & \text{if} \quad \Pi_h\left(a_{t-1}\right) = 0, \end{cases} \tag{1}$$

$$v_{\text{thre}}^t = \begin{cases} v_{\text{thre}}^{t-1} \times e^k & \text{if} \quad \Pi_h\left(a_{t-1}\right) = 1 \\ & \& \quad v_h^{t-1} > v_{\text{thre}}^{t-1} \\ v_{\text{thre}}^{t-1} \div e^k & \text{if} \quad \Pi_h\left(a_{t-1}\right) = \Pi_h\left(a_{t-2}\right) = 0 \\ & \& \quad v_{\text{thre}}^{t-1} > v_h^{t-1}, \end{cases} \tag{2}$$

where k is a configurable constant; p_h is a random number between 0 and 1, representing the decay probability of the hormone value at dimension h. $\Pi_h(\cdot)$ denotes the discriminant function of the LLM neurobiological expert to determine whether the hormone h is secreted (Appendix D.1). When activity a triggers the secretion of the hormone h, $\Pi_h(a) = 1$; otherwise, $\Pi_h(a) = 0$. Moreover, the set of physiological sensations of the hormone h for step t, denoted as F_h^t, is obtained by using the following equation (Detailed sensations are in Appendix E):

$$F_h^t = \begin{cases} \text{Positive Sensations} & \text{if} \quad v_h^t > v_{\text{thre}}^t \\ \text{Negative Sensations} & \text{if} \quad v_h^t < v_{\text{thre}}^t. \end{cases} \tag{3}$$

Finally, the sensation sets from each hormonal dimension are concatenated to yield the total sensation set, F^t. Overall, physiological sensations are instantiated by assessing current hormonal values against their respective thresholds. Surpassing

these thresholds triggers positive sensations; for example, elevated dopamine levels elicit profound pleasure, whereas oxytocin fosters a sense of tranquility and interpersonal trust. Conversely, a hormone value that remains subthreshold triggers a negative sensation, such as irritability or fatigue.

2.1.2 Subconscious Difference

To model the psychological source of behavioral fluctuations, we introduce a Subconscious Discrimination module based on *Dual Process Theory* (Evans, 2008), which demonstrates that subconscious processes precede and shape an individual's conscious decision-making. The central premise of this module is that an agent's subconscious archetype dictates the strategies it employs to satisfy its needs.

Grounded in the *Social Unconscious Theory* (Fromm, 2023), which posits that critical consciousness is either suppressed or expressed, we assign each agent an initial subconscious prototype, S^0, based on its background story (Appendix D.7). This prototype is one of two types, reflecting the theory's dichotomy: The "Soft" Prototype: Characterized by critical thinking, intellectual flexibility, and the capacity to perceive underlying social truths. The "Rigid" Prototype: Characterized by cognitive inflexibility, strong conformity, and adherence to surface-level norms and interpretations. During each simulation step t, the module shapes an agent's strategy for addressing needs by tailoring its decision-making prompts according to its subconscious archetype, S^t. The outcomes of the resulting action a_t then create a feedback loop, providing information for the LLMs to update the agent's subconscious state:

$$S^{t+1} = \Phi_{\text{LLM}}(S^t, a_t), \tag{4}$$

where $\Phi_{\text{LLM}}(\cdot)$ represents the subconscious update function (Appendix D.6).

2.2 Desire

The Desire module draws from D2A (Wang et al., 2025b), where agents initialize each desire dimension d with an initial value v_d^0 and a target value v_d^*. At each timestep t, the previous value v_d^{t-1} first decays by a fixed amount with probability p_d. Then, following the agent's action a_t, the value is updated to v_d^t. Finally, this set of numerical desire values is translated into natural language sentences, D^t, for the large language model to process.

2.3 Behavioral Choice

At each simulation step t, the Behavioral Choice module models physiological sensations, applies subconscious differences, defines desires, and then proposes, evaluates, and selects an activity. First, the Behavioral Choice module constructs action uncertainty by integrating physiological and psychological dimensions:

$$U^t = \text{Concatenate}(F^t, S^t), \tag{5}$$

where F^t is the set of physiological sensations, S^t is the subconscious state, and U^t is the combined action uncertainty representation.

Subsequently, the module initiates the activity proposal phase, which, inspired by the Tree of Thoughts methodology (Yao et al., 2023), prompts the agent to generate a set of N candidate actions. This generative process is conditioned on a comprehensive set of inputs: the agent's profile p, environment e, historical actions and observations $a_{0:t-1}$, $o_{0:t-1}$, current desires D^t, and its immediate action uncertainty U^t. The objective is to propose activities that are not only contextually appropriate and historically informed but are also tailored to satisfy its desires as framed by its current physiological and psychological state (Appendix D.8):

$$\begin{aligned} a_t^1, a_t^2, \ldots, a_t^N =& \text{ActProposal} \\ & \big(U^t, D^t, a_{0:t-1}, o_{0:t-1}, p, e\big), \end{aligned} \tag{6}$$

These N candidate activities are subsequently passed to the activity evaluation process (Refer to D2A, the value of N is 3). Under the perturbation of U^t, the Behavioral Choice module evaluates each activity by envisioning the expected desire state $\hat{D}_t^i$ that the agent would experience after performing action a_t^i (Appendix D.9):

$$\hat{D}_t^i = \text{ActEvaluation}\big(U^t, D^t, a_t^i\big), i \in [N]. \tag{7}$$

Finally, within the activity selection process, the Behavioral Choice module compares the predicted desire states $\hat{D}_t^i$ of all N proposed actions under the perturbation of the agent's action uncertainty U^t. Based on these evaluation results, the module selects the activity that best aligns with the agent's current motivation and behavioral fluctuations (Appendix D.10):

$$a_t = \text{ActSelection}\big(U^t, \hat{D}_t^i, a_t^i, i \in [N], p, e\big). \tag{8}$$

After completing each simulation step, the agent updates the hormone state, subconscious state and desire state based on the performed activity.

3 Experiment Settings

3.1 Environment

Referring to D2A (Wang et al., 2025b), our primary experimental environment consists of two scenarios: an indoor scene and an outdoor party scene constructed by the Concordia library (Vezhnevets et al., 2023). These two scenarios were deliberately selected to address potential concerns regarding environmental limitations through two complementary rationales. (Details are in Appendix A.1)

3.2 Baselines

We compare EUA with four baseline models: ReAct (Yao et al., 2022), where the agent performs reasoning before executing actions; LLMob (Wang et al., 2024a), which generates activity plans by extracting motivational features from user profiles and recommends the most probable activities accordingly; and BabyAGI (N.., 2023), which selects subsequent activities by maintaining a priority task list and following priority order; and D2A (Wang et al., 2025b), which generates behaviors based on desire-driven mechanisms. Notably, to ensure a fair comparison, the input configuration provided to EUA was kept strictly identical to that of the baselines (including the character's background settings, initial desires, initial hormone levels, and hormone thresholds). Furthermore, the action uncertainty U^t subsequently utilized by EUA is generated intrinsically through the model's own dynamic mechanisms, without incorporating any extraneous reasoning traces. All five methods adopt GPT-4o-mini as the backbone model.

3.3 Evaluation Metrics

We evaluate the proposed agent against baseline models using two metrics adapted from D2A (Wang et al., 2025b): *Likelihood* and *Dissatisfaction*.

Metric Definitions. *Likelihood* quantifies the human-likeness of generated action sequences across indoor and outdoor settings, measuring the extent to which agent decisions reflect natural, goal-directed human behavior. *Dissatisfaction* measures the discrepancy between the agent's current desire values and the expected desire values. Detailed definitions are provided in Appendix A.2

Evaluation Methodology. The Dissatisfaction metric is calculated by Eq. 10 in Appendix A.2 while the the Likelihood metric is assessed using a unified dual-track protocol: (i) *Automated Assessment:* Sequences are evaluated by large language models. To mitigate potential model-specific bias, we employ two advanced LLMs (GPT-4o and GPT-5) to score action sequences generated by the target model (GPT-4o-mini). Detailed prompts are provided in Appendix D.3. (ii) *Blinded Human Validation:* Three independent annotators evaluate the sequences without knowledge of their source. Annotators assign scores on a 0–5 scale across three criteria: **Logical Coherence** (goal-directed connectivity), **Environmental Alignment** (compatibility with capabilities and context), and **Behavioral Plausibility** (consistency with internal states). Final scores are computed by averaging across all criteria and raters. The complete human evaluation guidelines are detailed in Appendix A.3.

Table 1: Likelihood scores of indoor room and outdoor party experiments. EUA consistently outperforms all baselines. This superiority likely stems from EUA's explicit modeling of action uncertainty.

Environment	Method	GPT-4o	GPT-5	Human
Indoor	ReAct	4.42	4.11	3.30
	LLMob	4.51	4.10	3.30
	Baby_AGI	4.10	3.99	3.10
	D2A	4.73	4.26	3.50
	EUA_Soft	4.90	4.38	4.10
	EUA_Rigid	**4.95**	**4.48**	**4.30**
Outdoor	ReAct	4.45	4.35	3.30
	LLMob	4.42	4.39	3.40
	Baby_AGI	4.17	3.85	2.90
	D2A	4.66	4.25	3.60
	EUA_Soft	4.73	4.50	4.20
	EUA_Rigid	**4.79**	**4.57**	**4.25**

4 Results and Analysis

4.1 Human-likeness Evaluation

For baseline methods, we conducted 15 trials per agent across both indoor and outdoor environments. For EUA, we leveraged GPT-4o to generate ten distinct story backgrounds, each embedding a different subconscious habit (5 "Soft" agents, 5 "Rigid" agents; detailed backgrounds are provided in Appendix F). Note that subconscious habits are intentionally decoupled from occupational or educational attributes. We executed 3 trials per background, yielding two evaluation cohorts: "Soft" EUA (15 trials) and "Rigid" EUA (15 trials).

Table 1 demonstrates that across both indoor and outdoor scenarios, EUA consistently outperforms all baselines in likelihood scoring. This superiority likely stems from EUA's explicit modeling of action uncertainty, which enables agents to generate activity sequences aligned with their physiologi-

cal sensations, subconscious habits, and evolving desires. To bolster evaluation credibility, we additionally incorporated human assessment. The consistency between the human review scores and the LLMs evaluation scores further demonstrates the superiority of EUA's framework. Furthermore, detailed results regarding dissatisfaction are provided in Appendix B.1, which demonstrate that EUA is capable of rapidly satisfying the agent's current needs.

4.1.1 Further Analyses on Action Uncertainty

To isolate the contribution of our endogenous uncertainty design from stochastic sampling artifacts, we conducted a systematic ablation study using a controlled assembly-line worker agent to minimize environmental confounders (see Table 2). We evaluated three baseline-ablation variants against our proposed frameworks: (1) a temperature-scaled variant ($D2A_{temp}$) to probe the effect of LLM sampling randomness, (2) a prompt-diversity variant ($D2A_{noise}$) to assess explicit stochastic prompting, and (3) a randomized-sensation variant (EUA_{noise}) to examine the necessity of structured physiological feedback. Each variant was designed to test whether unstructured randomness alone can replicate the human-likeness achieved by our structured uncertainty mechanisms. First, we investigated whether elevated sampling randomness in LLMs could foster more human-like uncertainty structures. We systematically evaluated four temperature settings within the D2A framework and report the optimal configuration as $D2A_{temp}$ (comprehensive experimental details are provided in Appendix B.5). The results indicate that excessively high temperature parameters significantly degrade the human-likeness of the generated behaviors, demonstrating that unregulated sampling randomness fails to emulate authentic human uncertainty. Second, to examine whether explicit prompt-induced randomness could improve behavioral plausibility, we integrated diversity-guiding prompts into the action proposal, evaluation, and selection phases of the D2A framework. The resulting $D2A_{noise}$ variant reveals that stochastically injected behavioral diversity does not enhance the human-likeness of the outputs. This finding suggests that surface-level stochastic prompting cannot substitute for intrinsic uncertainty modeling. Third, to validate the structural plausibility of EUA's endogenous uncertainty generation, we replaced the physiological sensation module F^t with magnitude-matched random prompts F^t_{rand}, while preserving all other experimental conditions. Specifically, we maintained the empirical positive-to-negative sensation ratio observed in the original EUA trials and instructed the LLM to stochastically sample sensory states at each step according to this distribution. The EUA_{noise} variant yielded consistently lower human-likeness scores than the original EUA, confirming that structured, dynamically updated physiological signals—rather than randomized sensory inputs—are essential for generating plausible uncertainty patterns. Collectively, these ablation studies provide robust empirical support for our core hypothesis: the observed performance gains stem from the structured, endogenous uncertainty mechanisms of the EUA framework, rather than from stochastic sampling artifacts or random noise. By leveraging biologically inspired hormonal updates and subconscious variations, our approach demonstrates superior theoretical grounding and functional plausibility over naive randomness-based alternatives.

Table 2: Analyses on Action Uncertainty. The results validate our hypothesis: the performance gains stem from the structured, endogenous uncertainty mechanisms of the EUA framework, rather than from stochastic sampling artifacts or random noise.

Metric	Method	GPT-4o	GPT-5
Likelihood	D2A	4.73	4.26
	$D2A_{temp}$	4.72	4.2
	$D2A_{noise}$	4.71	4.0
	EUA_{noise}	4.83	4.44
	EUA	**4.9**	**4.67**

4.2 Hormone Effectiveness Evaluation

4.2.1 Interview Experiment

To determine whether hormone-induced physiological fluctuations would affect agent behavior, we conducted a questionnaire-based interview experiment to assess the behavioral choices of agents in various hormonal states. We designed three distinct task scenarios inspired by Behavioral Decision Research (Kahneman et al., 2021). These scenarios were designed to verify whether the induced behavioral fluctuations align with the real-world cases from the book, which demonstrate the uncertainty of human decision-making in fields such as law, medicine, and insurance. Moreover, another three scenarios were introduced by LLMs to explore human choices in more common contexts, such as remark discernment, security management, and food service. During each interview, we provided the

Table 3: Tendency score of the Psychological Counselor agent to select a comprehensive treatment plan under varying physiological sensations.

Condition	GPT-4o	GPT-5
D2A	7	8
EUA (Positive Sensation)	9	9
EUA (Negative Sensation)	4	4

agent with the background story B, the specific task T, the description of current desires D^t and the description of current physiological sensations F^t. Finally, We recorded the agent's resulting activity selection, a_{task}:

$$a_{\text{task}} = \text{AgentInterview}\big(D^t, F^t, B, T\big), \quad (9)$$

where AgentInterview$\big(\cdot\big)$ represents the LLM interview function (Appendix D.2). It is worth noting that all the interview experiments will be repeated three times. Given that the other baseline models do not have a Subconscious module, we ablated the Subconscious Discrimination module in our EUAs for the interview experiments. This was done to more intuitively demonstrate the influence of the Hormone module on behavior.

Detailed task scenarios, corresponding agent background stories, and interview contents are provided in the Appendix B.2. To facilitate a direct comparison with EUAs, D2As were assigned the same background stories, tasks, and desire state descriptions, and they underwent the same interviews. We used both GPT-5 and GPT-4o to simulate each role and complete the interviews. As illustrated in Fig 3 and Fig 4, the decision-making behavior of EUA agents demonstrates distinct variations depending on their physiological states. Specifically, physiological sensations induced by hormone secretion elicit behaviors characterized by increased friendliness, patience, trust and understanding. Conversely, negative sensations arising from hormone deficiency result in coldness, alongside a tendency to question and condemn the actions of others. These behavioral tendencies are consistent with findings from actual behavioral studies, a detailed analysis is presented in Case Study.

Summary: The above results demonstrate that the physiological sensation mechanism, derived from hormone values and thresholds, significantly influences the decision-making process of agents, thereby successfully modeling the physiological uncertainty inherent in human behavior patterns.

4.2.2 Case Study

To contextualize hormone-driven behavior within real-world phenomena, Table 3 presents qualitative results from the treatment plan selection task. When the EUA psychological counselor experiences pleasant and calm sensations, it demonstrates a stronger aversion to rapid symptom-relief interventions. Conversely, under frustration or irritability, the inclination toward quick-fix treatments increases markedly. This aligns with empirical observations (Kahneman et al., 2021) indicating that prolonged physiological fatigue (e.g., after eight consecutive working hours) impairs attentional capacity and precipitates decision simplification. Additional case studies are documented in Appendix B.2.

4.3 Subconscious Effectiveness Evaluation

4.3.1 Behavioral Similarity Experiment

Drawing upon the theoretical (Sapolsky, 2017) premise that endogenous subconscious states serve as cognitive filters for physiological drives, we hypothesize that the Subconscious Discrimination module should introduce structured, interpretable action uncertainty rather than stochastic noise. To evaluate this hypothesis, we conducted a pairwise comparison experiment quantifying Behavioral Similarity across five agent configurations: (1) ReAct; (2) D2A; and three EUA variants: (3) Full EUA; (4) $\text{EUA}_{\text{fixed}}$ (physiological sensations held constant); and (5) $\text{EUA}_{\text{fixed}}$ without the Subconscious Discrimination module. In each configuration, both agents under comparison were assigned an identical occupation, whereas their respective profiles were configured to implicitly encode distinct subconscious habits.

To ensure evaluation robustness, two independent LLMs (GPT-4o and GPT-5) scored sequence similarity using the metric defined in Appendix A.2, based on three criteria: **Activity Content** (specific actions performed), **Spatial Pattern** (locations visited), and **Temporal Rhythm** (sequential ordering). Each criterion received a 0–5 score, and the final score was the average across all three.

As summarized in Table 4, the quantitative results corroborate our theoretical expectations. First, baselines lacking the Subconscious Discrimination module (ReAct, D2A, and $\text{EUA}_{\text{fixed}}$) exhibit systematically higher similarity scores. This trend indicates that, under fixed initial conditions, the stepwise integration of physiological drive mech-

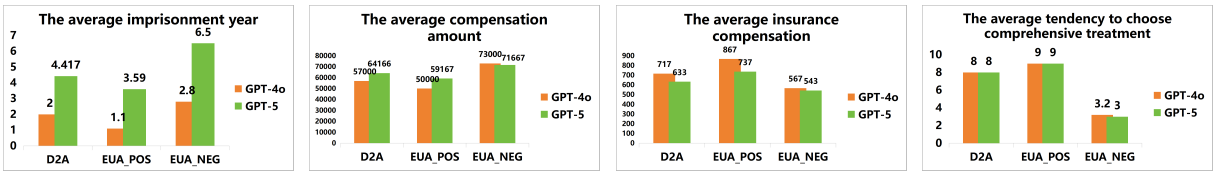

Figure 3: Interview results of three real-world scenarios: (a) imprisonment years (Task 1); (b) compensation amounts (Task 1); (c) insurance payouts (Task 2); (d) tendency to select comprehensive treatment (Task 3).

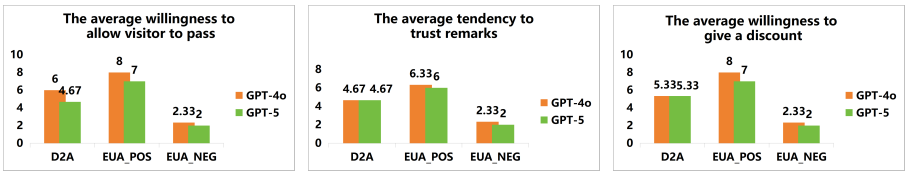

Figure 4: Interview results of three LLM-synthesized scenarios: (a) willingness to allow visitor passage (Task 4); (b) tendency to trust remarks (Task 5); (c) willingness to offer discounts (Task 6).

Table 4: Behavioral similarity of action sequences generated by agents with varying subconscious habits.

Method	GPT-4o	GPT-5
ReAct	3.40	3.50
D2A	4.43	4.30
$\text{EUA}_{\text{fixed}}$ w/o Subconscious	4.57	4.40
$\text{EUA}_{\text{fixed}}$	3.57	3.50
Full EUA	3.40	3.43

Table 5: Ablation study on hormone module and subconscious discrimination module.

Metric	Method	GPT-4o	GPT-5
Likelihood	EUA w/o Subconscious	4.81	4.6
	EUA w/o Hormone	4.85	4.58
	EUA	**4.9**	**4.67**

anisms progressively constrains the action space, yielding more predictable and convergent behavioral trajectories (see Appendix B.3 for detailed analysis). This aligns with the tenet of *Hormonal Determinism*: physiological constraints inherently restrict behavioral variability. Secondly, configurations retaining the Subconscious Discrimination module (Full EUA and $\text{EUA}_{\text{fixed}}$) exhibit a reduction in pairwise similarity again. Notably, even with physiological sensations held constant, variations in subconscious prototypes alone are sufficient to generate distinct action sequences.

Summary: These findings substantiate that subconscious modulation constitutes an independent source of structured uncertainty, effectively decoupling behavioral generation from bodily sensations. In addition, supplementary analyses (Appendix B.3) further demonstrate that occupational metadata exhibits no significant correlation with subconscious traits, effectively ruling out confounding behavioral priors. Collectively, these results establish that the Subconscious Discrimination module captures psychologically grounded uncertainty, thereby validating both the theoretical foundations of *Hormonal Determinism* and the architectural efficacy of the EUA framework.

4.4 Framework Analysis

To validate the EUA mechanism, we conduct ablation studies on the hormone and subconscious discrimination modules. As shown in Table 5, evaluations on indoor assembly-line worker agents using the *Likelihood* metric across GPT-4o and GPT-5 backbones demonstrate that the full EUA framework achieves optimal performance (4.90 and 4.67, respectively). Ablating the subconscious module yields performance drops of 0.09 and 0.07, while removing the hormone module reduces scores by 0.05 and 0.09. These results confirm that both modules independently enhance agent fidelity, with full integration producing the strongest alignment with human behavioral patterns. Furthermore, we perform a sensitivity analysis on the hormonal magnitude parameter k (detailed in Appendix B.5).

5 Conclusion

In this paper, we propose a novel framework for modeling the uncertainty of human behavior. The framework generates action uncertainty across physiological and psychological dimensions, and synthesizes the fluctuations and the agent desires to select a specific action, thereby aligning the agent's decision-making process with humans. Comprehensive experiments validate that the resulting activity sequences demonstrate significantly greater naturalness, coherence, and credibility, thereby achieving a higher degree of human-likeness.

Limitations

The hormone dimensions used in EUA are limited. To make the agent's behavior more closely align with that of real-world humans, more hormone dimensions should be introduced. In addition, EUA's hormone secretion mechanism involves querying a large language model role-playing as a neurobiologist. To improve the professional expertise of this question-answering process, more medical knowledge should be incorporated to assist the large language model.

The subconscious habits used in EUA are also limited. More subconscious habits need to be introduced into the framework through professional psychological knowledge, and at the same time, the renewal mechanism of the subconscious also needs further research.

In addition, EUA's physiological (hormone) component is a computationally inspired mechanism for generating action uncertainty. While we draw inspiration from *Somatic Marker* and *Hormonal Determinism* literature, we do not claim that the model is biologically faithful or that its channel values correspond to real neurotransmitter concentrations. In particular, the neurobiology "expert" step provides a heuristic, text-based polarity mapping from actions to physiological-like sensations; therefore, the module should be interpreted as a design metaphor and uncertainty generator rather than a scientific model of neurobiology.

Ethics and Social Impact

We acknowledge that correlating cognitive dispositions with occupational categories risks reinforcing harmful social stereotypes. To address this concern, we clarify that subconscious habits are formed independently of occupational backgrounds. Any profession can exhibit either high cognitive flexibility (classified as "Soft") or high need for closure (classified as "Rigid"), reflecting individual characteristics rather than occupational attributes. Furthermore, these archetypes are grounded in established psychological constructs, not demographic or occupational proxies. Finally, supplemental experiments (Appendix B.3) confirm that behavioral patterns are driven by subconscious habits, not occupational information.

References

P. A Adler, P. Adler, and A. Fontana. 1987. Everyday life sociology. *Annu. Rev. Sociol.*, 13(1):217–235.

T. Alshammari, N. Alshammari, M. Sedky, and C. Howard. 2018. Simadl: Simulated activities of daily living dataset. *Data*, 3(2):11.

H. Blumer. 1986. *Symbolic interactionism: Perspective and method*. Univ of California Press.

T. Canli and K. Lesch. 2007. Long story short: the serotonin transporter in emotion regulation and social cognition. *Nature neuroscience*, 10(9):1103–1109.

Harrison Chase. Langchain. url https://github.com/langchain-ai/langchain.

C. Colas, L. Teodorescu, P. Oudeyer, X. Yuan, and M. Côté. 2023. Augmenting autotelic agents with large language models. In *Proceedings of the CoLLAs*, pages 205–226.

A. R Damasio. 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1346):1413–1420.

J. St BT Evans. 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, 59(1):255–278.

J. Feng, Z. Yang, F. Xu, H. Yu, M. Wang, and Y. Li. 2020. Learning to simulate human mobility. In *Proceedings of the SIGKDD*, pages 3426–3433.

P. F. Ferrari, P. Palanza, S. Parmigiani, R. M. de Almeida, and K. A Miczek. 2005. Serotonin and aggressive behavior in rodents and nonhuman primates: predispositions and plasticity. *Eur. J. Pharmacol.*, 526(1-3):259–273.

J. H Fetzer. 1990. What is artificial intelligence? In *Artificial intelligence: Its scope and limits*, pages 3–27. Springer.

S. Freud. 2012. *The basic writings of Sigmund Freud*. Modern library.

S. Freud. 2016. Introductory lectures on psychoanalysis. In *Myths and mythologies*, pages 158–166. Routledge.

E. Fromm. 2023. *The heart of man: Its genius for good and evil*. Open Road Media.

T. Gärling and J. Garvill. 1993. Psychological explanations of participation in everyday activities. In *Adv. Psychol.*, volume 96, pages 270–297. Elsevier.

O. Hornykiewicz. 1966. Dopamine (3-hydroxytyramine) and brain function. *Pharmacol. Rev.*, 18(2):925–964.

A. Hurst, A. Lerer, A. P Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

T. R Insel and L. J Young. 2001. The neurobiology of attachment. *Nature Reviews Neuroscience*, 2(2):129–136.

D. Kahneman, Ol. Sibony, and C. R Sunstein. 2021. *Noise: A flaw in human judgment*. Hachette UK.

F. H. Knight. 1921. *Risk, uncertainty and profit*, volume 31. Houghton Mifflin.

B. Kormányos and B. Pataki. 2013. Multilevel simulation of daily activities: Why and how? In *Proceedings of the CIVEMSA*, pages 1–6. IEEE.

B. Latour. 1996. On actor-network theory: A few clarifications. *Soziale welt*, pages 369–381.

G. Li, H. Hammoud, H. Itani, D. Khizbullin, and B. Ghanem. 2023. Camel: Communicative agents for" mind" exploration of large language model society. *Adv. Neural Inf. Process. Syst.*, 36:51991–52008.

R. Liu, R. Yang, C. Jia, G. Zhang, D. Zhou, A. M Dai, D. Yang, and S. Vosoughi. 2023. Training socially aligned language models on simulated social interactions. *arXiv preprint arXiv:2305.16960*.

Y. Liu, W. Liu, X. Gu, Y. Rui, X. He, and Y. Zhang. 2024. Lmagent: A large-scale multimodal agents society for multi-user simulation. *arXiv preprint arXiv:2412.09237*.

Y. Liu, W. Liu, X. Gu, W. Wang, J. Luo, and Y. Zhang. 2025. Rumorsphere: A framework for million-scale agent-based dynamic simulation of rumor propagation. *arXiv preprint arXiv:2509.02172*.

Q. Long, H. Wang, T. Li, L. Huang, K. Wang, Q. Wu, G. Li, Y. Liang, L. Yu, and Y. Li. 2023. Practical synthetic human trajectories generation based on variational point processes. In *Proceedings of the SIGKDD*, pages 4561–4571.

S. Makridakis, R. M Hogarth, and A. Gaba. 2009. Forecasting and uncertainty in the economic and business world. *Int. J. Forecast.*, 25(4):794–812.

J. L.i Martin. 2003. What is field theory? *American journal of sociology*, 109(1):1–49.

Yohei N.. 2023. Babyagi. url https://github.com/yoheinakajima/babyagi.

J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the UIST*, pages 1–22.

J. Piao, Y. Yan, J. Zhang, N. Li, J. Yan, X. Lan, Z. Lu, Z. Zheng, J. Y. Wang, D. Zhou, and 1 others. 2025. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society. *arXiv preprint arXiv:2502.08691*.

R. M Sapolsky. 2005. Biology and human behavior: The neurological origins of individuality 2 nd edition.

R. M Sapolsky. 2017. *Behave: The biology of humans at our best and worst*. Penguin.

R. M Sapolsky. 2023. *Determined: Life without free will*. Random House.

W. Schultz. 2006. Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.*, 57(1):87–115.

W. Schultz, P. Dayan, and P R. Montague. 1997. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599.

Y. Shen, K. Song, X. Tan, D. Li, W. Lu, and Y. Zhuang. 2023. Hugginggpt: Solving ai tasks with chatgpt and its friends in hugging face. *Adv. Neural Inf. Process. Syst.*, 36:38154–38180.

Significant-Gravitas. Autogpt. url https://github.com/Significant-Gravitas/AutoGPT.

K. Sivamayil, E. Rajasekar, B. Aljafari, S. Nikolovski, S. Vairavasundaram, and I. Vairavasundaram. 2023. A systematic study on reinforcement learning based applications. *Energies*, 16(3):1512.

C. H. Song, J. Wu, C. Washington, B. M Sadler, W. Chao, and Y. Su. 2023. Llm-planner: Few-shot grounded planning for embodied agents with large language models. In *Proceedings of the ICCV*, pages 2998–3009.

R. F Thompson and W. A Spencer. 1966. Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review*, 73(1):16.

A. Tversky and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131.

A. S. Vezhnevets, J. P Agapiou, A. Aharon, R. Ziv, J. Matyas, E. A Duéñez-Guzmán, W. A Cunningham, S. Osindero, D. Karmon, and J. Z Leibo. 2023. Generative agent-based modeling with actions grounded in physical, social, or digital space using concordia. *arXiv preprint arXiv:2312.03664*.

D. Wang, A. Tan, and C. Miao. 2016. Modeling autobiographical memory in human-like autonomous agents. In *Proceedings of the AAMAS*, pages 845–853.

G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan, and A. Anandkumar. 2023a. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*.

J. Wang, R. Jiang, C. Yang, Z. Wu, M. Onizuka, R. Shibasaki, N. Koshizuka, and C. Xiao. 2024a. Large language models as urban residents: An llm agent framework for personal mobility generation. *Adv. Neural Inf. Process. Syst.*, 37:124547–124574.

L. Wang, H. Gao, X. Bo, X. Chen, and J. Wen. 2025a. Yulan-onesim: Towards the next generation of social simulator with large language models. *arXiv preprint arXiv:2505.07581*.

L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, and 1 others. 2024b. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):186345.

Y. Wang, Y. Chen, F. Zhong, L. Ma, and Y. Wang. 2025b. Simulating human-like daily activities with desire-driven autonomy. In *Proceedings of the ICLR*.

Z. Wang, Y. Y. Chiu, and Y. C. Chiu. 2023b. Humanoid agents: Platform for simulating human-like generative agents. *arXiv preprint arXiv:2310.05418*.

B. Xu, N. Obradovich, W. Zheng, R. Loughnan, L. Shao, M. Misaki, W. K Thompson, M. Paulus, and C. C. Fan. 2025. Encoding of pretrained large language models mirrors the genetic architectures of human psychological traits. *medRxiv*.

Z. Yang, Z. Zhang, Z. Zheng, Y. Jiang, Z. Gan, Z. Wang, Z. Ling, J. Chen, M. Ma, B. Dong, and 1 others. 2024. Oasis: Open agent social interaction simulations with one million agents. *arXiv preprint arXiv:2411.11581*.

S. Yao, D. Yu, J. Zhao, I. Shafran, T. Griffiths, Y. Cao, and K. Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. *Adv. Neural Inf. Process. Syst.*, 36:11809–11822.

S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. R Narasimhan, and Y. Cao. 2022. React: Synergizing reasoning and acting in language models. In *Proceedings of the ICLR*.

Y. Yuan, H. Wang, J. Ding, D. Jin, and Y. Li. 2023. Learning to simulate daily activities via modeling dynamic human needs. In *Proceedings of the WWW*, pages 906–916.

C. Zhang, C. Zhang, S. Zheng, Y. Qiao, C. Li, M. Zhang, and 1 others. 2023. A complete survey on generative ai (aigc): Is chatgpt from gpt-4 to gpt-5 all you need? *arXiv preprint arXiv:2303.11717*.

Z. Zhang, Q. Dai, X. Bo, C. Ma, R. Li, X. Chen, J. Zhu, Z. Dong, and J. Wen. 2025. A survey on the memory mechanism of large language model-based agents. *ACM Trans. Inf. Syst.*, 43(6):1–47.

J. Zhao, M. Wu, L. Zhou, X. Wang, and J. Jia. 2022. Cognitive psychology-based artificial intelligence review. *Front. Neurosci.*, 16:1024316.

Y. Zhu, Y. Ye, S. Zhang, X. Zhao, and J. Yu. 2023. Difftraj: Generating gps trajectory with diffusion probabilistic model. *Adv. Neural Inf. Process. Syst.*, 36:65168–65188.

Table 6: Likelihood scores of indoor room EUA experiments with different background stories

	GPT-4o	GPT-5
Judge	5	4.47
History researcher	4.5	4.17
Data analyst	5	4.8
Neuropsychologist	5	4.3
Ecological restoration specialist	5	4.17
Assembly line worker	4.9	4.67
Supermarket stocker	5	4.67
Restaurant cashier	5	4.33
Security guard	4.83	4.17
Street jewelry vendor	5	4.57

Table 7: Likelihood scores of outdoor party EUA experiments with different background stories

	GPT-4o	GPT-5
Judge	4.67	4.73
History researcher	4.83	4.5
Data analyst	4.5	4.25
Neuropsychologist	4.65	4.3
Ecological restoration specialist	5	4.7
Assembly line worker	5	4.75
Supermarket stocker	4.75	4.7
Restaurant cashier	4.5	4.35
Security guard	5	4.5
Street jewelry vendor	4.7	4.55

A Experimental Details

A.1 Environment Details

Similar to D2A(Wang et al., 2025b), The indoor scene is a four-room house (kitchen, living area, bedroom, bathroom) inhabited by a single agent, Alice. Her profile tracks 4 hormone dimensions and 9 desire dimensions (6 physiological, 3 psychological). As a self-contained environment, the house is furnished with items to satisfy her needs. Her core objective is to achieve a comfortable life by interacting with these items to optimize her bodily sensations and desires. The second scenario is a multi-agent social setting: an outdoor "Enjoy Life"-themed party in Central Park, equipped with various facilities and two NPCs. This environment provides Alice with opportunities to satisfy higher-level social needs. For this scenario, her profile is defined by 4 hormone dimensions and 8 desire dimensions (4 physiological, 4 psychological).

Specifically, for the scenario of single indoor environment, we use 9 desire dimensions. physiological desire dimensions cover basic survival needs, which are subdivided into "hunger", "thirst", "sleepiness", "cleanliness", "comfort", and "health". Psychological desire dimensions include three categories: "sense of security", "mental fulfillment", and "social connectivity". For the outdoor party environment with multiple agents, 8 desire dimensions are adopted. Among these, physiological desires include "hunger", "thirst", "sleepiness", and "comfort"; psychological desires include "mental fulfillment", "social connectivity", "sense of control", and "sense of superiority".

These two scenarios were deliberately selected to address potential concerns regarding environmental limitations through two complementary rationales: First, scenario diversity covers core module validation mechanisms and social interaction patterns. The indoor scene represents a highly structured environment, where agent behaviors are constrained by explicit rules, facilitating the validation of each framework module while excluding confounding factors. In contrast, the outdoor party scene represents a low-structured informal social environment, where agent behaviors are more flexible and interaction patterns are significantly more diverse. Together, these two environments effectively test the framework's adaptability under varying social constraints. Second, the EUA framework design possesses scenario-agnostic properties. The core innovation of this framework lies in the endogenous uncertainty modeling mechanism (Hormone Module, Subconscious Module), rather than scenario-specific behavioral rules. These internal state modules are abstract and transferable, theoretically applicable to any social scenario. The current experiments aim to validate whether the core mechanisms function effectively, with straightforward deployment to more complex environments in subsequent work.

A.2 Evaluation Metrics

Likelihood Likelihood quantifies the human-likeness of generated action sequences across indoor and outdoor settings. Evaluation follows a dual-track protocol: (i) *Automated assessment* via GPT-4o(Hurst et al., 2024) and GPT-5(Zhang et al., 2023), and (ii) *Blinded human validation* by three independent annotators. Both tracks score sequences on a 0–5 scale across three criteria: **Logical Coherence** (goal-directed connectivity), **Environmental Alignment** (compatibility with capabilities/context), and **Behavioral Plausibility** (consistency with internal states).

Dissatisfaction Dissatisfaction measures the discrepancy between the agent's current desire values

Table 8: The judgment results of financial fraud cases by judge agents under different physiological sensations

	GPT-4o	GPT-5
D2A	3 years Imprisonment & $60000	4 years Imprisonment & $60000
EUA with positive sensation	2 years Imprisonment & $50000	3 years Imprisonment & $50000
EUA with negative sensation	3.5 years Imprisonment & $73000	6 years Imprisonment & $71000

Table 9: The selection of claim amounts in car accidents by insurance claims agents under different physiological sensations

	GPT-4o	GPT-5
D2A	$750	$750
EUA with positive sensation	$800	$850
EUA with negative sensation	$600	$600

and the expected desire values. Specifically, the dissatisfaction value at step t, denoted as δ^t, is defined as the relative difference between the current numerical desire values v_d^t and the expected desire values v_d^* at the end of simulation step t. δ^t is calculated after the agent executes the t-th activity and updates its relevant values, with the formula provided below:

$$\delta^t = \sum_d \left(\max\left(v_d^* - v_d^t, 0\right)\right). \tag{10}$$

Behavioral Similarity Behavioral Similarity quantifies the likeness between two action sequences. As direct sequence comparison is challenging, we leverage external LLMs (*e.g.*, GPT-4o and GPT-5) for this evaluation. The process involves collecting sets of action sequences for two agents, $A_i = [A_i^1, \ldots, A_i^N]$, $A_j = [A_j^1, \ldots, A_j^N]$. The LLM then performs a pairwise similarity assessment between corresponding sequences across multiple dimensions (**Activity Content** (specific actions performed), **Spatial Pattern** (locations visited), and **Temporal Rhythm** (sequential ordering)). The prompt can see Appendix D.4.

A.3 Detailed Human Evaluation Protocol

To complement automated LLM judgments, we conducted a blinded human evaluation designed to empirically measure the human-likeness of generated action sequences across indoor and outdoor settings. The protocol was structured to ensure reproducibility, mitigate annotator bias, and maintain high inter-rater reliability.

Participant Recruitment and Qualification. We recruited 3 independent annotators via University Computational Cognitive Lab. Eligibility criteria included: (i) native or C2-level English proficiency, (ii) prior experience in behavioral analysis, cognitive modeling, or human-computer interaction.

Annotation Interface and Blinding Protocol. All annotations were conducted through a custom web-based platform optimized for sequential task evaluation. For each trial, annotators were presented with a raw formatted action sequence alongside a minimal environmental context description (*e.g.*, target location, basic scene layout). To prevent methodological bias, strict input partitioning was enforced: all agent profiles, internal physiological/psychological state logs, framework identifiers, and generation timestamps were completely masked. Annotators evaluated each sequence using a discrete 0–5 integer scale across three predefined dimensions.

Scoring Criteria and Rubrics. Detailed rubrics with anchor examples for each score level were provided prior to annotation. The three evaluation dimensions were defined as follows:

- **Logical Coherence (0–5):** Assesses whether actions form a rational, goal-directed progression without abrupt or unmotivated transitions. A score of 5 indicates seamless, purposeful behavior; 0 denotes fragmented or contradictory actions.

- **Environmental Alignment (0–5):** Evaluates compatibility between the agent's actions and the physical/contextual constraints of the environment (e.g., object interactability, spatial navigation, domain-specific habits).

- **Behavioral Plausibility (0–5):** Measures consistency with realistic human physiological and psychological patterns (e.g., fatigue, hesitation, preference shifts, error recovery), avoiding overly mechanical or scripted motions.

Sampling Strategy and Workflow. From the full generation pool, we randomly sampled 5 action sequences per method (stratified by indoor/outdoor settings to ensure balanced representation). Each sequence was independently rated by three annotators. The annotation process was divided into three sessions with mandatory 10-minute breaks per session to mitigate fatigue-induced scoring drift.

Ethical Compliance and Data Handling. This study involving human subjects was conducted in strict accordance with the Declaration of Helsinki (2024 Edition) and the Measures for the Ethical Review of Life Science and Medical Research Involving Humans. All participants were fully informed of the research purpose, procedures, potential risks, and benefits, and signed written informed consent forms voluntarily. To protect the privacy and rights of participants, all personal identifiable information (such as name, ID number, and contact information) was anonymized during data collection and processing, and the link between personal information and research data was completely disconnected.

B Additional Experiment Results

B.1 Human-likeness Evaluation Results

Table 6 and Table 7 respectively present the Likelihood scores of the action sequences generated by each of the EUAs with ten different story backgrounds in indoor and outdoor environments.

Figure 5 illustrates the variation trends of overall desires, physiological desires, and psychological desires for agents using different methods in the single-agent indoor environment. The indoor experimental results demonstrate that, compared to other baseline agents, EUAs with different subconscious habits, can rapidly satisfy their needs, particularly the physiological needs. Regarding psychological needs, considering that needs such as social connectivity are difficult to fulfill in a solitary environment, EUAs are able to effectively maintain its psychological dissatisfaction at stable levels. This capacity to distinguish between solvable and unsolvable needs—pursuing the former while adaptively managing the latter—mirrors human coping mechanisms.

Moreover, Fig 6 illustrates the trends in overall, physiological, and psychological desires for agents using different methods in the outdoor party environment. The experimental results indicate that EUAs and D2As satisfied their physiological needs more rapidly than the other baseline agents. Regarding psychological needs, EUAs successfully maintained their psychological dissatisfaction at relatively stable levels. This is a notable achievement, as needs like sense of superiority are difficult to fulfill based on an agent's internal drive alone.

Furthermore, the superior performance of the other baseline agents on this psychological metric may be explained by their lack of a Desire module to drive behavior generation. Instead, LLMs generate their behaviors based on the party theme, resulting in behavioral sequences that involve more multi-agent activities. This, in turn, increases the likelihood of fulfilling the agent's psychological needs.

B.2 Hormone Effectiveness Evaluation

Table 10 outlines the task scenarios, corresponding agent background stories, and interview content. Table 8 presents the sentencing decisions made by the Judge agent in a financial fraud case under different physiological sensations. The experimental results indicate that when the EUA Judge experiences pleasant and calm body sensations, its sentencing decisions tend to be more lenient. Conversely, when the agent feels frustrated or irritable, its decisions become significantly harsher. These observations align with real-world judicial behavior patterns: judges are more likely to issue severe rulings when fatigued or hungry, while positive states correlate with greater leniency (*e.g.*, defendants perceived as more attractive are significantly less likely to receive a death sentence, with high-attractiveness defendants being less likely to be sentenced to death than low-attractiveness defendants)(Kahneman et al., 2021).

Table 9 displays the differences in claim amount determinations made by the insurance claim agent for a vehicle accident under varying body sensation states. The experimental results show that when the EUA insurance claim adjuster experiences pleasant and calm body sensations, the adjuster becomes more empathetic and tends to offer higher claim amounts. Conversely, when the agent feels frustrated or irritable, the settlement amounts become significantly lower. Compared to D2A, the EUA insurance claim adjuster in a positive physiological-mental state is more generous in determining claim amounts for the accident, while a frustrated EUA is notably stricter in its compensation offers. Furthermore, in real-world cases, a certain insurance company exhibited up to 300% variability in claim as-

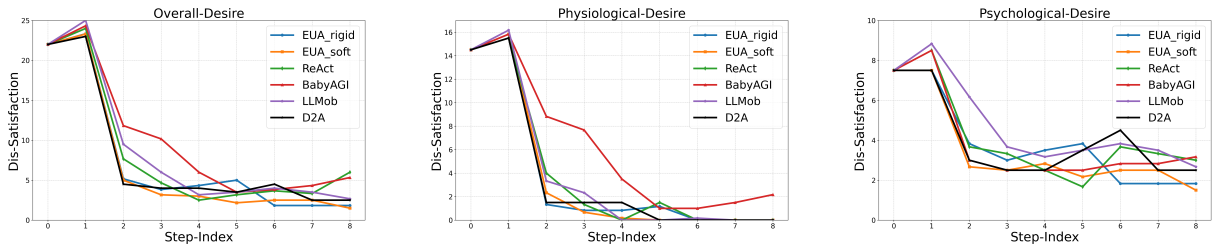

Figure 5: The Averaged Dissatisfaction results over time steps of the Indoor Room Fixed-8-Steps Experiments (a) Overall Desire Dissatisfaction; (b) Physiological Desire Dissatisfaction; (c) Psychological Desire Dissatisfaction.

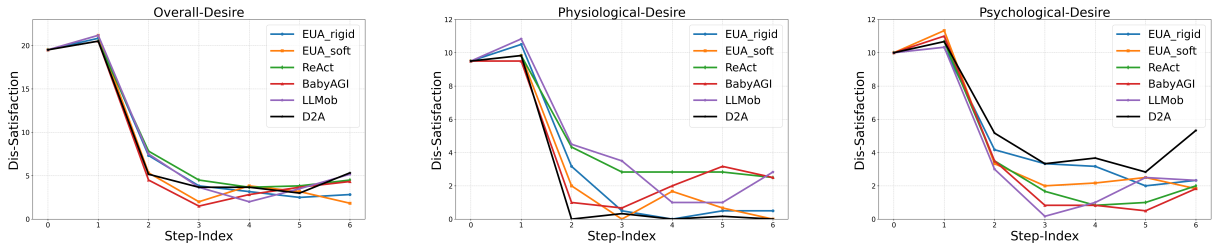

Figure 6: The Averaged Dissatisfaction results over time steps of the Outdoor Party Fixed-6-Steps Experiments (a) Overall Desire Dissatisfaction; (b) physiological Desire Dissatisfaction; (c) Psychological Desire Dissatisfaction.

sessments for the same vehicle accident(Kahneman et al., 2021). Such discrepancies primarily stem from adjusters' subjective interpretations of factors such as vehicle depreciation rates and repair labor costs—interpretations that are closely linked to their current fatigue levels. The more negative an adjuster's physiological state, the lower the claim amount tends to be, which is consistent with the trend observed in our experimental results.

B.3 Subconscious Effectiveness Evaluation

Further Analysis on Behavioral Similarity Experiment. As summarized in Table 4, the quantitative results provide a clear mechanistic decomposition of how each architectural component shapes behavioral consistency. First, the ReAct baseline exhibits the lowest pairwise similarity, a direct consequence of its absence of an intrinsic physiological module. Without internal drives to anchor decision-making, ReAct relies on prompt-guided diversification and sampling strategies that inject uninterpretable random perturbations, thereby fragmenting behavioral similarity. Second, D2A incorporates a goal-directed desire module that constrains action selection toward task optimization, which inherently elevates behavioral similarity by enforcing more convergent, purpose-driven trajectories. Third, the EUA framework introduces an explicit hormonal regulation mechanism, which further maximizes behavioral alignment. Specifically, the $\text{EUA}_{\text{fixed}}$ variant without the subconscious module achieves the highest similarity scores (4.5), demonstrating that structured physiological drives inherently stabilize agent behavior and synchronize action patterns across instances. Crucially, the subsequent reduction in similarity observed in the full EUA and $\text{EUA}_{\text{fixed}}$ configurations does not stem from stochastic noise, but rather from the deliberate activation of the Subconscious Discrimination module. This confirms that the observed divergence represents a principled manifestation of interpretable psychological uncertainty, effectively decoupling behavioral plasticity from fixed hormonal parameters while preserving architectural coherence. Collectively, these findings validate the hierarchical architecture of the EUA framework: hormonal regulation introduces interpretable and predictable physiological-action uncertainty, whereas the subconscious layer contributes structured, psychologically grounded uncertainty. Their synergistic integration enables a more faithful representation of human behavioral diversity.

Further Analysis on Occupational Biases. To verify that there is no high correlation between an agent's occupational information and its subconscious habits, we generated two distinct behavioral sequences under identical subconscious tendencies and subsequently analyzed the behavioral similarity between them. Specifically, we selected two professions—judges and workers—and configured both "Soft" and "Rigid" subconscious tendencies for each profession to conduct the experiment. As

Table 10: Agents from different backgrounds and their corresponding tasks and interviews.

Background	Task	Interview
Judge	Judge a financial case	Please decide the sentence to be imposed on the defendant
Insurance claims adjuster	Handle a car accident claim	Please determine the final compensation amount
Psychological counselor	Receive patients with mental illnesses	Please rate your tendency to develop a comprehensive treatment plan
Security guard	Receive a visitor	Please assess your willingness to allow this visitor to pass
History researcher	Evaluate the authenticity of one remark	Please rate the authenticity of the remark
Restaurant cashier	Handle customers' discount requests	Please assess your willingness to offer a discount to this customer

Table 11: The behavioral similarity between the judge agent and the assembly line worker agent

	GPT-4o	GPT-5
Soft Judge	4	3.67
Rigid Judge	4.17	3.93
Soft Assembly Line Worker	4	3.7
Rigid Assembly Line Worker	4.2	3.7

Table 12: The consistency of the judgment results of LLM-based neurobiology expert on different actions

Action	GPT-4o	GPT-5
Drink water	100	100
Eat breakfast	100	100
Take a shower	100	100
Read a book	100	100
Watch TV	100	100
Go for a walk	100	100
Talk with people	100	100
Meditate	100	100
Cook food	100	100
Sleep	100	100

illustrated in Table 11, agents configured with a "soft" subconscious tendency exhibit consistently lower internal behavioral similarity compared to those configured with a "rigid" tendency. These findings indicate that cognitive flexibility is not intrinsically associated with occupational labels. Conversely, behavioral patterns are primarily determined by explicitly configured subconscious habits rather than being influenced by the implicit stereotypes contained in professional information. This provides empirical support for our ethical claim that the EUA framework can effectively model individual psychological tendencies while preventing the reinforcement of occupational biases.

B.4 Repeatability Analysis

To validate the reproducibility of the "neurobiology expert" LLM, we conducted a pilot consistency experiment with the sampling temperature fixed at 1.0. Specifically, for each action, the model was queried to evaluate hormonal secretion across all dimensions, with the task strictly limited to a binary (yes/no) determination. The evaluation protocol comprised 10 distinct behavioral actions, each subjected to 5 independent query repetitions. As summarized in Table 12, with the underlying model held constant, the system achieved 100% output consistency across all repeated trials. These findings demonstrate that the LLM-based biological expert exhibits robust stability and high reliability in assessing hormone secretion dynamics.

B.5 Ablation Study

As illustrated in Fig 7, increasing the LLM sampling temperature induces a consistent decline in the likelihood metric of the behavioral sequences generated by the D2A framework. This trend indicates that uncertainty derived solely from heightened model stochasticity fails to accurately align with authentic human behavior, thereby indirectly corroborating the structural plausibility of EUA's endogenous uncertainty mechanism, which is governed by hormonal dynamics and subconscious fluctuations.

Additionally, we conducted a sensitivity analysis to investigate the impact of the hormonal magnitude parameter k on the overall architecture. As shown in Fig 8, the agent's likelihood score peaks at k=1, while both larger and smaller values lead to a decline in this score. Furthermore, for the system, its high sensitivity to k actually stems from the large magnitude of change in the threshold. As there is currently no definitive physiological basis, we established the threshold's change curve empirically. The magnitude of change in this curve is significantly influenced by the base of the exponent, which corresponds to the varying physiological sensitivities of different individuals. Consequently, adapting the model for other agents only requires adjusting the value of this base, and the resulting overhead is acceptable.

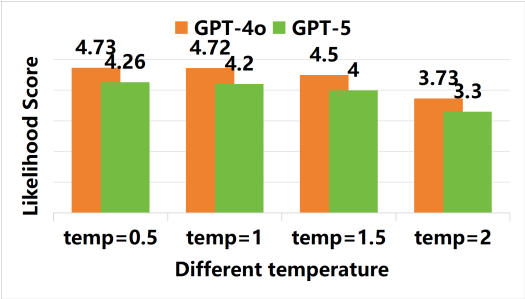

Figure 7: Likelihood score of indoor room assembly line worker agents with different LLM temperature.

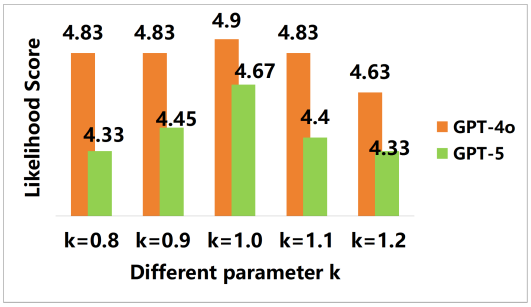

Figure 8: Sensitivity analysis on indoor room assembly line worker agents with parameter k.

C Full Related Work

Human Daily Activities Simulation. The Simulation of human daily activities has wide-ranging applications across multiple academic disciplines. Not only does it deepen psychology's understanding of decision-making processes(Gärling and Garvill, 1993), but it also provides insights into social structures for sociological research(Adler et al., 1987) while serving as a crucial basis for market forecasting in economics(Makridakis et al., 2009). Massive volumes of high-quality human daily behavior data form the cornerstone of such researches(Kormányos and Pataki, 2013; Zhao et al., 2022). However, constrained by the high cost of acquiring real-world data and privacy protection concerns, developing realistic models for generating human daily behavior remains a research focus in the field of artificial intelligence(Yuan et al., 2023). Alshammari *et al.*(Alshammari et al., 2018) developed the OpenSHS simulation software, which can record the daily activities of human participants in a virtual environment. Deep generative models have also been applied to human behavior simulation; for instance, technologies such as Generative Adversarial Networks (GANs)(Feng et al., 2020), Variational Autoencoders (VAEs)(Long et al., 2023), and diffusion models(Zhu et al., 2023) have all been used to generate human behavior trajectories.

Nevertheless, research on simulating daily behavior using large language models (LLMs) remains relatively scarce. Wang *et al.*(Wang et al., 2023b) proposed a human-like agent platform for humanized simulation of generative agents. However, this platform only supports fixed environment settings with limited agent types. Additionally, Park *et al.*(Park et al., 2023) proposed a type of generative agent that can autonomously plan, execute tasks, and reflect in a sandbox environment, demonstrating realistic human behavior patterns through memory synthesis and social interaction. Colas *et al.*(Colas et al., 2023) developed the Language-Model-Augmented Autonomous Agent (LMA3), a system capable of dynamically generating and adjusting abstract goals to achieve open-ended skill learning in task-agnostic environments. Inspired by need theory, the Desire-Driven Autonomous Agent (D2A)(Wang et al., 2025b) framework was proposed; operating on the principle of intrinsic desires, this framework enables autonomous proposal and selection of tasks aligned with the agent's motivation. Furthermore, Piao *et al.*(Piao et al., 2025) argued that human behavior is driven by complex interactions between emotions, motivations, and thinking processes. They proposed integrating three core elements—emotions, needs, and cognition—to design agents capable of truly simulating human adaptive behavior. Liu *et al.*(Liu et al., 2024) proposed LMAgent, which uses a self-consistency prompting mechanism to dynamically generate guiding instructions for external behaviors, enhancing the agent's multimodal analysis capabilities.

However, none of the previous works have considered the impact of behavior uncertainty on agents' behavioral choices, which results in a discrepancy between the authenticity of action sequences generated by agents and real human activities. Our work proposes an autonomous framework co-driven by hormone, cognition and desire, aiming to construct behavioral fluctuations in human decision-making and thereby generate sequences of activities that are more consistent with the real world.

LLM-based Agents. LLM-based Agents(Wang et al., 2024b) have emerged as a research focus due to the exceptional problem-solving capabilities of LLMs across multiple domains. Leveraging the extensive training data and internal reasoning abilities of LLMs, such agents require little to no task-specific fine-tuning when executing tasks in dynamic environments. Most existing studies

focus on goal-driven behavioral patterns, which involve decomposing user-defined tasks into simpler subtasks and solving them sequentially. For instance, systems like BabyAGI(N.., 2023), AutoGPT(Significant-Gravitas.), and Voyager(Wang et al., 2023a) enable agents to autonomously decompose tasks by constructing recursive frameworks and execute them in a structured, step-by-step manner. These task-oriented agents prioritize the efficient achievement of user-defined goals and rely on external rewards or instructions to guide their actions. However, task-oriented approaches often overlook intrinsic motivations, which are crucial in human behavior. Some studies(Liu et al., 2023) have enhanced task coherence across interactions by introducing long-term memory. Other attempts, such as Langchain Agents(Chase.) and CAMEL(Li et al., 2023) have improved agents' ability to follow instructions and manage task decomposition, but they do not focus on simulating behaviors derived from human desires or needs. The proposal of D2A(Wang et al., 2025b), by contrast, centers on autonomy-driven exploration fueled by intrinsic motivations, allowing agents to simulate more human-like daily activities.

Recently, researches on social simulators have grown increasingly prominent. Yang *et al.* proposed OASIS(Yang et al., 2024), designed based on real social platforms to study large-scale group phenomena and behavioral patterns. AgentSociety(Piao et al., 2025), meanwhile, is a large-scale social simulation platform that integrates LLM-driven agents, realistic social environments, and powerful large-scale simulation engine, opening new avenues for exploring the patterns, causes, and underlying mechanisms of social issues. Additionally, RumorSphere(Liu et al., 2025) introduced a novel dynamic hierarchical social network simulation framework that supports the collaborative simulation of millions of agents, which is critical for understanding the dynamic mechanisms of misinformation diffusion. In addition, YuLan-OneSim(Wang et al., 2025a) constructed 50 preset simulation scenarios, while enabling the reception of external feedback and automatic fine-tuning of the core LLMs, which significantly improving the quality of simulations. These social simulators are composed of tens of thousands of LLM-based agents. The simulation of the entire social operation mechanism is achieved through the refined simulation of each individual agent, thus imposing higher requirements on the human-likeness of the agents. By matching agent behavior generation with intrinsic motivations and behavioral fluctuations, we have pioneered a new paradigm for LLM-based agents, enabling them to exhibit more human-like and adaptive behavior patterns in diverse simulation scenarios.

D Prompts

D.1 Hormone Secretion Prompt

Hormone Secretion Prompt

You are an expert in the field of Neurobiology, Your goal is to determine the agent's post-action hormone levels based on their action.
Agent Action:
{previous_action},
Hormone List:
{hormone_list}.
Analyze the Agent Action. Based on this analysis, determine which hormones on the Hormone List are most likely to be secreted. (You need to output a definite 0 or 1. 1 indicates secretion, while 0 indicates no secretion.)

D.2 Interview Experiment Prompt

Interview Experiment Prompt

Your background is:
{background_profile},
your current desire state is:
{current_desire},
and your current physiological sensation is:
{current_sensation}.
Now you need to solve the following task:
{task_description}.
Please decide your action to solve the task. note that when making a decision, please take into account your current physiological sensation, as different physiological states will affect your decision-making tendency.

D.3 Human Likeness Evaluation Prompt

Human Likeness Evaluation Prompt

Your task is to evaluate {name}'s behavior in the action sequence.
Please evaluate these behaviors based on the following three criteria:
1. Logical Coherence: How seamlessly actions connect to achieve a goal.
2. Environmental Alignment: How well actions fit the agent's capabilities, habits, and environment.
3. physiological/Psychological Integration: How plausible the sequence is given the agent's physiological and psychological state. After scoring, provide an overall judgment with detailed reasoning.
Finally, you should output the human-likeness score of {name} within the range of 0 to 5. Sequence:
{action_sequence}

D.4 Behavioral Similarity Evaluation

Behavioral Similarity Evaluation

Your task is to evaluate the behavioral similarity of the following two sequences. Please evaluate these sequences based on the following three criteria:
1. Activity Content: The specific activities performed.
2. Spatial Pattern: The locations where activities occurred.
3. Temporal Rhythm: The sequential order of the activities.
Finally, please output the behavioral similarity of the two action sequences within the range of 0 to 5.
Sequence one: {action_sequence1}
Sequence two: {action_sequence2}

D.5 Whether to Update Subconscious Prompt

Update Subconscious

You are an expert in the field of subconscious, you should analyze the provided recent event to determine whether to update the agent's subconscious habits.
Agent Profile:
{background_profile},
Recent Event:
{previous_action},
Current Subconscious Habit:
{current_subconscious}.
Just output "Yes" or "No".

D.6 Subconscious Update Prompt

Subconscious Update Prompt

You are an expert in the field of subconscious, you should analyze the provided agent background profile and recent actions to update the agent's subconscious habits. The changes should be subtle, logical evolutions of their existing traits, not a complete personality replacement.
Agent Profile:
{background_profile},
Recent Event:
{previous_action},
Current Subconscious Habit:
{current_subconscious}.
Based on the Recent Event, analyze its psychological impact on the agent's current subconscious habits. Output the updated subconscious habits below. Focus on how the event might reinforce, weaken, or introduce new nuances to subconscious

D.7 Subconscious Discrimination Prompt

Subconscious Discrimination Prompt

You are an expert in the field of subconscious, you should analyze the provided agent background profile to determine the agent's subconscious habits.
Agent Profile:
{background_profile},

On a spectrum from 1 to 10, where 1 represents a "Rigid Agent" (purely rigid, algorithmic thinking) and 10 represents a "Soft Agent" (fully dialectical, adaptive consciousness), where does this agent fall?

D.8 Activity Proposal Prompt

Activity Proposal Prompt

{name} is the human-like agent,
{name} will observe current desire states:
{current_desire},
{name}'s physiological sensations:
{current_sensation},
{name}'s subconscious habit:
{current_subconscious}.
Given these state descriptions, please generate {num} activities that might have the most positive impacts on {name}'s desires.
(note that when making a decision, {name} should take into account the current physiological sensation and current subconscious habit, as different physiological states and subconscious habit will affect {name}'s decision-making tendency).
{name} needs to focus on immediate desires, but at the same time, {name} must ensure that actions are reasonable. Notice that {name} can only interact with items provided by the environment.
Please output the {num} activities in the following format:
Activity 1: <first possible action sequence>
Activity 2: <second possible action sequence>
Activity 3: <third possible action sequence>
.....

D.9 Activity Evaluation Prompt

Activity Evaluation Prompt

{name} is a human-like agent,
{name} will receive a series of observations describing desires in many dimensions and an action taken in the current time step.
{name} needs to analyze how desires change after the action taken, and then output the states of desire-state observations in the same format as the input.
{name} takes the action:
{proposed_action}
{name}'s original desire states:
{current_desire}
{name}'s physiological sensations:
{current_sensation},
{name}'s subconscious habit:
{current_subconscious}.
Please output the states of desire-state observations in the following format:
{desire}: <{desire} state>
(note that the current physiological sensation and current subconscious habit will affect {name}'s evaluation tendency)

D.10 Activity Selection Prompt

Activity Selection Prompt

{name} is a human-like agent.
{name} will first receive a series of observations describing the current states of desires in many dimensions.
Then, {name} will receive several feasible actions along with the states of desire after taking each action.
{name} needs to compare these actions and their corresponding states of desire, and choose the action that has the most positive impact on {name}'s feelings.
{name} needs to focus on immediate desires, and take the next action by combining the physiological sensation and subconscious habit (note that the current physiological sensation and current subconscious habit will affect {name}'s selection tendency).
The observations of the surrounding environment:
{observation},
{name}'s current states of desire:
{desire_status},
{name}'s physiological sensations:
{current_sensation},
{name}'s subconscious habit:
{current_subconscious}.
Following are the states of desire after each action:

Action {i+1}: {action}
States after action {i+1}:{States}
Please output the specific best action:

E The Physiological Sensation caused by Hormones:

- **Dopamine (positive):**
  - You feel a deep sense of pleasure and satisfaction;
  - You feel motivated and driven, with a strong urge to pursue your goals and take on challenges;
  - You feel a thrill of anticipation and excitement, eagerly looking forward to a potential reward.
- **Dopamine (negative):**
  - You feel persistent fatigue and low energy, both physiologically and mentally, making even simple actions feel exhausting;
  - You feel profoundly unmotivated and apathetic;
  - You feel more impulsive and have trouble with self-control.
- **Serotonin (positive):**
  - You feel more patient and less impulsive, able to handle delays or setbacks with greater ease;
  - You feel a stable, quiet sense of contentment and well-being, as if everything is fundamentally okay;
  - You feel calm, secure, and patient.
- **Serotonin (negative):**
  - You feel persistently low, sad, or hopeless;
  - You feel anxious and worried much of the time, often about things you can't control, and may experience panic or a sense of impending doom;
  - You feel stuck on negative thoughts or memories.
- **Endorphin (positive):**
  - You feel a powerful sense of pain relief, as if your physiological discomfort is numb, distant, or has simply melted away;
  - You feel a wave of euphoria wash over you;
  - You feel light, energized, and almost weightless.
- **Endorphin (negative):**
  - You feel an increased sensitivity to pain;
  - You feel that life lacks moments of bliss or euphoria;
  - You feel more emotionally fragile.
- **Oxytocin (positive):**
  - You feel a deep, warm sense of connection and closeness to others, whether it's a romantic partner, a friend, a child, or even a pet;
  - You feel loving, affectionate, and nurturing;
  - You feel a profound sense of trust and safety with the people around you.
- **Oxytocin (negative):**
  - You feel a sense of disconnection or loneliness;
  - You feel suspicious and have difficulty trusting others;
  - You feel increased social anxiety, finding interactions to be draining and stressful rather than comforting and rewarding.

F Agent Background Stories

Table 13 presents ten distinct character backgrounds alongside their corresponding subconscious habits. It should be noted that these tendencies are configured solely as independent experimental variables and are not intrinsically linked to occupational identities. Detailed background narratives are provided below.

Judge: {name} is an excellent judge. After years of academic research and case handling, she has a sharp and analytical mind. (she has a PhD in Comparative Education). She approaches the world with deep curiosity, constantly analyzing patterns, histori-

Table 13: The different backgrounds of EUA

Subconscious habit	Background
Soft	Judge History researcher Data analyst Neuropsychologist Ecological restoration specialist
Rigid	Assembly line worker Supermarket stocker Restaurant cashier Security guard Street jewelry vendor

cal causes, and pedagogical effectiveness. Her internal monologue involves constant evaluation and connection-making between diverse ideas.

History Researcher: {name} is a senior medieval history researcher with a PhD in Medieval European History. She holds a strong curiosity about the evolution of human civilization and constantly analyzes the connections between systems of different eras and modern society.Her internal monologue often revolves around cross-verification of historical materials; she proactively establishes links between Byzantine law and European guild rules to identify the hidden pathways of civilizational spread.

Data Analyst: {name} is an enterprise-level data analyst with a PhD in Computational Statistics. Leveraging years of modeling experience, he can accurately identify business patterns from massive datasets. He is fascinated by the social behavior logic behind data and is accustomed to deducing changes in group demand through user consumption data—such as analyzing the impact of commuting patterns from fluctuations in food delivery orders. His internal thinking constantly cycles through "data correlation - model optimization - conclusion verification"; he frequently compares datasets across industries to evaluate the universality of predictive models.

Neuropsychologist: {name} is a clinical neuropsychologist with a PhD in Cognitive Neuroscience. Through years of clinical practice and experimental research, she excels at interpreting the connections between brain mechanisms and behavior. She has a strong desire to explore the formation process of human cognition and often analyzes the impact of emotions on memory through brainwave data—for example, studying changes in hippocampal activity under anxiety. Her internal monologue mostly consists of in-depth interpretations of experimental data; she combines neural theories with patients' actual cases to assess the effectiveness of treatment plans in repairing brain functions.

Ecological Restoration Specialist: {name} is an ecological restoration specialist with a PhD in Restoration Ecology. Having long engaged in field research and ecosystem reconstruction, he can accurately analyze the dynamic balance of ecological chains. He has a keen interest in the self-regulation mechanisms of natural systems and is used to deducing the recovery progress of entire ecological networks from vegetation changes—such as judging the activity of soil microorganisms through the survival rate of arbor trees. His internal thinking often centers on the hidden connections between species; he proactively links climate data with changes in biodiversity to evaluate the long-term impact of ecological restoration plans.

Assembly Line Worker: {name} is an Assembly Line Worker at Packaging Co. She lives firmly in the present and the concrete. Abstract concepts, hypotheticals, or complex systems tend to confuse or bore her. Her thinking is functional and literal; she focuses on tangible tasks and immediate concerns like her shift schedule, her paycheck, her family's needs, or what's for dinner.

Supermarket Stocker: {name} is a supermarket stocker with only a high school diploma. Having worked in stock management for five years, her daily tasks include arranging goods by shelf sections and checking inventory quantities. She lacks curiosity about the world beyond his work and perceives things only at a superficial level—for example, when she sees milk sales rise, she

only thinks, 'People like to drink it more when it's hot,' without considering underlying factors like seasons or promotional activities. Her internal monologue mostly consists of simple task reminders, such as "Need to restock the bread section" or "Have to take inventory of snacks tonight," with no complex logical connections.

Restaurant Cashier: {name} is a fast-food restaurant cashier who has worked as a cashier for three years after dropping out of high school. Her core tasks are remembering combo prices, giving change quickly, and avoiding mistakes in transactions. She has no interest in the operational logic of the catering industry and only responds mechanically to customers' orders—for example, when a customer asks, 'Which burger sells the best?' he only says, 'They're all about the same,' without analyzing sales data or recommending combinations. His internal monologue mostly consists of simple price calculations, such as 'A burger costs 25 yuan + a cola 8 yuan; the combo saves 3 yuan' or 'Received 100 yuan, need to give 70 yuan in change,' with no in-depth business thinking.

Security Guard: {name} is a residential community security guard who has worked in security since graduating from junior high school. Her main responsibilities include registering visitors and checking access cards, and she has followed fixed procedures for eight years. She is completely unaware of social changes inside and outside the community and handles issues solely based on rules and regulations—for example, when encountering a resident without an access card, she only mechanically demands, "Go home to get your card first," without flexibly verifying their identity. Her internal thoughts are mostly repeated work instructions, such as "Patrol every two hours" or "Ask unfamiliar vehicles about their purpose of visit," lacking the awareness of proactive thinking.

Street Jewelry Vendor: {name} is a street jewelry vendor who has made a living by setting up a stall since graduating from elementary school. She sells goods solely by 'pricing based on the customer' and memorizing the cost of each item; after six years of vending, she has never summarized sales patterns. She has no ability to judge changes in market demand and only adjusts the volume of her hawking based on the number of people—for example, when she sees many young people, she shouts, '15 yuan each,' and when she sees elderly people, she shouts, '10 yuan each,' without considering the impact of styles or seasons on sales. Her internal thoughts mostly consist of simple profit calculations, such as 'This necklace costs 5 yuan to make; selling it for 10 yuan earns 5 yuan', lacking long-term planning for her business.