scientific reports

Check for updates

OPEN

Extending homeostasis to thought dynamics for a comprehensive explanation of mind-wandering

Kazushi Shinagawa[1,2]✉ & Kota Yamada[3]

Our thoughts are inherently dynamic, often wandering far from the current situation. Mind-wandering (MW), which is these thought transitions, is crucial for understanding the nature of human thought. Although previous research has identified various factors influencing MW, a comprehensive framework integrating these findings remains absent. Here, we propose that homeostasis has the potential to explain MW and validate the idea through simulations by replicating previous findings. We employed a homeostatic reinforcement learning model where independent drives for the task and others were assigned, and drive reduction became a reward and trained under sustained attention to the response task. To demonstrate that HRL agents can replicate key findings on MW, we had them perform a task widely used in MW research. We then analyzed their response tendencies and response times for validation. We confirmed that HRL agents behave consistently with the empirical results reported in human experiments, which suggest that MW could be under homeostatic control. Finally, we discuss the behavioral and neurobiological commonality between human thought and animal behavior and the possibility that the same principle, homeostasis, controls these phenomena.

The thoughts of human beings are highly dynamic, and we often think about what is disengaging from the current situation, such as events that occurred in the past, which might happen in the future or never happen. These thoughts shift from the primary task to unrelated thoughts—what is termed mind-wandering (MW)[1]. Since MW is an ordinary phenomenon in which people spend 30–50% of their waking time, it has important roles in human cognitive function[2,3].

Various factors encourage or discourage us from engaging in MW. The first is the participants' motivation toward the main task and task-unrelated thoughts (TUTs). As the task proceeds, participants' motivation declines, leading to more MW and poorer task performance[4,5]. Motivation for internal thought also influences MW. The content of TUT often directs to the future, which suggests that individual worries and concerns drive MW[6–8]. Task difficulty has a non-monotonic effect on MW. When the task is easy, it allows participants to engage in TUT, and as the difficulty increases, engagement in TUT decreases[9,10]. Conversely, when the task is too difficult, participants engage in TUT more[11–13]. Although various factors are involved in MW, and several hypotheses have been proposed to explain them, a comprehensive framework that unifies these findings is still absent.

TUTs or MW might not be unique to humans but rather a continuum of animal behavior. TUT occurs particularly frequently in situations where behavior is severely restricted, such as during classes, work, or laboratory experiments[14]. In environments with no such strong constraint on behavior, the state of boredom is reduced by observable changes in behavior[15–17]. These facts suggest that TUT may play an alternative role in behavior. Like humans engaged in TUTs in an experimental situation, animals also display task-unrelated behaviors, even when they engage in an experimental task[18–24]. Even mice, when their behavior is constrained by head-fixation, show a state characterized by immobility and a decrease in reactivity to external stimuli instead of engaging divergent behaviors[25,26]. In addition to the superficial similarity in MW, the shift of behavior in humans and animals might share a common function, such as escaping from a state of boredom, exploring the environment, and promoting learning[27]. These behavioral and functional similarities imply a possible common underpinning behind MW and behavior of organisms. Physiological requirements, such as hunger and thirst, are a primordial source that drives animals to engage in specific behavior. The idea of maintaining a steady physiological state, or homeostasis, has been applied to associative learning, habituation and sensitization, social behavior, and various behavioral phenomena[28–32], suggesting that it has potential to be applicable to human thoughts.

[1]Keio University Global Research Institute, Tokyo, Japan. [2]Department of Information Medicine, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan. [3]Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan. ✉email: kazushi_shinagawa@keio.jp

Homeostatic reinforcement learning incorporates the idea of homeostasis in how agents make decisions to keep their physiological state stable[30]. This model defines reward as the proximity to a desired value for the internal state (i.e., setpoint). Therefore, the reward changes dynamically according to the agent's physiological state. For example, when an animal is hungry, food reduces its hunger and reinforces its behaviors. However, if the animal is full, excessive foods make the animal nausea and punish its behavior. HRL is a model learning the policy to prevent the physiological state from diverging by modeling changes in the reward according to the setpoint. Key parameters of this model in this study are setpoint and time decay. Since setpoints reflect the desired value for the internal state, higher setpoints lead to a larger deviation between the homeostatic state and setpoints. Time decay specifies the natural decrease of homeostasis state (body temperature, satiety), which reflects the drive's recovery speed. Then, these parameters could directly control motivation for behavioral options. To show that engagement in the task and MW can be explained by homeostatic control in our simulations, we assumed that the homeostatic state is the amount of engagement in each behavior rather than the actual physiological state. Therefore, the setpoint is interpreted as the ideal amount of engagement in each behavior.

We compare our model with previous research to evaluate the consistency and novelty. Most previous computational models of MW used Adaptive Control of Thought-Rational (ACT-R). These studies set "task engagement" and "MW" as a chunk of the goal and stochastically choose which is selected at each time point[33,34]. By assuming that the representation of task engagement declines over time, ACT-R replicates the natural occurrence of MW during the task. These models often assume no explicit cognitive process as they set up engagement in MW and return to task engagement as stochastic[33,34]. While the model is highly extensible, it does not provide a cognitive account for the natural cycle of attentional states during the task, such as the occurrence of MW and return of task engagement. A more explanatory model, which extends the ACT-R, calculates the activation of representations for "task engagement" and "MW" and represents that our mental representations are in constant competition[35]. The competition of representations is explicitly assumed, and the model can deal with task engagement and MW comparisons. It focuses on explaining MW with a particular emphasis on underload, which is the hypothesis that surplus resources are allocated to MW. Therefore, the hypothesis that MW occurs from the aspect of failure to maintain representations may not be handled well, such that MW is more likely to occur when the task difficulty is extremely high[9–13].

The study in the context of meditation implements the dynamics of attention during tasks by explicitly introducing a tendency toward MW[36]. The state of focus on the task at a given time point is defined as the difference between the state before the time point and the bias toward the MW, expressing the tendency toward the MW. The dynamics of attention are implemented by continuously calculating the difference between the target value and current state as a function to revive attention when the difference becomes large. Although the model arises from the idea that MW occurs when there is more benefit than task focus[37], no clear comparison is made between focus and MW. Furthermore, when active engagement in MW occurs, it does not persist, and when a deep MW occurs, it is set to end immediately owing to the effect of a function that revives attention. One problem with both studies is that, although MW is such a common phenomenon that it occupies much of people's daily time, the model is not constructed in such a way as to account for its active engagement. We aim to create a model that implements the natural stream of attention during the task and can also reproduce active MW and the nonlinear relationship between task difficulty and MW frequency.

In this study, we replicated earlier findings by applying sustained attention to response task (SART), widely used in MW research, to HRL agents. In Simulation 1, we showed that MW occurs in HRL agents during the task and analyzed the model's internal variables to identify the mechanism underlying MW occurrence in the model. Subsequently, we qualitatively replicated results of existing studies to validate the use of HRL as a comprehensive framework for MW. In Simulation 2, we manipulated the parameter that determined motivation for the main task and replicated that changes in motivation affect the proportion of MW during the task. In Simulation 3, we replicated the results by which highly motivating events drive TUT by manipulating the motivation for TUT. In Simulation 4, we manipulated task difficulty and replicated the findings by which MW decreases as the task becomes more difficult but increases when the difficulty reaches an extreme level. Additionally, these results imply that human-specific phenomena, like thought dynamics, are also governed by biological principles across physiology and behavior.

Results

In the present study, we attempted to replicate the results reported in the MW study by conducting SART on agents implementing HRL[30]. In SART, fixation and numbers are presented alternately, and participants are required to respond only when a specific number is presented (Fig. 1A). Response time to the target stimulus and its variance is increased during MW[38–45]. We simplified this task to include only "fixation" and "stimulus" presentations for applying this task to agents (Fig. 1B). HRL agents could choose an action at each time step from three options: "response", "focus", and "TUT". "Response" indicates a response to the stimulus; "focus" indicates that the agent is engaged in the task but does not respond, and "TUT" indicates that the agent is engaged in the TUT. In this simulation, we defined MW as transitioning from a task-focused state to TUT and maintaining it during the task. We constrained the agents' choice depending on the environment state and previous action so that agents would not choose "response" when engaging in "TUT" or that the environment state was "fixation" (Fig. 1C; see "Materials and methods" section for detail). In the simulations, one trial involved 40 timesteps comprising 30 "fixation" steps followed by 10 "stimulus" steps, and each agent conducted 200 trials. The homeostatic state of "TUT" is equal to the setpoint from the onset of the task, meaning that the reward for "TUT" is more likely to become negative during the task. However, since reward depends on all changes in homeostasis, if a particular homeostasis state deviates from its setpoint owing to a certain response, but other homeostasis states approach those setpoints, the reward will be positive. Agents choose an action at each time step based on the past rewards and actions and calculate the reward and reward prediction error (RPE)

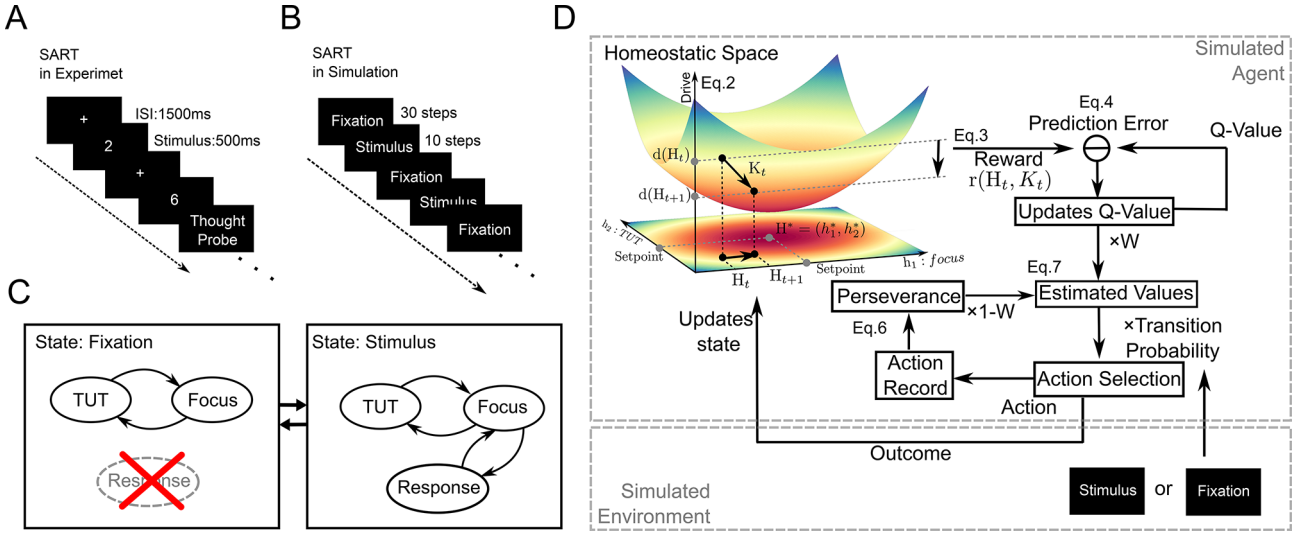

Fig. 1. The task and agent used in the simulation. (**A**) Schematic image of a typical SART recruited in an empirical MW study. (**B**) The image of SART in the simulation. The duration of the "stimulus" presentation was set to be similar to the actual task conditions, with each step set to be 50 ms. (**C**) The actions that would be selected under each task state in the simulation were as follows: During "fixation", the agents cannot respond and must either focus on the task and wait or be immersed in TUT. The agents can respond during "stimulus" presentation only if focused on the task. (**D**) Schematic image of HRL with a two-dimensional homeostasis space: When agents choose an action, agents receive an observation from the environment. The homeostatic state changes based on the observation, and the deviation from the setpoints is calculated; this proximity is defined as a reward. The model calculates RPE by comparing the reward and expected reward obtained by the chosen action, Q-value. RPE signal then updates the Q-value for the action. See Materials and Methods for details.

based on the observations and their homeostatic state (Fig. 1D; see "Materials and methods" section for details). The following four simulations share the same environment, model architecture, interaction between agent and environment, and flow of the simulation.

Simulation 1: exploration of the process by which MW occurs

The goal of simulation 1 is to show that the occurrence of MW depends on two independent processes: (1) a choice between "focus" and "TUT", and (2) independent dynamics of the drive behind MW. To achieve this goal, we simulated and analyzed the behavior of three models: a full model incorporating both processes, a no-TUT model, and a vanilla Q-learning model (VQ). The no-TUT and VQ models lacked either of two processes from the full model (see Materials and Methods for details). Comparing the proportion of each action, "TUT" was rarely chosen other than the full model (Fig. 2A). Response times in trials when agents engaged in "TUT" immediately before the "stimulus" onsets were prolonged in the full model and VQ compared to in trials, when agents engaged in "focus" (Fig. 2B). Furthermore, as indicated in many recent studies[40,42,43], the variance of response time was larger during "TUT" than the "focus" only in the full model (Fig. 2C). Additionally, the VQ's response times were scattered around 50 ms when the agents selected "TUT" upon presenting a "stimulus" (Fig. 2B). Considering that each time step is 50 ms and that the agents cannot choose "response" at the "fixation" (Fig. 1C), the agents chose "response" at 50 ms as the fastest way from "TUT". In the VQ agents, the probability of "TUT" engagement rapidly decreased from the start of the session and rarely occurred from the session's middle phase (Fig. 2D). These results suggest that two independent processes—(1) a choice between "focus" and "TUT", and (2) independent dynamics of the drive behind them—are necessary to replicate the pattern of response times and shifts in thought observed in typical SART. All detailed statistical results are provided in the Supplementary Materials.

We analyzed the model's internal variables, such as homeostatic states, q-values, and RPEs, during the task to elucidate the mechanism by which MW occurs only in the full model. In HRL, rewards and RPEs for "focus" and "TUT" varied throughout the session, while they remained constant except during the initial period in VQ (Fig. 2F, G). RPEs systematically changed depending on the homeostatic state of each action and the other (Fig. 2E, H). Specifically, the RPEs for the "focus" increased with the homeostatic state of "focus" decreased, and that for the "TUT" increased (Fig. 2H top). This relationship was reversed for "TUT" (Fig. 2H bottom). Furthermore, we analyzed changes in RPEs through successive engagement for a specific action. RPEs were initially positive at the onset of engagement for "focus" and "TUT", but they decreased as successive engagement in each action and turned to be negative (Fig. 2I). Successive engagement in a specific action led to saturation for the action and punished it. In contrast, unsaturated actions reduced the saturation of different actions and acquired a rewarding effect. In trials in which the agents were engaged in the "TUT" upon "stimulus" onset, the saturation in the homeostatic state of "focus" during the "fixation" period caused the reward and RPE to

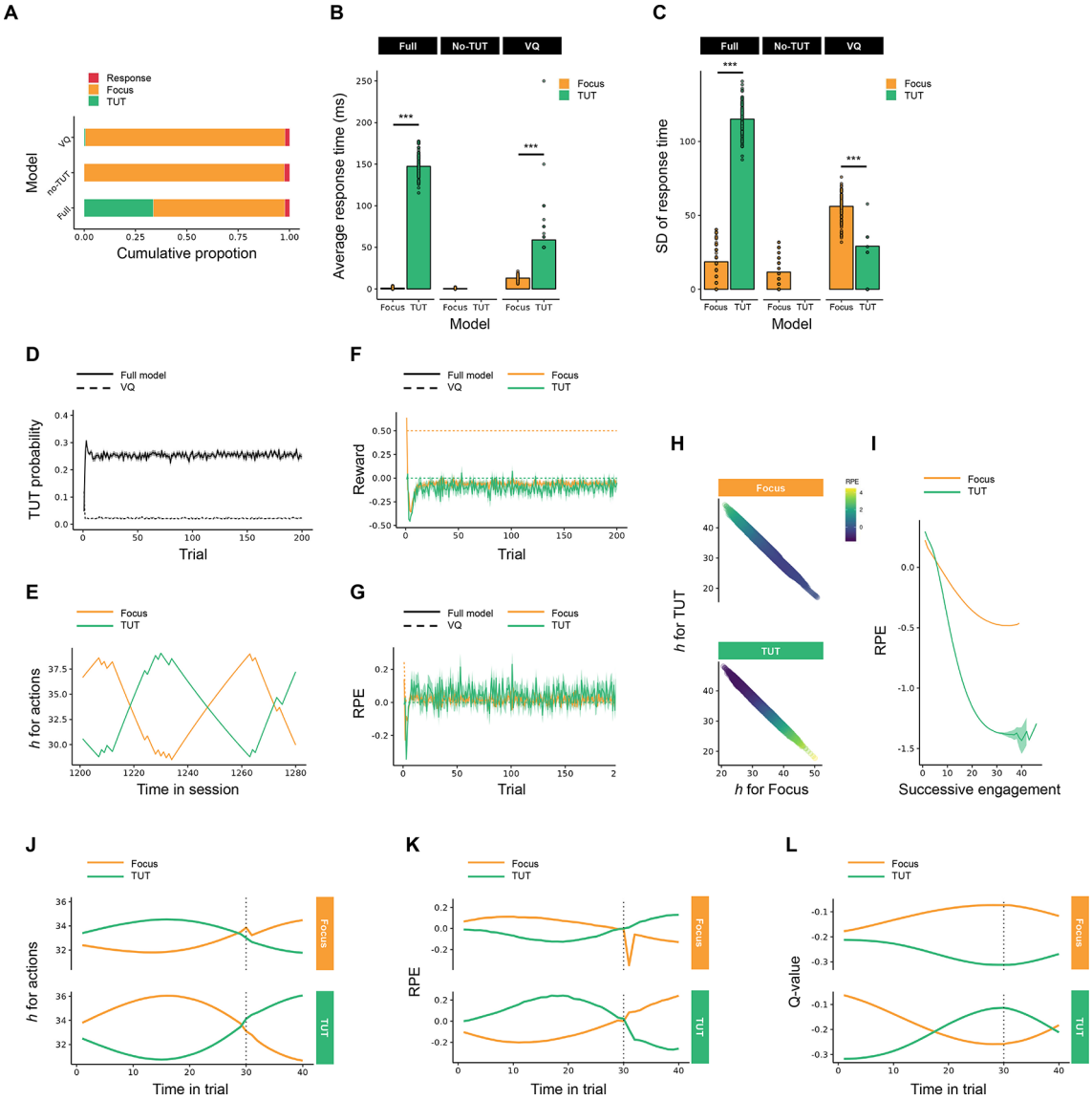

Fig. 2. Results of simulation 1. (**A**) Proportion of three actions, "response" (red), "Focus" (orange), and "TUT" (green), simulated by three models, the full model (bottom), no-TUT model (middle), and Vanilla Q-learning (top). (**B**) Bar plots show response time from "stimulus" onset, averaged across agents, for each response immediately before "stimulus" onset. Because agents could not take "response" in the "fixation" period, we showed results that agents took "focus" or "TUT" immediately before the "stimulus" onset. Individual panels show the results from each of the three models (from left to right: full model, no-TUT, VQ). Scattered dots above and below the bars indicate results from individual agents. (**C**) Bar plots show standard deviation of response time immediately before "stimulus" onset among agents for each three models. (**D**) TUT probability, averaged across agents, at each trial throughout the session for the full model (solid line) and VQ (dashed line). The grayed shades indicate standard error across agents. (**E**) An example of the dynamics of the homeostatic state, *h*, for "focus" (orange) and "TUT" (green) in the full model. The data from 1,200 to 1,280 timesteps for one agent are shown. (**F**) Trial-by-trial changes in rewards obtained by "focus" (orange) and "TUT" (green) simulated by the full model (solid line) and VQ (dashed line). (**G**) Trial-by-trial changes in RPEs occurred by "focus" (orange) and "TUT" (green) simulated by the full model (solid line) and VQ (dashed line). (**H**) The relationship between homeostatic states, *h*, for "focus" and "TUT" and RPEs generated by each two actions. The scatter plots show the RPEs resulting from the chosen action in the homeostatic state of the "focus" and "TUT" at a given time step. The upper and bottom panels show the RPEs generated by "focus" and "TUT", respectively. As the point color becomes brighter, the magnitude of RPE increases. (**I**) The lines show that changes in RPEs are caused by successive engagement in the same action. We calculated the number of times "focus" or "TUT" is chosen and averaged RPEs for each time across agents. (**J–L**) Within-trial dynamics of homeostatic state, *h* (**J**), RPEs (**K**), and q-values (**L**), for "focus" (orange) and "TUT" (green) averaged over agents and trials. The upper and bottom panels show the data from the trial where agents choose "focus" and "TUT" immediately before stimulus onset, respectively. Asterisks indicate significance levels: $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

turn negative, which reduced the q-value (Fig. 2J–L). The saturation of the homeostatic state of the "focus" was alleviated by engaging in "TUT", which turned the RPEs of "TUT" positive and increased the q-value. These results indicated that MW was caused by punishment due to saturation of "focus" and rewarding "TUT" due to reduced saturation, and these relationships were reversed in the "TUT" trials. In summary, MW continuously occurred throughout the session in HRL as "focus" and "TUT" alternated to reduce the divergence of homeostatic states for each behavior (Fig. 2E).

In simulation 1, comparing the HRL including "TUT" as an option of action with two models that omit specific processes revealed that the increased response time and its variance and the proportion of "TUT" engagement observed in the SART stem from the choice between task-related actions and "TUT" engagement and distinct dynamics of the drive behind them. Examining the model's internal variables revealed that MW in our model is driven by two processes: the homeostatic state becomes saturated owing to successive engagement in "focus", and this saturation is alleviated through engagement in "TUT". The same process also mediates the change in choice from "focus" to engaging in "TUT", indicating that HRL replicated the cycle of attention during SART.

Simulation 2–4: qualitative replication of MW study using HRL

To demonstrate that HRL could unify previous findings and provide a comprehensive explanation for MW, we performed the following four simulations in which we manipulated the parameters of HRL to qualitatively replicate the results reported in empirical studies. In Simulations 2–1 and 2–2, we replicated the results that subjective motivation to the task affects the behavioral features, such as the "TUT" proportion, response time, its variance, "focus" and "TUT" sustainabilities, and their occurrence frequency by manipulating the setpoint (2–1) and the time decay of the homeostatic state (2–2) for the task-related actions. In Simulation 3, we replicated the hypothesis that current concerns trigger MW[6–8]. This was achieved by adjusting the setpoint for "TUT" to reflect varying interest in the MW content. In Simulation 4, we replicated the finding that MW was less likely to occur when task difficulty was high and more likely to occur when task difficulty was extremely high[9–13]. Assuming that the reward presentation probability corresponds to task difficulty, we examined the effect of manipulating the difficulty on the MW in the HRL model. Except for the manipulated parameters, the experimental environment and model parameters were the same as those in Simulation 1.

Simulation 2: TUT decreased as motivation for task increased

In Simulation 2–1, to replicate the findings that MW is more likely to occur with decreasing motivation for the task, we simulated the behavior when the setpoints for the task-related actions were manipulated. First, we revealed that the HRL agents also engaged more "TUT", with decreasing motivations for the task-related actions (Fig. 3A). While the response time and variance were not changed with motivation level when the agents engaged in "focus" upon the "stimulus" presentation, they increased when the agents engaged in "TUT" (Fig. 3B, C). When actions occur as clusters, called bouts, behavior has different aspects, such as interbout interval and number of actions involved in each bout. To clarify how setpoint manipulation influenced which aspects of behavior, we analyzed the microstructure of behavior. Since we generated data on action choices at each time point, we could analyze the dynamics of the drive behind the behavioral outputs and microstructure from the time series of behavioral choices. We counted the number of repetitions from the onset of action to the end and revealed that motivations for task-related actions did not affect the average successive engagement for "focus" but decreased "TUT" (Fig. 3D). Conversely, the average interbout interval (interval between the end of action to the next onset), decreased in "focus" but increased in "TUT" with increasing motivation levels (Fig. 3E). Thus, the higher setpoint shortened the successive engagement for "TUT" and made response time shorter. To clarify how setpoint manipulation influenced the successive engagement of "TUT", we conducted further analysis of the agents' internal variables. When the setpoint was higher, the homeostatic state converged at a higher level, leading to enhanced attenuation (Fig. 3F, G) owing to its dependence on the current homeostatic state and time decay (Eq. 1). This relevancy between setpoint and time decay suggests that recovery from saturation was faster in the higher setpoint condition. When the homeostatic state for "focus" was saturated, agents shifted to engage in "TUT" until the state of "focus" recovered from saturation. Thus, when the setpoint was higher, the homeostatic state for "focus" could recover from saturation faster, resulting in the short successive engagement in "TUT".

In Simulation 2–2, the effect of task motivation on the proportion of MW during tasks was also replicated by manipulating the time decay parameters. Similar to the results in Simulation 2–1, as time decay decreased (i.e., decreasing to maintain motivation for the task-related action), the agents engaged more "TUT" (Fig. 3H). Moreover, the average response time and variance upon "stimulus" presentation increased with the higher motivation level (Fig. 3I), replicating the relationship between changes in motivation for the task, the proportion of MW, and task performance. The microstructures of behavior, successive engagements, and interbout interval indicated the same trends as the results of simulation 2–1 (Fig. 3J, K). Although we manipulated different model parameters, setpoints, and time decay, they influenced agents' behaviors in the same way, which suggests that these two manipulations were functionally equivalent. In summary, we successfully replicated previous experimental findings that the motivation for the task affects the number of MW occurrences and response times and their variance through the manipulation of setpoints and time decay.

Simulation 3: TUT increased as motivation for TUT increased

In Simulation 3, we simulated the behavior when the setpoint of "TUT" was manipulated and replicated the influence of motivation for TUT on the proportion of MW during tasks. Figure 4A shows the proportion of action choices for each condition and reveals that HRL agents engaged more "TUT" with increased motivation for the "TUT". While the response time and variance when agents engaged "TUT" upon "stimulus" presentation

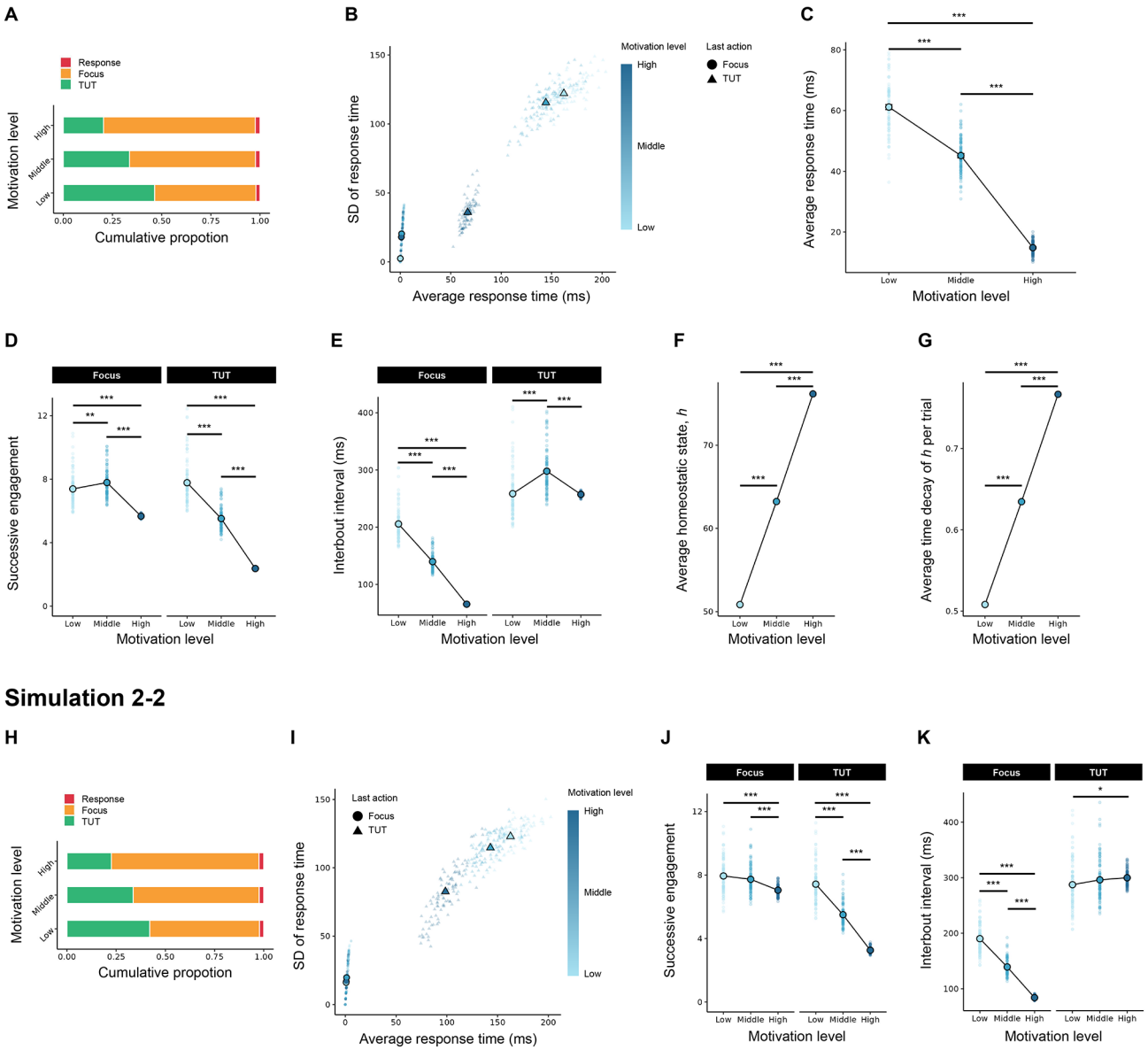

Fig. 3. Results of simulation 2–1 and 2–2. We presented the results when the motivation for the task was manipulated in simulations 2–1 (**A–G**) and 2–2 (**H–K**). (**A**,**H**) Proportion of three actions, "response" (red), "Focus" (orange), and "TUT" (green), when the motivation for the task was manipulated from low (bottom) to high (top). (**B**,**I**) Average response time and variance upon "stimulus" presentation for each motivation level and agent's action. The point shape indicates the selected action types upon "stimulus" presentation; as the point color becomes brighter, the motivation for the task decreases. (**C**) Average response time during the task for each motivation level. (**D**,**J**) Average successive engagement from the action initiation to the end of the action for each motivation level; we counted the number of repetitive choices for each action. (**E**,**K**) Average inter-action interval duration for each motivation level. (**F**) Average homeostatic state during the task for each motivation level. (**G**) Average time decay during the task for each motivation level. For (**B–G**,**I–K**) the black-edge points indicate the averaged value, and transparent points are individual data averaged across trials. Asterisks indicate significance levels: $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

increased with higher motivation in simulation 2, we did not observe the effect in simulation 3 (Fig. 4B). However, the average response time during the task prolonged as the setpoint increased (Fig. 4C). We conducted further analysis of behavioral microstructures, which revealed that setpoint manipulation did not influence successive engagement of actions. In contrast, the interbout interval was decreased as the setpoint for "TUT" became higher (Fig. 4D, E). While the proportion of "TUT" in the task varied across the motivation levels (Fig. 4A), the response time was not prolonged when the agent engaged in "TUT" upon "stimulus" presentation because the motivation level had no effects on successive engagement of "TUT" (Fig. 4D). In Simulation 2–1, the setpoints of

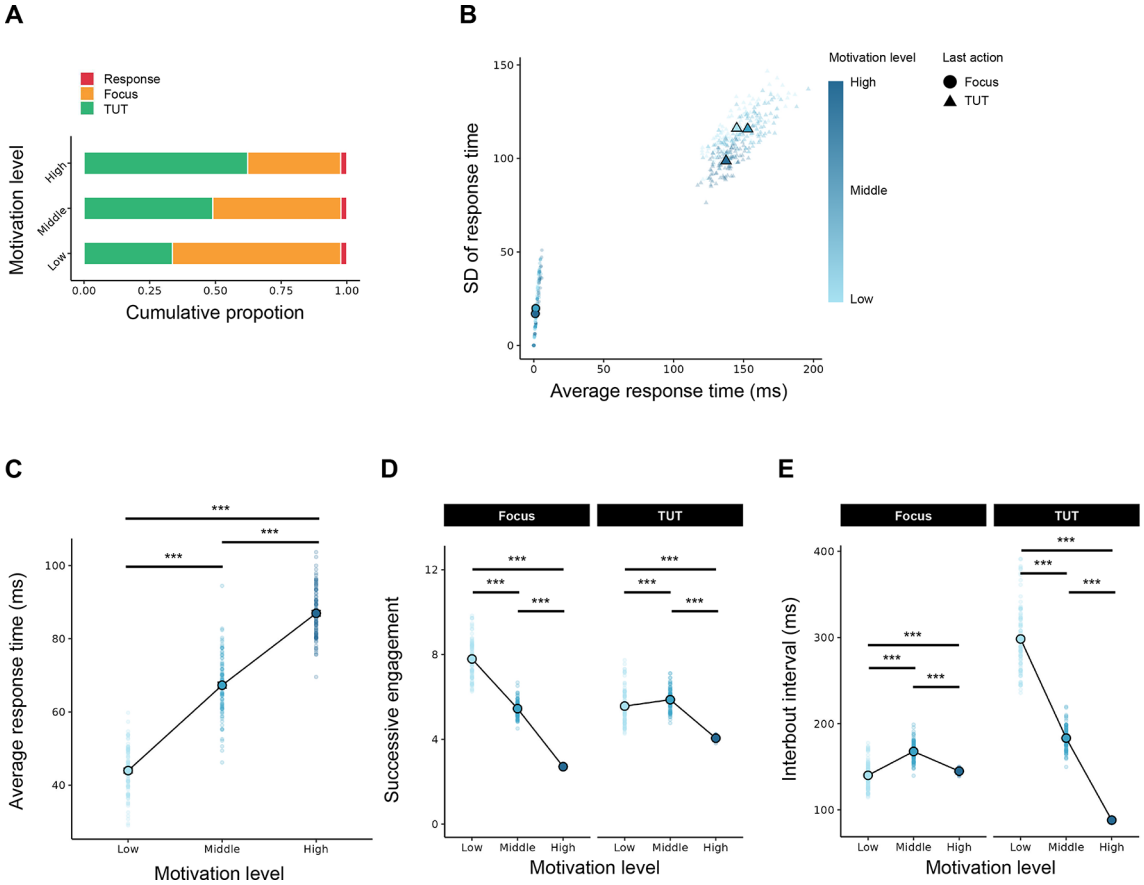

Fig. 4. Results of simulation 3. (**A**) Proportion of three actions, "response" (red), "Focus" (orange), and "TUT" (green), when the motivation for the "TUT" was manipulated from low (bottom) to high (top). (**B**) Average response time and variance upon "stimulus" presentation for each motivation level and agent's action. The shape indicates the selected action types upon "stimulus" presentation. As the point color becomes brighter, the motivation for the TUT decreases. (**C**) Average response time during the task for each motivation level. (**D**) Average sustained duration when action selection is initiated for each motivation level. (**E**) Average inter-action interval duration for each motivation level. For (**B**–**E**), the black-edge points indicate the averaged value for each motivation level, and transparent points are individual data averaged across trials. Asterisks indicate significance levels: $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

"focus" were always set as higher than the TUT, while this simulation included cases where the "TUT" was below or above the "focus", which suggests that the relative values of the setpoints affect the probability of occurrence for the other behavior. Taken together, we successfully reproduced the effect of MW motivation on behavioral features under the SART.

Simulation 4: non-monotonic effect of task difficulty on the TUT

In Simulation 4, to examine the effect of task difficulty on MW, we simulated the behavior of HRL agents by decreasing the probability of reward for "focus" based on the conditions of Simulation 1 (see "Materials and methods" section for details). While the proportion of "TUT" decreased as task difficulty increased, conversely, the proportion increased when the task difficulty was extremely high (Fig. 5A). As the task difficulty increased, the response time and variance when the agents engaged in "TUT" upon the "stimulus" presentation and average response time during the task decreased with the task difficulty. In contrast, the average response time and its variance were larger than the high level in the extremely high level (Fig. 5B, C). We further analyzed the behavior's microstructure, and the successive engagement and interbout interval for each action did not change monotonically (Fig. 5D, E). As shown in Fig. 5A–E, the model displayed non-monotonic changes in behavior for task difficulty, and we conducted further analysis on the model's internal variables to clarify the underlying mechanisms. When the task was easy, agents could obtain rewards at high probabilities, and this caused saturation in the homeostatic state for "focus" (Fig. 5F). If the homeostatic state for "focus" was saturated, engagement in "TUT" reduced the homeostatic state for "focus" (Fig. 5F), and it produced positive RPEs for "TUT" (Fig. 5G). In contrast, when the task was difficult, the homeostatic state for "focus" was not saturated. In that case, engagement in "TUT" led to deviation from the setpoint for "focus" (Fig. 5F), and it caused negative RPEs for "TUT" (Fig. 5G). However, the task was extremely difficult; agents could not obtain a reward at all, and the homeostatic state for "focus" remained constant regardless of agents' action (Fig. 5F). Thus, the "focus"

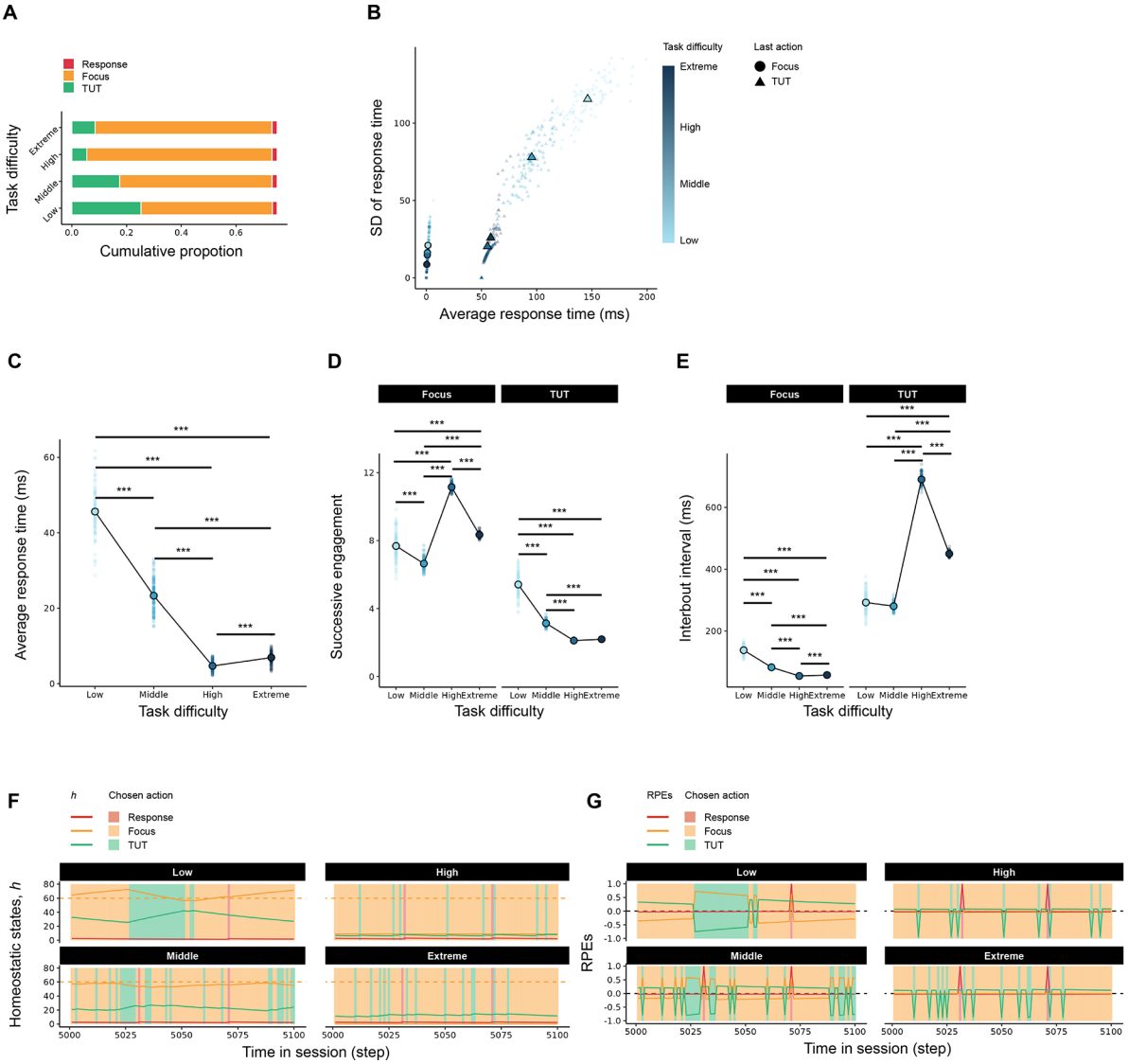

Fig. 5. Results of Simulation 4. (**A**) Proportion of three actions, "response" (red), "Focus" (orange), and "TUT" (green), when the motivation for the "TUT" was manipulated from low (bottom) to Extreme (top). (**B**) Average response time and variance upon "stimulus" presentation for each task difficulty and agent's action. The shape indicates the selected action types upon "stimulus" presentation. As the point color becomes brighter, the task difficulty decreases. (**C**) Average reaction time during the task for each task difficulty. (**D**) Average sustained duration when action selection was initiated for each task difficulty. (**E**) Average inter-action interval duration for each task difficulty. (**F**, **G**) Time series of homeostatic state and RPEs for each action in a specific agent across some trials. We plotted these values in solid lines and highlighted the interval depending on the actions chosen by agents. For (**B–E**), the black-edge points indicate the averaged value for each motivation level, and transparent points are individual data averaged across trials. Asterisks indicate significance levels: $^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

dimension produced no RPEs for "TUT" (Fig. 5G). The homeostatic state for "focus" converged different asymptotes across task difficulties, and they produced different magnitudes of RPEs from positive to negative. As a result, the task difficulty influenced MW in a non-monotonic way. These results indicate that Simulation 4 succeeded in replicating the effect of task difficulty on MW.

General discussion

The present study aimed to comprehensively explain the previous results related to MW occurrence using HRL through the simulations and provide evidence that action selection through homeostatic control is behind the dynamics of thoughts. We imposed the SART on HRL agents and examined response time, behavioral microstructures, and the internal variables as a function of parameter manipulations. The results showed that MW in HRL agents occurred when the homeostatic state for the task-focused was saturated, and the relative action value function of TUT increased (Fig. 2). Furthermore, we successfully replicated the effects of

motivation for the task and MW and of task difficulty on the proportion of MW reported in previous studies by manipulating the relevant parameters (Figs. 3, 4 and 5; Supplementary Table 1). In summary, HRL could provide a comprehensive explanation for MW, suggesting that human thought transitions follow the homeostatic control of action selection, that is, choices between the task and TUT and independent drives dynamics behind them.

In addition to the experimental results of MW validated through the simulations, the traits of MW observed in the HRL agents are consistent with many findings in previous studies. At the onset of MW in HRL, the reward for focusing on the main task is negative, and the TUT is driven to resolve this saturated homeostatic state. This functional aspect of MW is also consistent with the hypothesis that MW has a distracting effect that may alleviate the current boredom[46]. Temporal patterns in the proportion of MW within a session showed a low occurrence rate initially, which increased as the task continued, and similar results have been reported in experimental settings[4,5]. The decrease in the proportion of MW among older adults can be explained by fewer cues related to current concerns in the experimental setting[47], consistent with the findings when we manipulated motivation for TUT in the present study. Although not explicitly addressed in this study, HRL can also explain various aspects of MW.

Characteristics of thought processing, such as intentionality or automaticity, are considered important in MW research[2]. Although the model presented in this study does not explicitly address these aspects owing to the agent's inherent lack of subjectivity, it provides some valuable insights. Two types of MW emerged in this model: those intended to alleviate excessive engagement to the main task (Simulation 1) and those intended to actively engage in MW driven to fill in the gap between the current homeostatic state and setpoint (Simulation 3). Intentional MW tends to relate the contents to the future[48]. We assumed the active MW was driven by current concerns in Simulation 3; thus, active engagement in MW may be associated with intentional thought. Additionally, unintentional MW occurred more frequently over task time than intentional MW[49]. These results indicated that the unintentional MW may be consistent with the characteristics of MW, which is intended to alleviate excessive engagement in the main task and support the correspondence between intentionality and the HRL model's MW types.

The microstructure analysis revealed parameter manipulations that influence each aspect of MW, which enabled us to compare the results of our simulations and previous findings in detail. Higher motivation for the task decreases MW by mediating the effort to return attention to the task more quickly[50], which suggests that the duration of each TUT is shortened. When we manipulated motivation for the task in simulation 2, the proportion of MW decreased as task motivation increased, and it was mediated by shorter successive engagement in "TUT" (Fig. 2E and J); this suggests that HRL could capture the underlying mechanism behind thought transition. This is not limited to MW, but animal behavior showed a similar relationship between motivation and behavioral microstructure. Instrumental behaviors of animals show bout-and-pause patterns that are characterized by bursts of instrumental responses and longer pauses separating each burst[51,52]. Motivation, such as hunger, controlled interbout intervals and bout length[52–54]. These results suggest that MW have a similar microstructure to those observed in observable animal behavior, which supports the idea that thought phenomena share a common principle with the general behavior of organisms.

One of the concerns with employing HRL to provide an account for MW is that the model handles an observable action ("response") and unobservable ones ("focus" and "TUT") in the same line. However, this formulation is validated by the following three points. (1) Many MW studies recruited probe-caught methods, which ask participants to choose whether they have focused on the task or MW randomly during a task[55]. This method is similar to that adopted in the present study, where MW is treated along the same lines as task engagement as participants' thought options. (2) The fact that MW prolonged the response time in our simulations and empirical studies indicates that the TUT competes with observable response or task-focused states even in the behavioral output. (3) The hypothesis that MW uses the same cognitive resources as task engagement[8] is supported by many psychological and neuroscientific findings[56,57]. These findings and assumptions indicate that previous studies implicitly or explicitly treated internal thoughts on the same level as other observable behaviors, especially in assessing whether the agents were engaged in the task.

Another concern is to set a homeostatic state or setpoint for task-related actions and TUT. In the present study, the homeostatic state of behavior does not assume a corresponding physiological state, as with conventional HRL. However, several perspectives in prior research support the idea of treating the amount of engagement in a particular behavior as the homeostatic state without corresponding to the physiological state and that action choices minimize deviations of this state from setpoint. The motivation hypothesis, which explains the occurrence of MW, posits that psychological tasks are simple and task engagement restricts other behaviors, resulting in lost opportunities and an increase in the action value of alternative behaviors, the accumulation of which leads to attentional shift[4]. This perspective implicitly assumes that task engagement depends on the comparisons between the value of TUT and task engagement calculated from the amount of engagement and the threshold. Thought transitions in the neuroscientific perspective are explained by the fact that continued activity in the locus coeruleus, a brain region that contributes to sustained attention, leads to hyperactivity and a release of the task-focused state beyond the optimal activity point[58]. Although the process after the release of focus is unknown, a threshold of activity is assumed for the locus coeruleus, which would lead to the termination of a specific thought. Extending homeostasis beyond physiological states, such that engagement in a particular behavior or thought has some threshold for engagement, does not deviate from the assumptions and findings of previous studies. Indeed, some studies have postulated homeostasis in social interactions and habituation[28,31].

The explanation of MW using HRL also includes a perspective on interactions between brain regions among the thought transitions frequently discussed in the context of MW research. Three brain regions are thought to be related to task engagement and MW: (1) the central executive network (CEN), which is active during task engagement; (2) the default mode network (DMN), which is associated with internal thought; and (3) the salience network (SN) switching the transition between the two networks[2,59,60]. The transition from MW to task-focused

state is explained by the SN detecting salient stimulus in the internal and external environment and switching to CEN. However, the transition from the focused state to MW (i.e., switching when the DMN is activated) has yet to be fully explained[61–63]. The insular cortex, which constitutes the SN, is known to be involved with emotional and physical responses[61]. This region is also activated immediately before action selection changes, especially in environments with little change[64,65]. Our simulations revealed that the transitions from task engagement to TUT and from TUT to task engagement are driven by saturation of the currently engaged action (Fig. 2). If the SN tracks the internal state of each dimension of action, HRL can provide a comprehensive account of bidirectional thought transitions, such as the occurrence of MW (activation of DMN) and engagement in the task (activation of CEN) in the form of the detection of agents' unpleasant emotions and their associated physical responses by the SN. Furthermore, HRL can include the case that salient stimulus detections occur suddenly in the internal and external environment, which conventional brain network models have assumed. In this study, as the task time progressed, the internal states of both "focus" and "TUT" deviated from the setpoint, and the action value function converged to a negative value (Fig. 2). When a salient novel stimulus emerges, the action value function of that stimulus is relatively high compared to other alternatives. Thus, when any internal state dimension is saturated, a salient stimulus can be detected and processed. The degree of task engagement, which is manipulated by fatigue and motivation, affects the degree of suppression of task-irrelevant stimuli, which supports our hypothesis[66,67]. The behavior of HRL is consistent with the action-level shifts between MW and task engagement and the interactions among the relevant brain networks.

We showed that HRL, a model for associative learning of animals, could reproduce MW; this suggests that thoughts unique to humans could be explained from the view of animal behavior. Animals must satisfy several requirements, such as hunger, thirst, danger, and mating in natural environments. The facts lead us to assume a different principle from maximizing a single reward, as expected in a laboratory setting. Evidence for this has long been reported in animal behavior research, such as engaging in diverse behaviors not related to the experimental task[18,20,21,23,24]. Changes in the relative reward effects of behavioral opportunities, such as a less preferred behavior reinforcing a more preferred behavior under specific environmental settings[68], have been reported. Moreover, task-unrelated behaviors may influence task-related behavior[69–75]. Additionally, we showed that behavioral changes caused by motivational operation to HRL were similar to those of animal operant behaviors[52–54]. Although HRL could explain human's MW and animal behaviors, whether they have the same biological underpinnings remains unclear. In fact, theories explain MW other than HRL, and the correspondence with the neural basis is still underexplored. These facts suggest that MW may be driven by a mechanism other than homeostasis control, or the two may be working in parallel. Therefore, the biological fidelity of HRL requires rigorous comparison with existing and future theories and models as well as verification through actual neurobiological research.

Our model can comprehensively address not only the findings of the MW study but also boredom, which is the adjacent research area. Boredom is defined as an affective signal that we have deviated from an optimal zone of cognitive engagement[76]. The nonlinear relationship between task difficulty and MW appears to be explained in terms other than HRL as non-optimal difficulty generates feelings of boredom and task-irrelevant behavior. Since our model is a potentially comprehensive model of boredom, these ideas are likely to be natural occurrences. That is, deviations from the optimal state are thought to reflect feelings of boredom and associated physical reactions, which in turn promote subsequent behavioral transitions. These relationships are supported by behavior and neurological findings. Boredom-like behavior in which animals and humans tend to accept aversive stimulus, such as electric shocks or air puffs, in a poor environment, and its neural substrates support our idea[65,77]. When the environment is poor, agent homeostatic states for available actions should satiate, and aversive stimulus should become a reward by reducing satiation. In such a situation, the insular cortex, a part of SN, was involved in the boredom-like behavior, and mesolimbic dopaminergic neurons were excited around the aversive stimulus presentations[65]. We showed that "TUT" reduced the saturation of the homeostatic state for "focus", and it works as a reward for "TUT" (Fig. 2K, L). The process that drives "TUT" is similar to the mechanism of boredom-like behavior of mice. This is not limited to conceptual similarity, but neural evidence that the insular cortex is involved in boredom-like behavior also supports our idea. Thus, the present study comprehensively addresses boredom and MW in the form of increased negative reward prediction error and the behavioral transitions driven by it. Additionally, our results suggest that animal behaviors and the dynamics of thought—a phenomenon unique to humans—may have a common basis in homeostatic behavior control.

We simulated HRL behavior under the SART where participants cannot behave freely owing to laboratory settings. Thus, the HRL agent has only three alternatives as their choice: "focus", "TUT", and "response". When the participants became satiated or bored with the experiment, they could not be distracted other than with the TUT in laboratory settings; thus, the participants and HRL agent engaged in the TUT. However, humans can have more alternatives in non-laboratory settings, and they distract such state by their shift of behavior[15]. This fact suggests an interesting relationship between satiation and exploration. Generally, when a particular physiological need is strong, the animal will behave in a way that satisfies that physiological need. However, what happens once that need is satisfied? One possibility is that it may not be linked to reward, but the animal will move on to exploration. In fact, the level of hunger is strongly related to the exploratory nature of animals. When the level of hunger is high, animals do not explore but exploit the choice options that are linked to rewards; however, as hunger is satisfied, they begin to explore[78–80]. If a certain homeostasis state is satiated in HRL, the reward value will turn negative, punishing the behavior associated with that reward and encouraging other behaviors. Therefore, HRL can also be seen as controlling exploratory behavior through dynamic adjustment of the reward value. With this in mind, MW may be considered a form of exploration without movement, in a situation where movement is restricted.

In daily life, we get lost in thoughts of various matters, such as issues or problems we face, our future, and past mistakes. What drives these thoughts to transition from one thing to another? Here we provide a clear answer

to the question. The difference between the setpoint and amount of engagement for options decides the action to be engaged at the next step. Despite the simplicity, we successfully replicated the previous findings in the MW research through our simulations and provided a comprehensive explanation for MW. Although not addressed in this study, it has high applicability to spontaneous thought transition phenomena in general as it can be extended not only to bidirectional transitions between task and MW but also to sudden events as described above and to more segmented thoughts of TUT (i.e., to assume each dimension for the future and past). Furthermore, we revealed the similarity between the process of MW occurrence and animal behavior by analyzing the agents' internal dynamics. Taken together, we could explain human thought transition by extending a model of animal learning and behavior based on homeostatic control, including not only MW and task focus but also boredom-like behaviors and other spontaneous thoughts that occur during tasks. This suggests that human thought shares a more fundamental principle with animal behavior than we had usually assumed.

Methods

Sustained attention to response task in the simulation

In vivo

Many MW studies have used SART as the main task, in which fixation and numbers (e.g., 1–9) are presented alternately. In our study, participants were asked to press a key as quickly as possible when the number was presented; moreover, they were asked not to respond to a target number (e.g., 5; Fig. 2A). As inter-stimulus intervals were set as random, SART required them to hold the sustained attention to the task. At random time points during the task, participants were asked about their attentional state, such as whether they were focused on the task or MW. By measuring the response time, brain activity, and physiological indices before the MW reports, previous research has revealed the traits of MW. Many studies have pointed out the delay of response time and increase in variance during MW; thus, we used these indices as the measure of MW in the present study.

In silico

We used a simplified version of SART because the focus of the present study is describing the relationship between action choice and the internal states underlying it. Agents were presented with the "fixation" or "stimulus" at each time step and required to respond as quickly as possible when the task state changed to "stimulus". In this situation, the agents focused on the task, engaged TUT or responded to the "stimulus" at each time point (Fig. 2B, C).

The inter-stimulus interval (ISI), the presentation time of the "fixation", was 1,500 ms; the "stimulus" presentation time was 500 ms, and each timestep was 50 ms, resulting in 40 steps per trial. In all simulations, 200 trials were conducted—in other words, 2,000 ms x 200 trials = 400 s (about 6.5 min). We ran 100 simulations with the same setting for each condition in all simulation sections, assuming we collected data from 100 participants. Since ISI is 30 steps (1,500/50 ms), the agents could respond when the time step is between 31 and 40 steps. When we presented the results of simulations, time steps were changed to actual time in seconds.

Homeostatic reinforcement learning

HRL aims to minimize the deviation of the internal homeostatic state from the setpoint, assuming homeostasis is maintained throughout the reinforcement learning. The agent's physiological state, or homeostatic state, is formulated $H_{i,t} = (h_{1,t},\ h_{2,t}, \ldots\ h_{N,t})$ and defined as a multidimensional space consisting of time-varying body temperature, blood glucose level, blood pressure, etc. The *i*-th internal state at a time point *t* is denoted by $h_{i,t}$. Each homeostatic state decays exponentially through time according to the following:

$$H_{t+1} = \left(1 - \frac{1}{\tau}\right) H_t \tag{1}$$

τ is a time constant that defines the decay speed in Eq. (1). In this homeostatic space, the drive function $D(H_t)$ is defined as the distance of the homeostatic state from the setpoint h_i^* as follows:

$$D(H_t) = \sqrt[m]{\sum_{i=1}^{N} |h_i^* - h_{i,t}|^n} \tag{2}$$

In Eq. (2), m and n are free parameters that induce nonlinear effects on the mapping between homeostatic deviations and their motivational consequences. Rewards are defined as a reduction in drives when agents get observations from environments, which implies that changes in homeostatic space work as rewards. The model does not receive a reward for correctly responding to the target number in the SART. This design reflects the fact that, in the SART in vivo, no feedback is provided for responses. The reward is defined as follows:

$$r\left(H_t,\ K_t\right) = D\left(H_t\right) - D\left(H_{t+1}\right)\ = D\left(H_t\right) - D\left(H_t + K_t\right) \tag{3}$$

To maximize cumulative reward over time, agents must learn state–action mapping. Organisms in the real world cannot survive by consuming only food but must satisfy multiple physiological needs, such as hunger and thirst. In the same way, maximizing cumulative reward in HRL means minimizing deviation from all setpoints; thus, it is different from maximizing a single reward with a fixed value, as in vanilla reinforcement learning. The Q-learning model was employed to learn action value functions. In this model, the values of action $Q_t\left(a\right)$ is updated based on the reward prediction error.

$$Q_{t+1}\left(a\right) = Q_t\left(a\right) + \ \alpha^{\,Q}\left(r_t - Q_t\left(a\right)\right) \tag{4}$$

In Eq. 4, α is the learning rate for action values and determines how significantly the prediction error is evaluated. The choice decided depends on the probability calculated from the soft-max function.

$$P_t\left(a^k\right) = \frac{exp\left(\beta \cdot Q_{sum_t}\left(a^k\right)\right)}{\sum_j exp\left(\beta \cdot Q_{sum_t}\left(a^j\right)\right)} \tag{5}$$

$P_t\left(a^k\right)$ is the probability of an action to be selected at time t, and β is the inverse temperature, which controls the randomness of choice. The overarching architecture of HRL is outlined above, with several processes added to the conventional HRL model to implement the SART on agents (Fig. 2D).

Homeostatic reinforcement learning for sustained attention to response task

Here, we describe how we applied HRL to the SART. SART involves two different environment states: "fixation" and "stimulus", and available actions differ in the environment states. Agents can choose three actions: "response", "focus", and "TUT". "Response" is the response to the stimulus; "focus" is the agent engaged in the task but not responding to the stimulus, and "TUT" is a task-unrelated thought. "Response" is only available when the environment state is "stimulus", but others are available in both states. To impose constraints on agents' choice depending on the environment state, we set different transition probability matrices for each state and previous action choice and weighted the action values calculated by Eq. 4 (Fig. 2D). Moreover, the agents were required to constantly focus on the task; thus, we modeled using shared action values across states of the environment to ensure that the action value function did not change depending on whether the stimulus was presented. While the behavioral value function did not vary across environmental states, multiple states were prepared to reproduce the situation assumed in the SART (i.e., focusing on the task in "fixation" and response in "stimulus").

We also assumed an additional process to the HRL, that is, the persistence of action. In general, thoughts and actions are not considered to be in immediate and frequent transition, but rather, are maintained through time[43,81,82]. We incorporate the persistence to the HRL by having agents to choose action based on not only action values but previous actions. First, we introduce an action trace, which represents how many times each action has been selected in the past timesteps. The perseverance, denoted as Q_{per_t}, is calculated by the following equation:

$$Q^k_{per_t} = a^k_{t-1} + \alpha\ \left(a^k_t - Q^k_{per_{t-1}}\right) \tag{6}$$

The action, a^k_t, is represented by a binary; if the action k is chosen in a timestep t, then a^k_t is 1; otherwise, it is 0. If agents choose action k repeatedly, Q_{per_t} is increased; otherwise, it decreases. Finally, agents choose their action based on the q-values (Eq. 4) and persistence (Eq. 6) by calculating their weighted sum as follows:

$$Q_{sum_t} = w * Q_t + (1 - w) * Q_{per_t} \tag{7}$$

Parameter settings for each simulation

Simulation 1

We showed the specific parameter settings used in each simulation. Simulation 1 was conducted to reveal that MW occurrence depends on two independent processes: (1) a choice between task focus and TUT, and (2) independent dynamics of the drive behind them. To achieve this goal, we simulated and analyzed the behavior under SART of three models: a full model incorporating both processes, a no-TUT model, and a vanilla Q-learning model (VQ). In the full model where an action choice includes "TUT", the transition probabilities for "focus" and "TUT" were set to be the same during "fixation", and which one is chosen depends on the action value (Table 1). In a no-TUT model, the overall structure was the same as the full model but controlled using transition probabilities to restrict the choice of "TUT". We set the transition probability to select "TUT" as 0 from the task-related actions (Table 2). When the agents selected a "response", we applied the same transition probability as the "fixation" to avoid consecutive responses to prevent agents from choosing "response" many times in one trial. We also examined the behavior of a VQ in SART to show the independent dynamics of the drive behind "focus" and "TUT", which is a necessary process for MW occurrence. In the VQ model, no change occurred in the homeostatic states; thus, rewards were clipped at constant through time. We set the observed reward from the environment to 1 for all models and simulations.

The other parameter settings of the HRL agent used in Simulation 1 are shown in Tables 3 and 4. First, parameters such as learning rate, m, and n were determined by referring to previous studies, while the inverse

Environment state		Response	Focus	TUT
Fixation	Response	0	0.5	0.5
	Focus	0	0.5	0.5
	TUT	0	0.5	0.5
Stimulus	Response	0	0.5	0.5
	Focus	0.5	0.25	0.25
	TUT	0	0.5	0.5

Table 1. Transition probabilities when task-unrelated thoughts are included in the action options.

Environment state		Response	Focus	TUT
Fixation	Response	0	1	0
	Focus	0	1	0
Stimulus	Response	0	1	0
	Focus	0.5	0.5	0

Table 2. Transition probabilities in environments where task-unrelated thought cannot be selected.

	α	β	m	*n*	w
values	0.05	10	3	4	0.95

Table 3. Parameter settings related to learning of HRL agents.

	Response	Focus	TUT
tau	1/100	1/100	1/100
setpoint	30	30	0

Table 4. Other parameter settings in simulation 1.

	High	Mid	Low
Response	60	30	5
Focus	60	30	5
TUT	0	0	0

Table 5. Setpoints for task related actions in each condition in simulation 2–1.

	High	Mid	Low
Tau	1/10	1/100	1/1000

Table 6. Time decay for task related in each condition in simulation 2–2.

temperature, learning rate, and weight parameter (w) for perseverance were determined based on the simulation results (Table 3). In all simulations, the initial values of the homeostatic state and Q-value were set to 0; thus, the first action was randomly chosen. Furthermore, Table 4 shows the settings of the setpoints and time decay for each action choice. The homeostatic state of "TUT" is equal to the setpoint from the onset of the task, meaning that the reward for "TUT" more likely to becomes negative during the task. However, since reward depends on all changes in homeostasis, if a particular homeostasis state deviates from its set point due to a certain response, but other homeostasis states approach those set points, the reward will be positive.

Simulation 2
To replicate the finding that motivation for the task reduces the proportion of MW, we manipulated the setpoint (Table 5) and the speed of spontaneous attenuation in the homeostatic state (Table 6) for the task-related actions. Since higher setpoints led to a larger deviation between the homeostatic state and setpoints, this parameter could directly control motivation. Time decay specifies the drive's recovery speed and seemed to be motivation. The setpoints and time decay used in Simulation 1 were set as the reference (Mid), and we manipulated this value to lower (Low) or higher values (High). The other parameters are the same as those in the full model used in Simulation 1.

Simulation 3
To replicate the finding that highly concerned events drive MW, we manipulated the setpoint to the "TUT" (Table 7). As mentioned above, the setpoints directly control the action's motivation. We increased only the setpoint for "TUT" from that used in the full model in Simulation 1, and the other parameters were the same.

Simulation 4
To replicate the relationship between the proportion of MW and task difficulty, we manipulated the change probability of homeostatic state depending on the action choice (Table 8). The agent does not fulfill the task as

	Response	Focus	TUT
Low	30	30	0
Mid	30	30	30
High	30	30	60

Table 7. Setpoints for task-unrelated thought in each condition in simulation 3.

	Extreme	High	Mid	Low
Response	1	1	1	1
Focus	0	0.1	0.75	1
TUT	1	1	1	1

Table 8. Setting the probability of homeostatic state change for the action choice for each condition of simulation 4.

expected by manipulating the probability of homeostatic state updates. We tested the effect of task difficulty by setting the probability used in the full model of Simulation 1 to the reference condition (Low) and reducing the probability from there. Additionally, to reproduce a situation in which the task was extremely difficult, we used a condition in which the probability of homeostatic state change was set to 0, in which the agents did not feel fulfilled regardless of the extent to which they engaged in the task. The other parameters are the same as those in the full model of Simulation 1. We generated data in all simulations using Julia (version 1.7.2).

Data availability

All relevant data are within the paper (Figs. 2, 3, 4 and 5) and the data and Figures were generated using author's scripts (see Code availability).

Code availability

The codes generating and analysing data were available via GitHub at https://github.com/shina-k/HRL-MW.git.

Received: 26 June 2024; Accepted: 28 February 2025
Published online: 13 March 2025

References

1. Smallwood, J. & Schooler, J. W. The science of Mind wandering: empirically navigating the stream of consciousness. *Annu. Rev. Psychol.* **66**, 487–518 (2015).
2. Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N. & Andrews-Hanna, J. R. Mind-wandering as spontaneous thought: a dynamic framework. *Nat. Rev. Neurosci.* **17**, 718–731 (2016).
3. Killingsworth, M. A. & Gilbert, D. T. A wandering Mind is an unhappy Mind. *Science* **330**, 932 (2010).
4. Brosowsky, N. P., DeGutis, J. & Esterman, M. Mind wandering, motivation, and task performance over time: evidence that motivation insulates people from the negative effects of Mind wandering. *Psychol. Conscious. Theory Res. Pract.* **10** (4), 475–486. https://doi.org/10.1037/cns0000263 (2020).
5. Zanesco, A. P., Denkova, E. & Jha, A. P. Mind-wandering increases in frequency over time during task performance: an individual-participant meta-analytic review. *Psychol. Bull.* https://doi.org/10.1037/bul0000424 (2024).
6. Baird, B., Smallwood, J. & Schooler, J. W. Back to the future: autobiographical planning and the functionality of mind-wandering. *Conscious. Cogn.* **20**, 1604–1611 (2011).
7. Poerio, G. L., Totterdell, P. & Miles, E. Mind-wandering and negative mood: does one thing really lead to another? *Conscious. Cogn.* **22**, 1412–1421 (2013).
8. Smallwood, J. & Schooler, J. W. The restless Mind. *Psychol. Bull.* **132**, 946–958 (2006).
9. Seli, P., Konishi, M., Risko, E. F. & Smilek, D. The role of task difficulty in theoretical accounts of Mind wandering. *Conscious. Cogn.* **65**, 255–262 (2018).
10. Thomson, D. R., Besner, D. & Smilek, D. In pursuit of off-task thought: Mind wandering-performance trade-offs while reading aloud and color naming. *Front. Psychol.* **4**, 360 (2013).
11. Barrington, M. & Miller, L. Mind wandering and task difficulty: the determinants of working memory, intentionality, motivation, and subjective difficulty. *Psychol. Conscious. Theory Res. Pract.* https://doi.org/10.1037/cns0000356 (2023).
12. Kahmann, R., Ozuer, Y., Zedelius, C. M. & Bijleveld, E. Mind wandering increases linearly with text difficulty. *Psychol. Res.* **86**, 284–293 (2022).
13. Xu, J. & Metcalfe, J. Studying in the region of proximal learning reduces Mind wandering. *Mem. Cogn.* **44**, 681–695 (2016).
14. Varao-Sousa, T. L., Smilek, D. & Kingstone, A. In the lab and in the wild: how distraction and Mind wandering affect attention and memory. *Cogn. Res. Princ. Implic.* **3**, 1. https://doi.org/10.1186/s41235-018-0137-0 (2018).
15. Bench, S. W. & Lench, H. C. On the function of boredom. *Behav. Sci.* **3**, 459–472 (2013).
16. Newberry, A. L. & Duncan, R. D. Roles of boredom and life goals in juvenile delinquency. *J. Appl. Soc. Psychol.* **31**, 527–541 (2001).
17. Phillips, S. W. Police discretion and boredom: what officers do when there is nothing to do. *J. Contemp. Ethnogr.* **45**, 580–601 (2016).
18. Breland, K. & Breland, M. The misbehavior of organisms. *Am. Psychol.* **16**, 681–684 (1961).
19. Eisenberger, R., Karpman, M. & Trattner, J. What is the necessary and sufficient condition for reinforcement in the contingency situation? *J. Exp. Psychol.* **74**, 342–350 (1967).
20. Falk, J. L. Schedule-induced polydipsia as a function of fixed interval length. *J. Exp. Anal. Behav.* **9**, 37–39 (1966).
21. Gentry, W. D. Fixed-ratio schedule-induced aggression. *J. Exp. Anal. Behav.* **11**, 813–817 (1968).

22. Kachanoff, R., Leveille, R., McLelland, J. P. & Wayner, M. J. Schedule induced behavior in humans. *Physiol. Behav.* **11**, 395–398 (1973).
23. Levitsky, D. & Collier, G. Schedule-induced wheel running. *Physiol. Behav.* **3**, 571–573 (1968).
24. Skinner, B. Superstition in the pigeon. *J. Exp. Psychol.* **38**, 168–172 (1948).
25. Ashwood, Z. C. et al. Mice alternate between discrete strategies during perceptual decision-making. *Nat. Neurosci.* **25**, 201–212 (2022).
26. Reimer, J. et al. Pupil fluctuations track fast switching of cortical States during quiet wakefulness. *Neuron* **84**, 355–362 (2014).
27. Lin, Y. & Westgate, E. C. *The Origins of Boredom: The Oxford Handbook of Evolution and the Emotions* 317–338 (Oxford University Press, 2024).
28. Eisenstein, E. M. & Eisenstein, D. A behavioral homeostasis theory of habituation and sensitization: II. Further developments and predictions. *Rev. Neurosci.* **17**, 533–557 (2006).
29. Juechems, K. & Summerfield, C. Where does value come from? *Trends Cogn. Sci.* **23**, 836–850 (2019).
30. Keramati, M. & Gutkin, B. Homeostatic reinforcement learning for integrating reward collection and physiological stability. *Elife* **3**, 11. https://doi.org/10.7554/eLife.04811 (2014).
31. Lee, C. R., Chen, A. & Tye, K. M. The neural circuitry of social homeostasis: consequences of acute versus chronic social isolation. *Cell* **184**, 1500–1516 (2021).
32. Uchida, Y., Hikida, T. & Yamashita, Y. Computational mechanisms of osmoregulation: A reinforcement learning model for sodium appetite. *Front. Neurosci.* **16**, 857009 (2022).
33. Held, M., Minculescu, A., Rieger, J. W. & Borst, J. P. Preventing mind-wandering during driving: Predictions on potential interventions using a cognitive model. *Int. J. Hum. Comput. Stud.* **181**, 103164 (2024).
34. van Vugt, M., Taatgen, N., Sackur, J. & Bastian, M. Modeling mind-wandering: a tool to better understand distraction. In *Proceedings of the 13th International Conference on Cognitive Modeling* 252 (University of Groningen, 2015).
35. Taatgen, N. A. et al. The resource-availability model of distraction and mind-wandering. *Cogn. Syst. Res.* **68**, 84–104 (2021).
36. Christian, I. & Agrawal, M. A computational model of unintentional mind wandering in focused attention meditation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 44 (2022).
37. Kurzban, R., Duckworth, A., Kable, J. W. & Myers, J. An opportunity cost model of subjective effort and task performance. *Behav. Brain Sci.* **36** (6), 661–679 (2013).
38. Henríquez, R. A., Chica, A. B., Billeke, P. & Bartolomeo, P. Fluctuating minds: spontaneous psychophysical variability during mind-wandering. *PLoS ONE* **11**, e0147174 (2016).
39. Irrmischer, M., van der Wal, C. N., Mansvelder, H. D. & Linkenkaer-Hansen, K. Negative mood and mind wandering increase long-range temporal correlations in attention fluctuations. *PLoS ONE* **13**, e0196907 (2018).
40. Leszczynski, M. et al. Mind wandering simultaneously prolongs reactions and promotes creative incubation. *Sci. Rep.* **7**, 10197 (2017).
41. Makovac, E. et al. Response time as a proxy of ongoing mental State: A combined fMRI and pupillometry study in generalized anxiety disorder. *Neuroimage* **191**, 380–391 (2019).
42. Seli, P., Cheyne, J. A. & Smilek, D. Wandering Minds and wavering rhythms: linking Mind wandering and behavioral variability. *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1–5 (2013).
43. Shinagawa, K., Itagaki, Y. & Umeda, S. Coexistence of thought types as an attentional state during a sustained attention task. *Sci. Rep.* **13**, 1581 (2023).
44. Shinagawa, K., Tanaka, Y., Terasawa, Y. & Umeda, S. Brain–body interactions influence the transition from mind wandering to awareness of ongoing thought. *BioRxiv* 1 (2024).
45. Smallwood, J., Beach, E., Schooler, J. W. & Handy, T. C. Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *J. Cogn. Neurosci.* **20**, 458–469 (2008).
46. Mooneyham, B. W. & Schooler, J. W. The costs and benefits of mind-wandering: a review. *Can. J. Exp. Psychol.* **67**, 11–18 (2013).
47. McVay, J. C., Meier, M. E., Touron, D. R. & Kane, M. J. Aging ebbs the flow of thought: adult age differences in Mind wandering, executive control, and self-evaluation. *Acta Psychol.* **142**, 136–147 (2013).
48. Seli, P., Ralph, B. C. W., Konishi, M., Smilek, D. & Schacter, D. L. What did you have in Mind? Examining the content of intentional and unintentional types of Mind wandering. *Conscious. Cogn.* **51**, 149–156 (2017).
49. Martínez-Pérez, V. et al. Propensity to intentional and unintentional mind-wandering differs in arousal and executive vigilance tasks. *PLoS ONE* **16**, e0258734 (2021).
50. He, H., Chen, Y., Li, T., Li, H. & Zhang, X. The role of focus back effort in the relationships among motivation, interest, and Mind wandering: an individual difference perspective. *Cogn. Res. Princ. Implic.* **8**, 43 (2023).
51. Gilbert, T. F. Fundamental dimensional properties of the operant. *Psychol. Rev.* **65**, 272–282 (1958).
52. Shull, R. L., Gaynor, S. T. & Grimes, J. A. Response rate viewed as engagement bouts: effects of relative reinforcement and schedule type. *J. Exp. Anal. Behav.* **75**, 247–274 (2001).
53. Podlesnik, C. A., Jimenez-Gomez, C., Ward, R. D. & Shahan, T. A. Resistance to change of responding maintained by unsignaled delays to reinforcement: a response-bout analysis. *J. Exp. Anal. Behav.* **85**, 329–347 (2006).
54. Shull, R. L. Bouts of responding on variable-interval schedules: effects of deprivation level. *J. Exp. Anal. Behav.* **81**, 155–167 (2004).
55. Weinstein, Y. Mind-wandering, how do I measure Thee with probes? Let me count the ways. *Behav. Res. Methods* **50**, 642–661 (2018).
56. Christoff, K., Gordon, A. M., Smallwood, J., Smith, R. & Schooler, J. W. Experience sampling during fMRI reveals default network and executive system contributions to Mind wandering. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8719–8724 (2009).
57. Randall, J. G., Beier, M. E. & Villado, A. J. Multiple routes to Mind wandering: predicting Mind wandering with resource theories. *Conscious. Cogn.* **67**, 26–43 (2019).
58. Mittner, M., Hawkins, G. E., Boekel, W. & Forstmann, B. U. A neural model of Mind wandering. *Trends Cogn. Sci.* **20**, 570–578 (2016).
59. Menon, V. & Uddin, L. Q. Saliency, switching, attention and control: a network model of Insula function. *Brain Struct. Funct.* **214**, 655–667 (2010).
60. Molnar-Szakacs, I. & Uddin, L. Q. Anterior Insula as a gatekeeper of executive control. *Neurosci. Biobehav. Rev.* **139**, 104736 (2022).
61. Schimmelpfennig, J., Topczewski, J., Zajkowski, W. & Jankowiak-Siuda, K. The role of the salience network in cognitive and affective deficits. *Front. Hum. Neurosci.* **17**, 1133367 (2023).
62. Seeley, W. W. The salience network: A neural system for perceiving and responding to homeostatic demands. *J. Neurosci.* **39**, 9878–9882 (2019).
63. Sridharan, D., Levitin, D. J. & Menon, V. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12569–12574 (2008).
64. Munn, B. R., Müller, E. J., Wainstein, G. & Shine, J. M. The ascending arousal system shapes neural dynamics to mediate awareness of cognitive States. *Nat. Commun.* **12**, 1–9 (2021).
65. Yawata, Y. et al. Mesolimbic dopamine release precedes actively sought aversive stimuli in mice. *Nat. Commun.* **14**, 2433 (2023).
66. Buetti, S. & Lleras, A. Distractibility is a function of engagement, not task difficulty: evidence from a new oculomotor capture paradigm. *J. Exp. Psychol. Gen.* **145**, 1382–1405 (2016).
67. Faber, L. G., Maurits, N. M. & Lorist, M. M. Mental fatigue affects visual selective attention. *PLoS ONE* **7**, e48073 (2012).

68. Allison, J. & Timberlake, W. Instrumental and contingent saccharin Licking in rats: response deprivation and reinforcement. *Learn. Motiv.* **5**, 231–247 (1974).
69. Baum, W. M. Rethinking reinforcement: allocation, induction, and contingency. *J. Exp. Anal. Behav.* **97**, 101–124 (2012).
70. Guthrie, E. R. Conditioning as a principle of learning. *Psychol. Rev.* **37**, 412 (1930).
71. Herrnstein, R. J. On the law of effect. *J. Exp. Anal. Behav.* **13**, 243–266 (1970).
72. Killeen, P. R. & Fetterman, J. G. A behavioral theory of timing. *Psychol. Rev.* **95**, 274–295 (1988).
73. Staddon, J. E. R. Operant behavior as adaptation to constraint. *J. Exp. Psychol. Gen.* **108**, 48–67 (1979).
74. Timberlake, W. & Allison, J. Response deprivation: an empirical approach to instrumental performance. *Psychol. Rev.* **81**, 146–164 (1974).
75. Yamada, K. & Toda, K. Habit formation viewed as structural change in the behavioral network. *Commun. Biol.* **6**, 303 (2023).
76. Danckert, J. & Elpidorou, A. In search of boredom: beyond a functional account. *Trends Cogn. Sci.* **27** (5), 494–507 (2023).
77. Wilson, T. D. et al. Just think: the challenges of the disengaged Mind. *Science* **345**, 75–77 (2014).
78. Baum, W. M. On two types of deviation from the matching law: bias and undermatching. *J. Exp. Anal. Behav.* **22**, 231–242 (1974).
79. Katz, K. & Naug, D. Energetic state regulates the exploration–exploitation trade-off in honeybees. *Behav. Ecol.* **26**, 1045–1050 (2015).
80. Corrales-Carvajal, V. M., Faisal, A. A. & Ribeiro, C. Internal States drive nutrient homeostasis by modulating exploration-exploitation trade-off. *Elife* **5**, e19920 (2016).
81. Bastian, M. & Sackur, J. Mind wandering at the fingertips: automatic parsing of subjective States based on response time variability. *Front. Psychol.* **4**, 573 (2013).
82. Zanesco, A. P., Denkova, E., Witkin, J. E. & Jha, A. P. Experience sampling of the degree of Mind wandering distinguishes hidden attentional States. *Cognition* **205**, 104380 (2020).

Acknowledgements

This research was supported by JSPS KAKENHI 24K16869 (KY) and 24KJ0069 (KY).

Author contributions

K.S., and K.Y. conceived the idea for the paper, analyzed data and wrote and edited the paper.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92561-0.

Correspondence and requests for materials should be addressed to K.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.