Hindawi
Complexity
Volume 2021, Article ID 3343494, 20 pages
https://doi.org/10.1155/2021/3343494

Research Article

On Cognitive Searching Optimization in Semi-Markov Jump Decision Using Multistep Transition and Mental Rehearsal

Bingxuan Ren,[1,2] **Tangwen Yin**,[1,2] **and Shan Fu**[1,2]

[1]*Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China*
[2]*Key Laboratory of System Control and Information Process, Ministry of Education, Shanghai 200240, China*

Correspondence should be addressed to Shan Fu; sfu@sjtu.edu.cn

Received 21 July 2021; Revised 25 August 2021; Accepted 1 September 2021; Published 6 October 2021

Academic Editor: Hamid Reza Karimi

Cognitive searching optimization is a subconscious mental phenomenon in decision making. Aroused by exploiting accessible human action, alleviating inefficient decision and shrinking searching space remain challenges for optimizing the solution space. Multiple decision estimation and the jumpy decision transition interval are two of the cross-impact factors resulting in variation of decision paths. To optimize the searching process of decision solution space, we propose a semi-Markov jump cognitive decision method in which a searching contraction index bridges correlation from the time dimension and depth dimension. With the change state and transition interval, the semi-Markov property can obtain the action by limiting the decision solution to the specified range. From the decision depth, bootstrap re-sampling utilizes mental rehearsal iteration to update the transition probability. In addition, dynamical decision boundary by the interaction process limits the admissible decisions. Through the flight simulation, we show that proposed index and reward vary with the transition decision steps and mental rehearsal frequencies. In conclusion, this decision-making method integrates the multistep transition and mental rehearsal on semi-Markov jump decision process, opening a route to the multiple dimension optimization of cognitive interaction.

1. Introduction

The human-computer cognitive interaction (HC^2I) process can be embodied to analyze human factors, interactive performance, and decision uncertainty. In terms of decision making, the decision solution space constructed by the estimation and searching for the solution path is under effect with the uncertainty of human decision [1]. In the HC^2I process, the chronologically ordered decision path based on human experience is composed of each decision action step which is uniquely determined under the estimation of the future decision path. From a prior perspective, due to the influence of decision jumpy intervals and the multiple estimation of decision paths, there are infinite possibilities while deciding the decision path from its solution space. It is necessary to reduce the impact of the exploration of solutions on the efficiency of decision making. In the optimization of the cognitive searching, human's high-level control hierarchy makes preparations for upcoming decision before people realize it [2]. When exploring the decision solution space, searching contraction optimization is used to show that people have subconsciously eliminated some decision paths that would not actually be made.

In order to analyze decision behaviors, human performance modeling (HPM) has been researched in the last few decades [3]. HPM tends to demonstrate the interactive relations through designing different inner structures. It evolves from the broad symbolism cybernetic approaches [4] to the new stage of computational rational modeling [5], involving human cognitive behavior at various decision hierarchies. Similar to the non-homogeneous sequential model, decisions are continuously generated in the chronological order. Through depicting the potentially possible distribution caused by differentiation structures, HPM essentially contracts and prunes the immense decision sequence formed by rehearsal.

Similar to the human-like behavior, multistep decision is mutually influenced during the periods rather than the

instant moment. The Markov property, strengthening the correlation in decision path, constricts that the selection of action elements is only relevant to the decision adopted at the previous moment. Based on the Markov decision principle, the policy iteration method calculates the one-step reward value by introducing the state of decision object to computation [6]. However, existing physical obstruction makes humans unable to access state parameters without sensor measurement in the interaction environment, which causes unavoidable deviation. To cover this shortage, the partially observable Markov decision method uses interactive object state as an uncertain estimation of observable state set, which also is the main difference from the observable state Markov decision method [7]. Another completely unobservable method named hidden Markov decision analyzes human cognitive behaviors through the state of the interactive object (such as machine). In the aforementioned methods, there exist similar deficiencies on depicting the transition state interval and calculating reward value in a locality. Therefore, the variant intervals and reward value combined by history decision steps and future decision steps are crucial.

On the other hand, adjacent action which lacks mutual influence is a shortcoming for Markov property [8]. It leads to being short of tightness interference in human-like decisions. Different from the single-state inference in Markov process, the semi-Markov decision focuses on the cross-correlation of transition interval, even though the transition state is not ergodic and innumerable. Its state is jumpy and changeable accompanying with the decision process.

In this paper, a semi-Markov jump decision method is proposed to optimize the human cognitive searching decision path through the multistep transition part and mental rehearsal part in a specified airplane pilot interaction scenario. We define a searching contraction index to represent the coverage degree. The coverage degree refers to a ratio between decision behaviors chosen subconscious and all accessible decision behaviors. For a more general situation, the semi-Markov decision process overcomes the restriction by adopting the time-varying transition rate. Thus, the sojourn time between each mode can be of any non-exponential distribution. Besides, the human making decision is different from a fixed step decision controller. The time interval in sequential decision is not a constant and is arbitrary. In addition, it cannot be modeled by noise like exponential distribution which obeys the Markov transition law. We consider the semi-Markov process and human-centered reinforcement Q-learning to realize the estimated decision solution. Depending on the state of inconsistent transition interval, the composition decision step accomplishes making decisions. As the core of decision making, the dynamical transition probability motivates state transition and action adopted. The bootstrap re-sampling frequency can abstract the mental rehearsal process by re-screening the transition probability. Finally, decision boundary influenced by the interactive object constricts the final human decision. Figure 1 briefly shows above compounding relation. To summarize, this paper puts forward the following four contributions:

(i) A semi-Markov jump cognitive decision method is proposed to evaluate the dynamical cognitive interaction process. Our method integrates the semi-Markov decision transition interval, the multiple decision path estimation, and the changeable decision solution space for jump state.

(ii) The transition interval and sojourn time, which are of vital importance characteristics, have been preferably reflected in our method. By adding mental rehearsal property, our method addresses the reduction of infinite-dimensional decision solution space and forward advances the dimension deduction to a smaller range.

(iii) An introduced index named searching contraction can efficiently reflect cognitive computation ability of human while exploring the decision solution space.

(iv) Our method incorporates the relation in decision time and depth, conforming to the human being's logic of deciding and the property of transition state jumpy property.

The rest of the paper is organized as follows. Section 2 briefly describes the related works about multistep transition, mental rehearsal, and dynamical searching dimension in decision. Section 3 and Section 4 emphasize the specific problem and illustrate how our decision method is built for decision making. Section 5 and Section 6 detail the experiments and the integral analysis of this model, including its shortcomings, and the future directions in this area. Section 7 summarizes this paper.

2. Related Work

2.1. Multistep Transition. The multistep transition happening in the decision making has been developed with many methods [9], such as reinforcement learning [10], utility selection theory [11], and networked control system [12]. Compared to the continuously accumulated and improved process, the common decision framework is to obtain an optimal decision strategy via the feedback effect [13] and evaluate the potential outcome values caused by events when decision is formulated by the feedback loop design [14]. Focusing on cognitive analysis, Yanco and Drury [15] modified a taxonomy of multiagent systems and treated the human-computer interactions as a process of two heterogeneous agent interactions. Moratz et al. [16] experimented with a comparison test between human-robot and human-human to illustrate the difference in spatial features. The comparative trials implicitly revealed that the complexity of cognitive space plays a prominent role in the interaction process. From the view of multiple timescales, such as cognition and decision, Purcell and Kiani [17] designed a hierarchy of multistep transition decision on processes to disambiguate the detrimental factors such as flawed information. All of the above works mainly focused on the differences between human and robot as the autonomous agent. They considered human as uncertain and non-

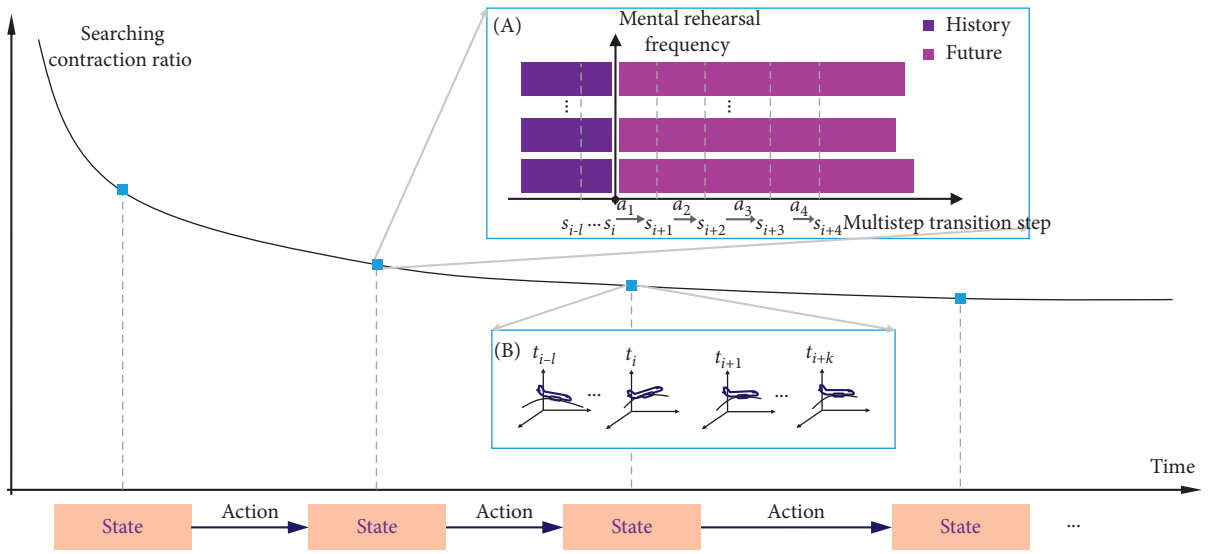


FIGURE 1: Semi-Markov jump decision in one kind of HC^{2}I process. The searching contraction ratio decreases with the decision time increasing. The ratio is calculated by different decision possibilities. The state is jumpy and changeable with time. Parts A and B show the mental rehearsal and multistep transition happening at one specified decision state. We use the vertical axes to represent mental rehearsal. It constructs the decision dimension together with the multistep decision.

monotonous agent by adding the stochastic-dynamic transition interval which accords with exponential distribution. Wu et al. [18] added the transition time restriction into the semi-Markov model while its finite system state was limited into the noninteractive modal. To the best of our knowledge, the non-exponential distribution of transition interval which can be used into cognitive decision making has received limited effort so far. Decision interval of Markov jumping provides a more general way to describe the multistep transition for cognitive decision.

2.2. Mental Rehearsal. Mental rehearsal, also known as mental simulation, is one of the cognitive strategies [19]. This strategy takes future action-practice without outer observed physical performance. In the typical task, it is regarded as one of the efficient methods to improve the decision performance of the psychomotor and sport. For example, Miranda et al. [20] used mental rehearsal to decrease depressive predictive certainty, which showed the gains in making optimistic predictions. Ignacio et al. [21] proposed that different health disciplines can utilize mental rehearsal strategy as a part of clinical training. Su et al. [22] designed incremental deep convolutional neural network process to demonstrate the human-like learning behavior. Moreover, researchers analyzed its different effects on the user's learning decision in the theory of working memory [23]. As a computational model, Oberauer and Lewandowsky [24] designed a time-based resource-sharing theory to derive unambiguous predictions about the effect of rehearsal on memory, which is beneficial for differentiating between varying forms of mental practice. Besides, mental rehearsal can be analyzed by parameterized formation. To demonstrate the advantage of rehearsal, Mazher et al. [25] found that rehearsal was beneficial for memorized long-term learning by discriminating the learning decision states using electroencephalography.

2.3. Dynamical Decision Dimension. Exploration-exploitation related to the dynamical decision dimension is a crucial aspect, especially for searching the feasible solutions [26]. The dimensional optimization method connects with the decision-making property such as the non-Markovian property [27], which is used to describe the cross-influence between different decision states. To reduce the dimension of decision searching, Engel et al. [28] considered the stochastic jumpy interval in human cognitive decision behaviors and handled it with a linearity weighted logic according to monotonically increased time [29]. To get the global optimal solution, the brute-force calculation method is used. But it is easily trapped into the plight to search the space in finite polynomial time, especially under the non-convex issues. Some proposed optimal algorithms such as best proximity points [30] and particle swarm optimization [31] were applied to evade the non-convex difficulties. However, there still exists an enormous gap between human physical simulation and computational simulation like emotions [32]. The state caused by human action is discretely jumpy rather than the inflexible inference from a fixed step to another. In addition, the uncertain human factors enlarge the difficulty covering decision solution space of all accessible scenes [33].

3. Problem Formulation

The HC^{2}I process is able to give people insight into and observe the state information from the interactive machine object. According to the state obtained by observation and the state of the historical decision path, people make new decision in limited period. Although related work provides

state of the art in terms of multistep decision, mental rehearsal, and dynamical decision dimension, existing methods cannot optimize the cognitive decision searching from dimension of time and depth. The current methods implicitly contain a flaw that the process of searching decision solution space depends on the partial history information. Limited to the single sample estimation, evaluation of decision also leads to losing the unbiasedness of decision. Additionally, cognitive searching optimization is involved with the human cognitive properties which have not been fully used in historical research, such as jumpy decision interval and multiple decision path estimation. Therefore, the problem in our work is to optimize the decision reward $\mathscr{R}$, generate efficient decision path b, increase the searching contraction ratio μ, and stabilize the decision solution space scope under the two human cognitive characteristics. We design a semi-Markov jump decision method using the hierarchical transition probability optimization from the different step lengths of multistep transition and mental rehearsal frequency. First, we design the semi-Markov process and reinforcement Q-learning to form the multistep transition on the basis of history fragment strategies and subjective estimation about future's expectation feedback. Then, we build the mental rehearsal optimization based on bootstrap re-sampling, which plays an essential role for human's subconscious simulation and optimizes transition probability. Besides, the decision space boundary ensures that decision admissible is developed.

4. Semi-Markov Jump Decision Method

In this section, we show the semi-Markov jump decision method for cognitive searching optimization. As shown in Figure 2, the interior of method can be divided into two hierarchies. The first hierarchy indicates the targeted decision inference block. During the decision process, this block limits the interaction target domain and illustrates the maximum-minimum reward for the decision process. The second hierarchy determines searching contraction block by estimating transition distribution. Human decision memory is not amnesic instantly once action completed while considering non-Markov property [34]. It implicitly indicates that decision making does not rely on a point but a fragment. Here we use block of semi-Markov process and human-centered reinforcement Q-learning decision maker to estimate the decision state on the limited length fragment, which is capable of jumpy transition interval belonging to non-exponential distribution. The bootstrap re-sampling controller block is designed to explore the optimal transition probability for the human mental rehearsal. Depending on flight phase, decision space boundary block limits decision to admissible scope for decision inference and flight dynamics. Besides, airplane flight simulating block receives the observable airplane state information from space $\mathscr{N}^*$, whereas it handles executable action parameters from the block of decision inference target.

4.1. Targeted Decision Inference on Receding Horizon. The whole cognitive interaction is defined in the HC^{2}I space $\mathscr{N}$. Interior of decision method is defined in decision solution space $\mathscr{N}^*$. Also, we define the inner bootstrap space as re-sampling with replaceable space $\mathscr{N}^\dagger$, which is a subset of space $\mathscr{N}^*$. Similar to a sliding surface forcing the system state in semi-Markov jump system [35], the targeted decision inference is addressed on receding horizon. In space $\mathscr{N}$, the aimless interaction decision is excluded in the scope of paper. We assume that the HC^{2}I process has preassigned target set Θ where its elements relate to machine (computer) state at each decision π. The constrained loss function is given below, which is similar to a filter using energy comprehensive index to get the optimal decision trail Π under the specified target.

$$\prod_{\pi^* \in \Pi} \arg\max_{\theta_1} \arg\min_{\theta_2} \mathscr{F}\left(\pi | \mathbb{E}_{\tau \sim p_\theta^\phi(\tau)}\left[\mathscr{R}_{\text{total}}(\pi)\right]\right), \tag{1}$$

where

$$p_\theta^\phi(\pi) = \mu(s_0) \prod_{t=1}^{T-1} \mathbf{p}_\phi(a_{t+1}) \pi_\theta(a_t, s_t), \tag{2a}$$

$$\begin{aligned} \mathbf{p}_\phi(a_{t+1}) &= P\left(a_{j+1} | a_1, a_2, \ldots, a_j\right) \\ &= P\left(a_{j+1} | a_{j-i}, a_{j-(i-1)}, \ldots, a_{j-1}, a_j\right), \quad i \geq 1, j = t, \end{aligned} \tag{2b}$$

$$\pi_\theta(a_t, s_t) = \sup\left\{\pi = x^* | \epsilon^*(x^*, F_n) = \widehat{\theta}(f_n^*) - \theta(f_n)\right\}, \tag{2c}$$

where $\mathscr{R}$ represents the synthesis reward function decided by decision process; it minimizes the energy consumed by interaction of cognitive and gets a sequence maximizing the computer performance while the decision π follows distribution of trajectory density function parameterized by both transition probability and stochastic error factor. $\mathbf{p}_\phi(a_{t+1})$ states that the HC^{2}I process of is capable of semi-Markov factor θ_1. Also, both the temporal fragment factor and jumpy factor make contribution and intervention in process. $\pi_\theta(a_t, s_t)$ states that mental rehearsal factor θ_1 contributes to the decision. We need to get the optimal policy under the minimum condition θ_2 to ascertain the maximum θ_1. θ_1 denotes the maximum value $\pi_\theta(a_t, s_t)$ process. θ_2 relates to the minimum of smaller worst cost function $\mathscr{R}$. At each time step t, the agent is in state $s_t \in S$ and must choose an action $a_t \in A$, transitioning it to a new state $s_{t+1}\, P(s_{t+1}|s_t, a_t)$ and yielding a reward $R(s_t, a_t)$. A policy π: $S \times A \longrightarrow [0, 1]$ is defined as a probability distribution over state-action pairs, where $\pi(a_t|s_t)$ represents the density of selecting action a_t in state s_t. Upon consequent interactions with the environment, the agent collects a trajectory τ of state-action pairs. The goal is to determine an optimal policy π^* by this loss function.

Besides, the constrained loss function satisfies two implicitly postulated conditions. The first condition indicates that the step number of decision is limited. It shows that the cognitive interaction exists in a terminal state. Also, the solution space is bounded by the environment tasks. The second assumption explains that the computer or machine state pattern is similar

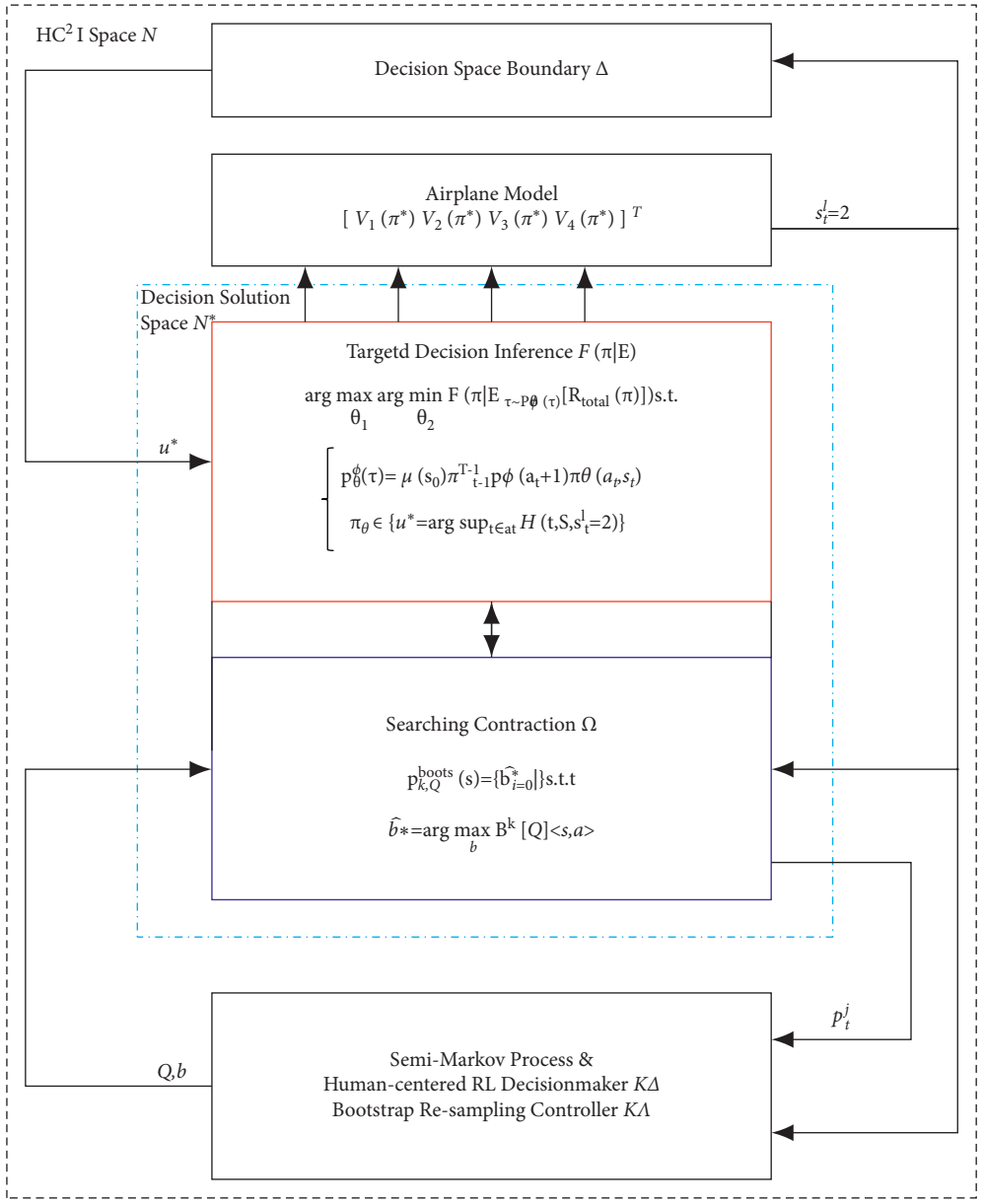


FIGURE 2: The architecture of semi-Markov jump decision method. The multistep transition and mental rehearsal are represented separately by the decision maker K_Δ and bootstrap re-sampling controller K_Λ in the bottom block. It receives the probability $\frac{j}{t}$ and calculates b and Q for searching contraction block. The airplane model block receives different control parameters from human decision making and outputs the observable state $s_t^{l=2}$. The decision space boundary calculates the admissible control scope for the final decision judgment.

during this trail process, and it assures that the process can be properly classified into several stages.

4.2. Semi-Markov Process and Human-Centered Q-Learning Decision Maker for Multistep Transition. Decision maker K_Δ is composed of a hybrid semi-Markov process with the forward human-centered Q-learning estimation. To determine Q parameter in $\mathscr{B}^k[Q]\langle s_i, a_i\rangle$ for equation (19), we consider composite decision maker K_Δ by the semi-Markov process and human-centered Q-learning. The former takes jumpy property and sojourn time of decision interval into consideration by the semi-Markov process, while the latter calculates future predictive estimation in $\mathscr{N}^\dagger$. Figure 3 shows the sketch of this part.

The sampled discrete state trail (specified value k) of state-action pair is $S_{\text{trail}} \triangleq \{S_0, S_1, \ldots, S_t, \ldots, S_n\}$. Subscript n is the number of state elements in trail. It is hidden left part of b^* from start of decision process. The action a_i is chosen from the action set. We define A as $A \triangleq \{a^1, a^2, \ldots, a^{\|m\|}\}$, where m is the number of admissible control elements in decision π_t. $O = \{o_0, o_1, o_2, \ldots, o_v\}$ is observable state variable from process $\widehat{M}_t$ and its subscript v is dimension for s_t^l when l refers to $\widehat{M}_t$. Based on the Markov property $P\{X_t = y|X_r, 0 \le r \le s\} = P\{X_t = y|X_s\}$ and jumpy transition probability

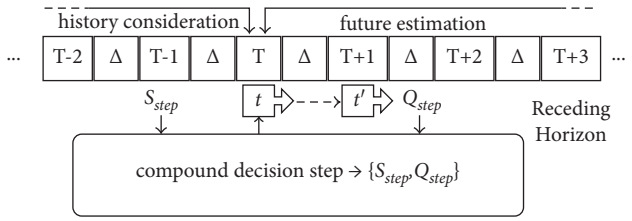


Figure 3: Semi-Markov process and human-centered Q-learning decision maker K_Δ for synthesis consideration. The figure shows how history consideration decision and future estimation decision make effect together on the time series names as receding horizon. Interval time Δ is jumpy transition between accord with the non-exponential distribution.

$P\{X_{t+\delta t} = x | X_t = y\} = \alpha(y, x)\delta t + o(\delta t), y \neq x$, the following functions state the semi-Markov relation between adjacent states s_i. Here, X_t is the stochastic process, $\alpha(x, y)$ denotes the jumpy transfer rate from state x to adjacent state y, δt is the decision interval following the non-exponential distribution, and $o(\delta t)$ denotes a high-order stochastic variable which is small.

$$p(s_t|\langle s_{t-1}, a_{t-1}\rangle, \ldots, \langle s_0, a_0\rangle) = p\Big(s_t|\langle s_{t-1}, a_{t-1}\rangle, \ldots, \langle s_{t-_{\text{step}}}, a_{t-_{\text{step}}}\rangle\Big), \tag{3a}$$

$$p(s_t|\langle s_{t-1}, a_{t-1}\rangle) = \alpha(s_t, s_{t-1})\delta t + o(\Delta t), \tag{3b}$$

$$\sum_S \sum_{\langle s_t, a_t\rangle} p\Big(s_t| \cup \Big\{\langle s_{t-1}, a_{t-1}\rangle, \ldots, \langle s_{t-S_{\text{step}}}, a_{t-S_{\text{step}}}\rangle\Big\}\Big) = 1. \tag{3c}$$

In space $\mathcal{N}^\dagger$, the quadruple (S, A, O, T) is the composite of elements in observable semi-Markov discrete process M_t. $T = [0, \infty)$ and t is discrete Lebesgue additive on the measure μ. The state transition is denoted as operation F: $O \times S \longrightarrow^A S\prime(\pi)$; then, the observable state o_t is from rigorous time homogeneous continuous Markov process $\widehat{M}_t$ whose tuple form is $(\widehat{S}, \widehat{A}, T)$ and $\widehat{F}$: $\widehat{S} \times \widehat{A} \longrightarrow \widehat{S}\prime(\widehat{\pi})$. According to m, dimension of action set a_t is dynamically changeable under the updating transition probability p. For $\widehat{M}_t$ process, the element of action set $\widehat{A}$ is the same as A and it receives the decision a_i from M_t. When a new decision is determined by M_t, a_t will be transmitted to $\widehat{M}_{t+1}$ after the correction $v(\cdot)$. $v(\cdot)$ is normal distribution represented by the transition error. Here we assume that $\widehat{M}_t$ does not exhibit parameter drift such as time delay factor. We write the actual action set form A and $\widehat{A}$ as

$$A_t = A\big(p(\cdot|t)|s_t\big) = \Big[a^1, a^2, \ldots, a^{m\prime}\Big], m\prime \in [1, m], \tag{4a}$$

$$\widehat{A}_t = v(A_t) = \Big[\widehat{a}_t^1, \widehat{a}_t^2, \ldots, \widehat{a}_t^{\prime m}\Big]. \tag{4b}$$

The accumulated state S and indicator state $\widehat{S}$ are calculated from state in the M_t and $\widehat{M}_t$. Therein, semi-Markov process M_t follows the non-exponential distribution sojourn time Δt but $\widehat{M}_t$ is continuous without sojourn time. We estimate it by the interval information entropy. Here $w(a_{t-1})$ is the weight coefficient of action, and superscript r' is the observable state dimension in $\widehat{M}_t$.

$$S_t = \text{diag}\big[s_t, o(\widehat{s}_t)\big], s_t = \sum_1^m w(A_{t-1}) a_{t-1}^{m'}, \tag{5a}$$

$$\widehat{S}_t = \widehat{s}_t = \Big[\widehat{s}_t^1, \widehat{s}_t^2, \ldots, \widehat{s}_t^{r'}\Big]. \tag{5b}$$

Also, the semi-Markov process M_t within human and regular Markov process $\widehat{M}_t$ within machine (computer) happen synchronously, while the former is discrete and the latter is continuous. Therefore, the intervention relations between M_t and $\widehat{M}_t$ can be represented as follows. Here ξ, ζ is the error variable that follows normal distribution. $\Delta\widehat{M}_t, \Delta M_t$ separately indicate the continuous and discrete interval period from different processes.

$$\mathrm{d}S = \mathrm{d}\widehat{S} + \xi + O(\widehat{S})\Delta\widehat{M}_t, \tag{6a}$$

$$\mathrm{d}A = \mathrm{d}\widehat{A} + \zeta + O(\widehat{A})\Delta M_t. \tag{6b}$$

For regular Markov process $\widehat{M}_t$, the expectation performance $\mathbb{E}(\widehat{M}_t)$ is obtained at decision π_t. According to Doeblin lemma [36], let $\widehat{P}$ be a transition probability matrix: $\forall i \in \mathcal{N}^\dagger$, $(P)_{ij_0} \geq \varepsilon$ when $j_0 \in \mathbb{S}$ and $\varepsilon > 0$, there exists only stationary probability vector $(\pi)_{j_0} \geq \varepsilon$ in P, and for all initial distribution μ, $\|\mu\mathbf{P}^n - \pi\|_v \leq 2(1-\varepsilon)^n, n \geq 0$. This lemma presents that the amnesic initial of distribution exists in $\widehat{M}_t$. On the other hand, the expectation performance $E(M_t)$ is an accumulated reward about state $S_t \in S$. It is a continuous additive from the initial of history consideration step to the current decision time, including jumpy transition interval and decision action time that follows normal distribution. Let $R_t^-(\xi) = t - \tau_{N_t}(\xi), \mathrm{R}_t^+(\xi) = \tau_{N_{t+1}}(\xi) - t$ where $N_t(\xi) = \sup\{n\colon \tau_n(\xi) \leq t\}$ is the number of jumps for function ξ up to a time t. Then, probability kernel function of the semi-Markov process is

$$\begin{aligned} P_x\big(R_t^- \geq r, R_t^+ \geq s, X_t \in S\big) &= \sum_{k=0}^{\infty} P_x\Big(X_{t_n} \in A, \tau_n \leq t - r, \tau_{n+1} \geq t + s\Big) \\ &\overset{n \longrightarrow \infty}{=} \int_0^{t-r} \int_S P_{x_1}(\tau_1 > t + 2 - s_1) U(x, \mathrm{d}s_1 \times \mathrm{d}x_1), \end{aligned} \tag{7}$$

where $U(x, [0, t] \times S) = \sum_{k=0}^{\infty} P_x(\tau_n \leq t, X_{\tau_n} \in S)$. P_x is a measure of intensity of random point field for fixed x. Assuming that $N(B \times S) = \sum_{n=0}^{\infty} I_{B\times S}(\tau_n, S_n)$ represent the number of discontinuity pairs belonging to a set

$B \times S (B \in \mathscr{B}((R)_+), S \in \mathscr{B}((S))$, we have expectation $\mathbb{E}(M_t)$.

$$\begin{aligned}\mathbb{E}_x(N(B\times S)) &= \sum_{n=0}^{\infty} E_x\left(I_{B\times S}(\tau_n, X_n)\right)\\ &= \sum_{n=0}^{\infty} P_x\left(\tau_n \in B, X_{\tau_n} \in S\right) = U(x, B\times S),\\ \mathbb{E}(M_t) &= \mathbb{E}_{x=t_n}\left(N_t(\xi), R_t^-(\xi), \alpha\right)\\ &= \int_{\min}^{\max} e^{-\delta t} P_{x=t_n}\left(R_t^- \geq r, R_t^+ \geq s, X_t \in S\right)\\ &= \mathbb{E}\left(M_{(t-S_{\text{step}})}\right) + \int_{t-S_{\text{step}}}^{t} \frac{1}{H}\mathbb{E}(s_t|\pi),\end{aligned} \tag{8}$$

where π is history decision determined and $H \equiv \sup_i R_i = \sup_i(-(\alpha_i i))$. In the multistep decision process, humans anticipate fuzzy assessment before making the decision [37]. Next, we use human-centered reinforcement Q-learning to estimate accumulated feedback and the maximized performance as the transition probability in short future period Q_{step}.

$\{b|b = \pi_1, \pi_2, \ldots, \pi_n\}$ denotes the optimization limited trail for future perdition decision sequences, and subscript $n = Q_{\text{step}}$ is predicted future decision step length. Given a future estimation process $\{X\colon x_k\} \subset \mathscr{N}^\dagger$, which is right continuous part of b, we write performance index as $Q_n = \sum_{k=0}^{k=n} V(x_k, a_k)$. Bellman optimal theory [38] illustrates that optimal decision sequence can be divided into several blocks staying in optimal state space. It makes sure the sufficiency for division of b. According to Bellman optimal theory, we derive Q_n to $Q_n^* = \min_{u_0}[V(x_0, u_0) + Q_{n-1}^*]$ [39].

$$\begin{aligned}V_T^\pi(x) &= E_\pi\left[\frac{1}{T}r_1 + \frac{T-1}{T}\frac{1}{T-1}\sum_{t=2}^{T} r_t|x_0 = x\right]\\ &= \sum_{a\in A}\pi(x,a)\sum_{x\prime\in X}\epsilon_{x\longrightarrow x'}^a\left(\frac{1}{T}R_{x\longrightarrow x'}^a + \frac{T-1}{T}E_\pi\left[\frac{1}{T-1}\sum_{t=1}^{T-1} r_t|x_0 = x'\right]\right)\\ &= \sum_{a\in A}\pi(x,a)\sum_{x\prime\in X}\epsilon_{x\longrightarrow x'}^a\left(\frac{1}{T}R_{x\longrightarrow x'}^a + \frac{T-1}{T}V_{T-1}^\pi(x')\right),\end{aligned} \tag{9}$$

where $R = 1_{\widehat{S}}$, ϵ is the exploration ratio, and r is the reward. To calculate the value function $V(x_k, a_k)$, we have the following derivations. First, we consider

$$V_T^\pi(x) = E_\pi\left[\frac{1}{T}\sum_{t=1}^{T} r_t|x_0 = x\right] = E_\pi\left[\frac{1}{T}r_1 + \frac{1}{T}\sum_{t=2}^{T} r_t|x_0 = x\right]. \tag{10}$$

Furthermore, we let $R_{x\longrightarrow x'}^a$ to denote r_1 and

$$V_{T-1}^\pi(x') = E_\pi\left(\frac{1}{T-1}\sum_{t=1}^{T-1} r_t|x_0 = x\prime\right). \tag{11}$$

Through the law of total probability expansion, the operation $E_\pi(\cdot)$ is substituted by the following expression:

$$\sum_{a\in A}\pi(x,a)\sum_{x\prime\in X}\epsilon_{x\longrightarrow \acute{x}}^a \tag{12}$$

Then, we can get the function equality equation (9). $\widehat{S}$ is an indicator function format.

$$1_{\widehat{S}_t} = \begin{cases} 1, & \text{for}\,\widehat{S} \subset \Theta^*,\\ 0, & \text{for}\,\widehat{S} \subset \Theta^*.\end{cases} \tag{13}$$

For prediction process X, we assume that discrete time sequence is equidistance $|T_j - T_i| = \text{const}$. According to the Bellman equation and formula of total probability, we have the recurrence accumulated feedback. Also, this predicted reward will be used to determine optimal transition probability on space $\mathscr{N}^\dagger$ in the next section.

$$Q_T^\pi(x,a) = \mathbb{E}(M_t) + \sum_{x\prime\in X}\epsilon_{x\longrightarrow \acute{x}}^a\left(\frac{1}{T}R_{x\longrightarrow \acute{x}}^a + \frac{T-1}{T}V_{T-1}^\pi(x\prime)\right). \tag{14}$$

4.3. Bootstrap Re-Sampling Controller for Mental Rehearsal. Bootstrapping was introduced as a flexible method to estimate the sampling distribution of an independent observation function [40]. It takes distribution F_n from sample data to substitute the global whole data, $R^*(x^*, F_n) = \widehat{\theta}(F_n^*) - \widehat{\theta}(F_n)$, and is useful for estimating of uncertainty in subspace identification. Figure 4 shows a segment that describes bootstrap re-sampling training to search the optimal transition probability where the bootstrap re-sampling controller K_Λ is used to determine the transition possibility.

We have $\mathscr{N}^\dagger \triangleq \{b_1 \cup b_2 \cup \cdots \cup b_k\}$ where the subscript k stands for the different sample index. To explore the theoretically infinite solution decision space $\mathscr{N}^*$, we assume that its probability distribution function is in accordance with $f(x, \theta)$. For each $\mathscr{N}^\dagger$, it stands for a brevity decision

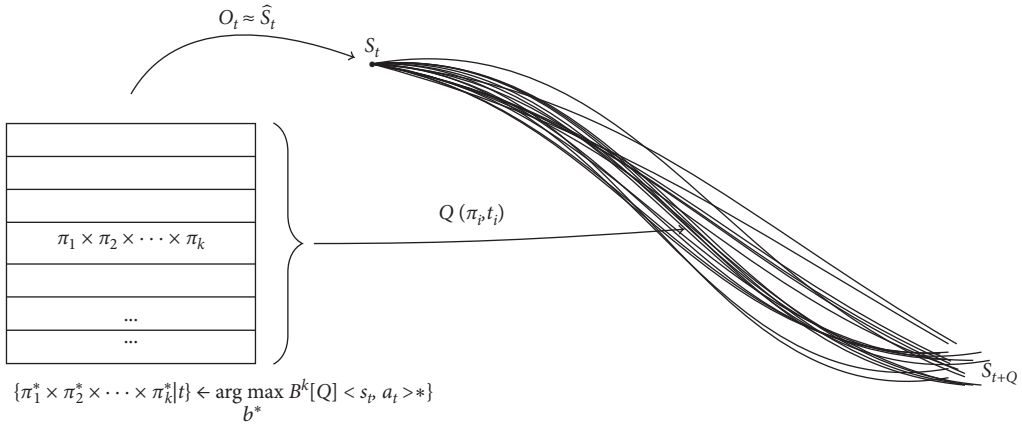


Figure 4: Bootstrap re-sampling controller K_Λ for transition probability through different mental rehearsal frequencies. The set of curves denotes the machine state composed of different mental rehearsal frequencies. Correspondingly, the left part is the human decision between future estimation step length Q.

sequence according to Q-learning estimation introduced in the next section. We use the limited re-sampling characteristic $\widehat{f}$ to represent the global big sampling range.

$$f(< s_t, a_t >, b) = \widehat{f}(< s_t, a_t >^*, b^*). \tag{15}$$

Here $\langle s_t, a_t \rangle$ is the state from observed aspect and decision in space $\mathcal{N}$; therein, $s_t \triangleq \mathrm{diag}(s_t^{l=1}, s_t^{l=2})$. The subscript t is the discrete step index in decision sequence. The superscript l stands for different observed states, such as human state and machine state.

We assume that transition probability discrete distribution Pr dominates the pair $\langle s_t, a_t \rangle$ transferring in decision sequence. To calculate $\Pr_t(\cdot | s_{t-1}, a_{t-1}, \ldots, a_{t-S}) \triangleq [\widehat{p}_t^1, p_t^2, \ldots, p_t^{\|r\|}]^T$, we first set initial probability $\Pr_0 = [p_0^1, p_0^2, \ldots, p_0^{\|r\|}]^T$. The capital subscript S stands for step length of history consideration decision. The superscript $\|r\|$ is the account of decision category in that different kinds of decisions are independently identically distributed. Next, each probability element is calculated by

$$p_t^j = g(\Pr_0 \Pr_t(s)) = \Pr_0 \frac{\sum 1_{p_t^j}}{\sum p_t^j}. \tag{16}$$

Here $1_{p_j^i}$ is an indicator function.

$$1_{p_t^j} = \begin{cases} 1, & \text{if } j \in [1, \|r\|], \\ 0, & \text{if } j \in [1, \|r\|]. \end{cases} \tag{17}$$

Furthermore, the searched optimal transition probability group $\mathbf{P}$ will be determined for searching contraction Ω. In bootstrapped sample k for future estimation, reward function r, which is assumed monotonically increasing as performance improves, is used to compare the accumulated reward.

$$\mathscr{B}^k[Q] < s_t, a_t >^* = \max_{a_{t-S}, \ldots, a_t} \left\{ r\langle s_{t-S}, a_{t-S} \rangle^* + \cdots + \gamma^{t-1} r\langle s_{t-1}, a_{t-1} \rangle^* + \gamma^t Q\langle s_t, a_t \rangle^* \right\}_k, \tag{18}$$

where γ is damping factor and Q is a Q-learning estimation from time i. After given the bootstrap sample estimation value, we derive the transition probability distribution and get the decision a_{t+1} from the optimal decision trail $\{b^*\}$ for the next decision step. Algorithm 1 shows the comprehensive block of searching contraction Ω.

$$\mathbf{P}_Q^k(s_t) = \left\{ \widehat{b}^*{}_{\pi_{i=1}} | \widehat{b}^* = \arg\max_{b^*} \mathscr{B}^k[Q]\langle s_t, a_t \rangle^* \right\}. \tag{19}$$

4.4. Decision Space Boundary for Admissible Decision. The decision space boundary block Δ in Figure 2 limits the accessible decision action set and relies on the airplane flight dynamic state variables. Based on the observed state $s_t^{l=2}$, admissible decision action a_t is the subset of action control scope $a_t = [{}_t^1, \ldots, {}_t^l]^T$. Meanwhile, allowable state space $\mathbb{S} = \widehat{s}_t$ is a hypercube field $\{\min V_1, \max V_1\} \times \cdots \times \{\min V_{r'}, \max V_{r'}\}$, which will inversely limit the decision action generated from the searching contraction method. This self-triggered policy caused by interaction object contributes to jumpy updating state and executing action by relying on the latest sampled state information [41]. In the sequel, we consider the flight longitudinal dynamic system V model as follows:

$$\begin{aligned} \dot{V}_1 &= \frac{1}{m}(T\cos(V_4) - D(V_4, V_3, \delta e) - G\sin V_2) \\ \dot{V}_2 &= \frac{1}{mv}(T\sin V_4 + L(V_4, V_3) - G\cos V_2) \\ \dot{V}_3 &= \frac{M(V_4, V_3, \delta e)}{I_y} \\ \dot{V}_4 &= V_3 - \dot{V}_2. \end{aligned} \tag{20}$$

(1) Initialize $\alpha, \epsilon, v, \delta t$, and $t \in T, a \in A, s \in S$
(2) Initialize $T, S, A, \widehat{S}, S_{\text{step}}, Q_{\text{step}}$
(3) **repeat**
(4) Calculate history reward $\mathbb{E}(M_t)$
(5) Initialize the boots frequency k
(6) **repeat**
(7) Calculate predicted reward $Q_T^{\pi}(x, a)$
(8) **until** Q_{step_j} and b = boots
(9) Calculate $\frac{j}{t}$
(10) Calculate $\mathscr{B}^k[Q]\langle s_t, a_t\rangle^*$
(11) **until** Temporal fragment decision completed at t
(12) Calculate $\mathbf{P}_Q^k(s_t)$
(13) Obtain the possible decision π^*

Algorithm 1: Searching contraction in space $\mathcal{N}^{\dagger}$.

$\dot{V}_1, \dot{V}_2, \dot{V}_3, \dot{V}_4$ are the various rates of airspeed, vertical speed, attack angle, and pitch angular velocity. $M(\cdot), I_y$ is the rotational inertia, T is the power of airplane, and δe is the engine mounting angle. We assume that airspeed is approximately equal to tangential velocity. To get the boundary block, the Hamilton–Jacobi function is needed to be solved as follows:

$$-\frac{\partial \mathbb{S}(a_t, t)}{\partial t} = \min\left\{0, H^*\left(a_t, \frac{\partial \mathbb{S}(a_t, t)}{\partial a_t}\right)\right\}. \tag{21}$$

Therein $H^* = \sup\{\nabla V^*\}^T V(a, s_t^{l=2})$, and the terminal boundary condition is set as $\mathbb{S} = \min\{\max\ V_1 - V_1, V_1 - \min V_1, \ldots,\ \max V_{r_l} - V_{r_l}, V_{r_l} - \min V_{r_l}\}$. Then, we can obtain the effective action u^* which stays in the scope of decision space boundary.

$$u* = \arg\sup_{a_t \in a_t} H\left(a_t, \mathbb{S}, s_t^{l=2}\right). \tag{22}$$

At each decision time updated, the bounded observable state $\mathbb{S}(_t, t)$ will be generated and compared with the boundary $\mathbb{S}$. Only until the decision $\pi^* \in u^*$, π can be transmitted into airplane simulating block and then the interaction can be completed. Below, we provide Algorithm 2 for the whole decision inference on the receding horizon.

(1) Initial target set Θ
(2) **repeat**
(3) Calculate the allowable state space $\mathbb{S}$
(4) **repeat**
(5) searching contraction Ω block
(6) **if** $\pi \in u^* = \text{argsup}_{_t \in a_t} H(_t, \mathbb{S}, s_t^{l=2})$ **then**
(7) output decision π
(8) **end if**
(9) **until** π
(10) **until** target set is completed

Algorithm 2: Targeted decision inference on receding horizon.

5. Experiment and Results

In this section, we present an experimental case and its results for airplane manipulating scenario. A typical task is to manually control the aircraft to descend altitude by the pilot. Some extra tasks are set on the designated altitude. Those particular subtasks require pilots to execute special operations. We apply our method into this experiment case. Results show that our method reflects the cognitive searching optimization from searching contraction index.

5.1. Experiment Setting. Table 1 shows the flight altitude descending stages from 11000 ft to 2000 ft in experiment. Stages 2, 4, and 6 require the pilot to complete the specific tasks on specified altitude scope, and stages 1, 3, and 5 are the normal descending procedures involving basic flight joystick and throttle control [42]. Table 2 lists the multiresource channels involved in the experiment. We calculate the situation channels occupied with equipment to assess the workload taken by action. Table 3 lists the correlation between control rules and related equipment. Table 4 points out that delay in decision process separately represents state interval time delay and action time delay. We use the Poisson distribution and normal distribution to represent those two types of delay. The Poisson distribution is a non-exponentiation distribution satisfying the semi-Markov property in the context of continuous time domain. μ and σ stand for the mean value and variance value, respectively [44, 45].

In experiments, we use the control rule to substitute the action in traditional decision action set. Each control rule corresponds to specified control equipment, which is chosen depending on the interactive process $\widehat{M}_t$. The number of actions can be one or more depending on observed states. The parameter mean in normal distribution of action time delay relates to the different equipment. Here we adopt parameters from the NASA timeline analysis report [46].

5.2. Simulation Results. In order to assess the pilot's cognitive decision ability, flight performance, human accumulated workload, and the number of manipulations are three indexes measuring our multistep decision method. Trends of dynamical dimension in $\mathcal{N}^{\dagger}$ show the searching

Table 1: Flight stages and tasks.

Flight stage	Flight height (ft)	Maneuvering apparatus	Specific task
1	11000–7000	Normal descending	
2	7000–6000	Switch	De-icing switch
3	6000–3500	Normal descending	
4	3500–3300	Communication	Tower communication
5	3300–3000	Normal descending	
6	3000–2000	Flaps	Adjust flaps

Table 2: Multiresource workload channels and weights [43].

Channel	Visual	Auditory	Balancing	Hand	Foot	Analysis
Weight	0.2	0.2	0.1	0.2	0.1	0.2

Table 3: Control rule, maneuvering apparatus, and channel occupation.

Location	Control rules (ru)	Maneuvering apparatus	Channel occupation
Rule 1	Pitch control	Elevator, throttle	1 0 1 1 0 1
Rule 2	Vertical-speed control	Elevator	1 0 1 1 0 1
Rule 3	Height control	Elevator	1 0 1 0 0 1
Rule 4	Configuration control	Flaps	1 0 1 0 0 1
Rule 5	Dynamical control	Throttle	1 0 1 0 0 1
Rule 6	Information obtain	Observe, scan, microphone	1 1 0 0 0 1
Rule 7	Button control	Switch	1 0 1 0 0 1
Rule 8	Idle control	Keep	1 1 1 0 0 0

Table 4: Transition interval and action time delay.

Series	Distribution	Parameter
State transition interval	Poisson distribution	$\mu = 3$
Action time delay	Normal distribution	$\mu = \{1.02, \ldots,\}, \sigma^2 = 0.025$
Action accuracy error	Normal distribution	$\mu = 1, \sigma^2 = 0.05$
Action transition error	Normal distribution	$\mu = 1, \sigma^2 = 0.05$

contraction results from infinity to limited. We use inner batch frequency to simulate the mental rehearsal in which the pilot chooses the suitable action from anticipating. Here inner batch frequency refers to rehearse times. By varying SMDP-step (semi-Markov decision property) and Q-step tuple, we simulate the time scale of history state influence and future prediction impact. Table 5 lists the main experiment results, and it shows the detailed contrast data in different combinations of parameters. Figure 5 shows flight height performance. Figure 6 depicts the workload accumulated speed increased with working time under different parameter configurations. Figure 7 shows the value of searching contraction ratio.

In Table 5, time of flight task, human accumulated workload, and steps of manipulation are three indexes to build the overall evaluation assessing the decision methods. The accumulated workload caused increases slightly when the parameter inner batch frequency is boosted. This result is consistent with the fact that human mental workload increased with cognitive time pressure [47]. The bigger batch frequency is, the more the time pressure is. It is worth noting that our experiment setting batch = 20 is an extreme situation, exceeding the ordinary [48]. Regarding the human cognitive, their capability enlarges as batch frequency increases. The compound step tuple is another critical parameter. By setting history consideration steps (SMDP-step) and future estimation steps (Q-step), we compose the different multistep decision methods. For example, when the SMDP-step and Q-step are equal to 1, the method is essentially a Markov decision process (MDP).

Figure 5 intuitively reflects airplane flight height effect. On the whole, descending trajectories show the optimal stationary distribution at batch = 5, where curve differences are less. The differences between different trajectories, decided by steps tuple, are significantly increased. This result illustrates that searching contraction is relevant to human mind rehearsing action. The bigger the rehearsing frequency, the more the difference caused by different multistep decision methods. From the view of trajectory smoothness, the descending trend is similar to flight stage 1. But the accumulated effect caused by different multistep decision methods starts to appear from stage 2. In Figure 5(a),

Table 5: Different decision method parameters in simulation.

Batch frequency	Method	SMDP-step	Q-step	Time (seconds)	Accumulated workload	Decision steps
Batch = 1	MDP (S1Q1)	1	1	256.48	223.99	69
	Q-learning (S1Q3)	1	3	238.14	205.81	65
	SMDP (S3Q1)	3	1	246.45	182.87	65
	S3Q4	3	4	211.52	191.93	57
	S4Q3	4	3	264.69	243.82	73
	S8Q3	8	3	213.46	202.12	61
	S3Q8	3	8	218.93	208.32	58
Batch = 5	MDP (S1Q1)	1	1	222.58	230.66	49
	Q-learning (S1Q3)	1	3	222.11	224.80	52
	SMDP (S3Q1)	3	1	218.79	181.34	43
	S3Q4	3	4	243.79	265.43	59
	S4Q3	4	3	242.76	215.07	52
	S8Q3	8	3	218.75	230.78	46
	S3Q8	3	8	229.04	230.62	51
Batch = 10	MDP (S1Q1)	1	1	216.87	225.46	47
	Q-learning (S1Q3)	1	3	257.72	218.67	51
	SMDP (S3Q1)	3	1	234.12	244.90	52
	S3Q4	3	4	256.71	258.67	63
	S4Q3	4	3	238.40	215.03	50
	S8Q3	8	3	230.78	220.35	53
	S3Q8	3	8	268.97	264.86	64
Batch = 20	MDP (S1Q1)	1	1	264.65	247.43	54
	Q-learning (S1Q3)	1	3	287.11	289.09	63
	SMDP (S3Q1)	3	1	262.07	236.37	53
	S3Q4	3	4	266.08	231.41	53
	S4Q3	4	3	258.91	239.23	55
	S8Q3	8	3	253.59	228.45	55
	S3Q8	3	8	236.54	228.89	57

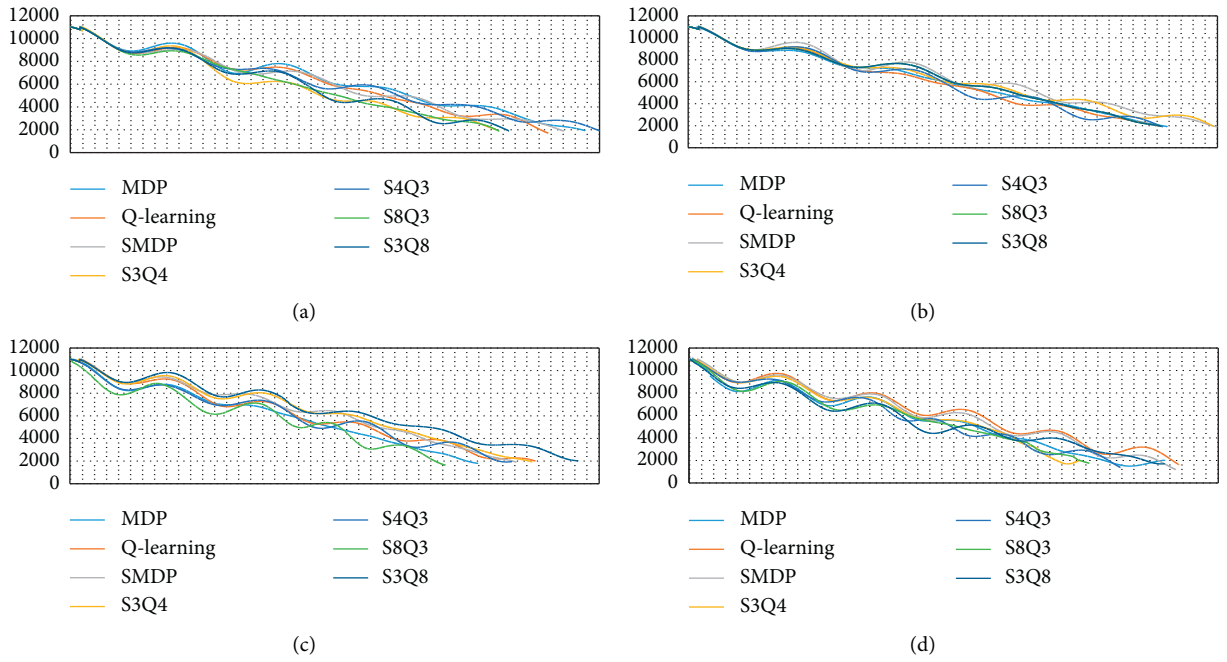


Figure 5: Airplane descending performance. The vertical axis represents the corresponding airplane height/ft, and the horizontal axis represents the horizontal flight distance/ft. (a) Batch = 1. (b) Batch = 5. (c) Batch = 10. (d) Batch = 20.

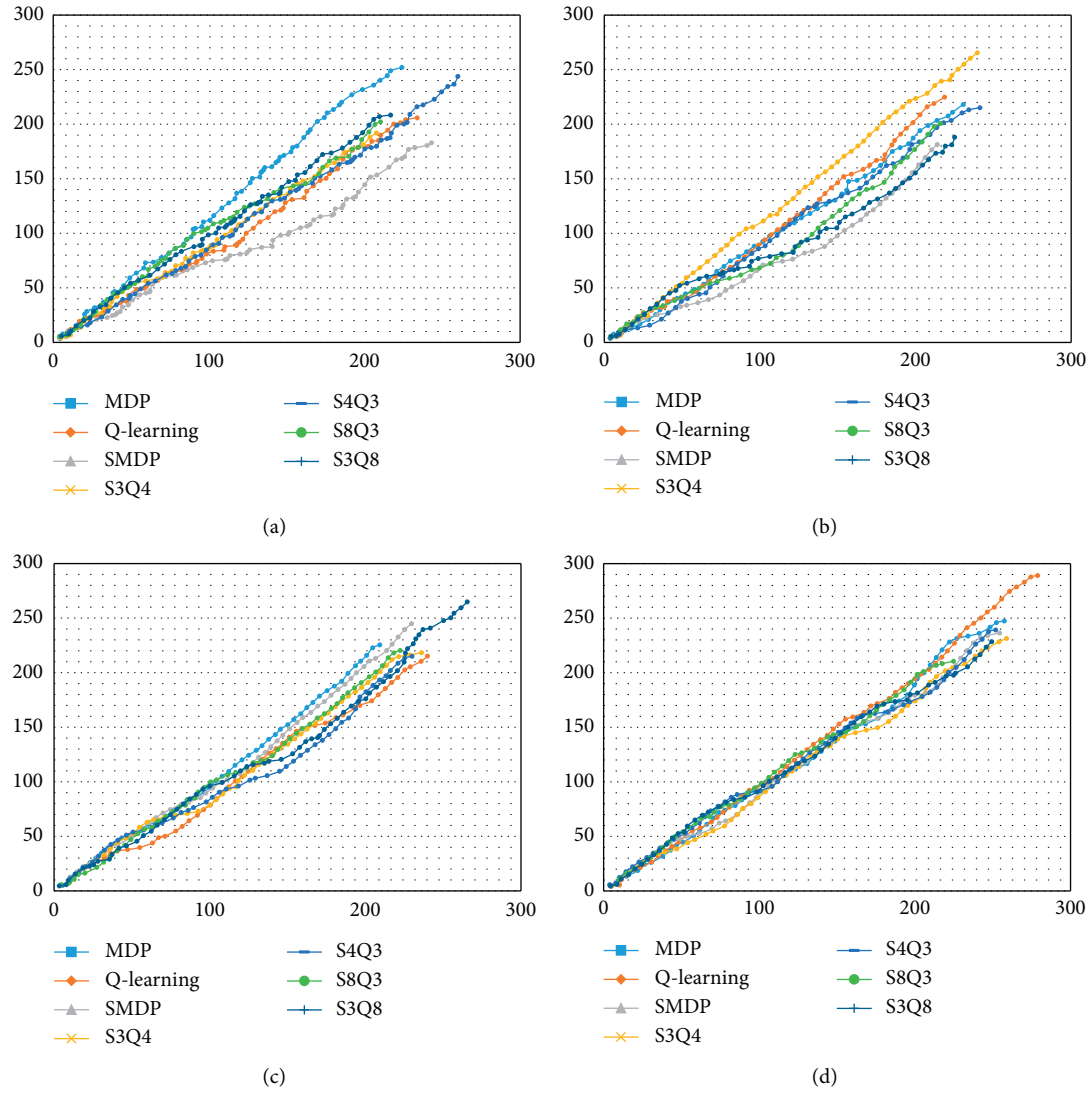


Figure 6: Value of continuously accumulated cost increased with working time. The horizontal axis represents the variable working time, and the vertical axis represents the reward value. (a) Batch = 1. (b) Batch = 5. (c) Batch = 10. (d) Batch = 20.

fluctuating range of S8Q3 is flatter compared with others. In Figure 5(b), fluctuating range of S3Q8 is flatter compared with others. In Figure 5(c), fluctuating range of MDP is flatter compared with others. In Figure 5(d), fluctuating range of S8Q3 is flatter compared with others. Corresponding to the rough scope of higher local value in each subfigure, flight stages 2, 4, and 6 which cover more control tasks show the hysteresis effect in descending trend curves. Also, it can be found that all multistep decision methods in experiments can converge airplane state to the target position.

Figure 6 shows a composite reward calculated by airplane flight performance and human cognitive workload performance. Reward value is more uniformly distributed when batch frequency is bigger. When human makes decisions after the repeated estimations, reward caused by manipulation tends to be similar. But higher repeated estimation times bring higher reward value, meaning that excessive anticipation leads to excessive cognitive workload. The reward accumulated speed increases with the batch frequency based on the slope of curves. Aside from Figure 6(d), the green line (S8Q3) and blue line (S4Q3) state the superior result from the horizontal axis (time, less is better) and vertical axis (reward value, less is better). The gray line (SMDP) at batch = 5 takes the best effect, which means that the multiple estimations also take effects on future estimation. The deepskyblue line (MDP) at batch = 10 takes the best effect, while the yellow line (S3Q4), green line (S8Q3), and the blue line (S4Q3) present proximate effect.

Figure 7 shows dimension variation percentage of searching contraction ratio. We calculate the ratio μ by the accumulated searching result of history decision space and predetermined searching scope at the parameter tuple (S_{step}, Q_{step}, batch or mental rehearsal frequency). The value of μ is smaller, and the searching contraction ratio is higher. Equation (23) calculates μ, where ru refers to the number of rules and 1_π is an indicator function.

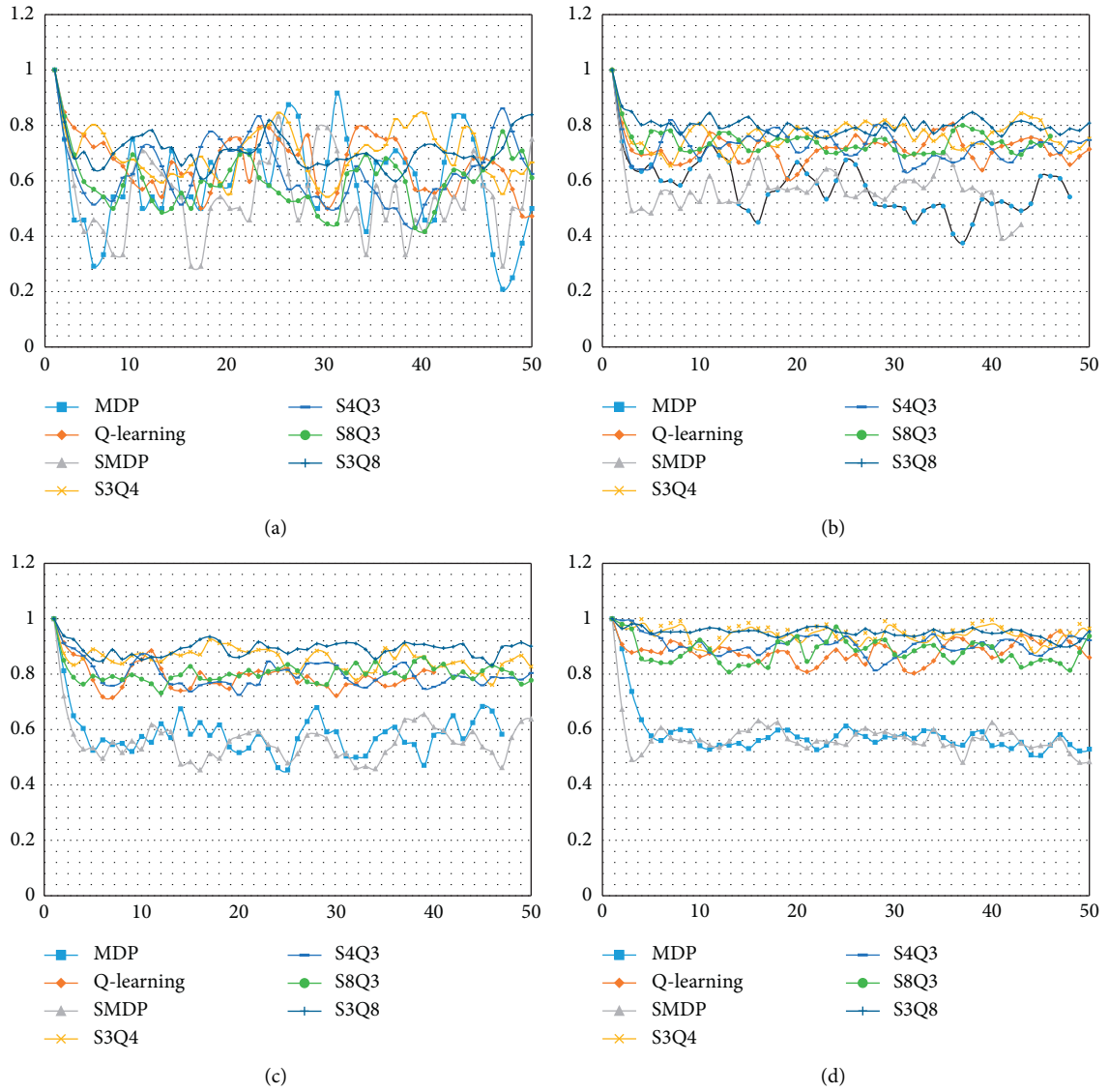
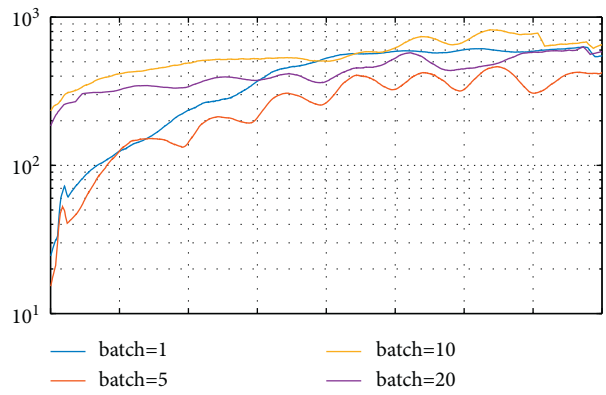


Figure 7: Dimension variation percentage for searching contraction in HC^{2}I process. The horizontal axis represents the serial number of decision steps, and the vertical axis represents the percentage of dimension contraction. In order to facilitate the comparison of values, we intercept the decision step number 50 as the maximum number of steps in the figure. (a) Batch = 1. (b) Batch = 5. (c) Batch = 10. (d) Batch = 20.

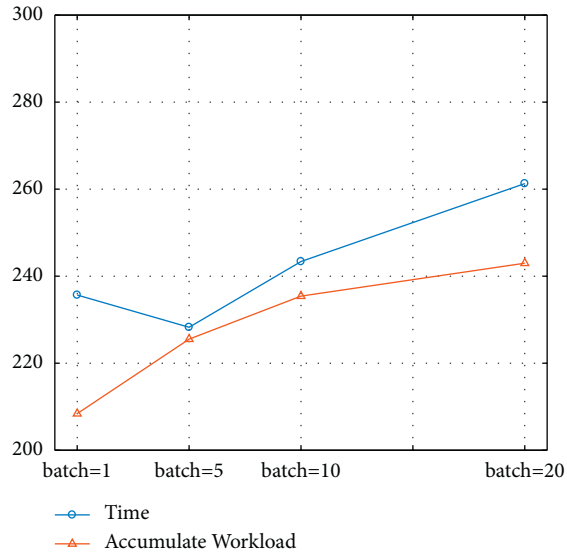
$$\mu = \frac{\sum 1_{\pi}}{\text{boots} \cdot \left(S_{\text{step}}^{\text{ru}} \cdot Q_{\text{step}}^{\text{ru}}\right)^{1/\gamma}} \times 100\%. \tag{23}$$

In this way, the initial infinity searching dimension is related to the history step influence, future step estimation, and bootstrapping frequency. The probability determined by the mental rehearsal and multistep transition, which is influenced by the dimension of decision, reflects the contraction effect of the search dimension. As shown in the results, cognitive searching optimization process shows a downward trend overall. When the batch frequency increases, the stability of the searching dimension contraction ratio gradually improves. Also, the ratio is distinguished according to different types of combined decision steps. For example, when future estimation step Q_{step} equals 1, such as MDP (S1Q1) and SMDP (S3Q1), the longitudinal change amplitude of dimension percentage changes intuitively from big to small within the batch frequency but independent from other methods.

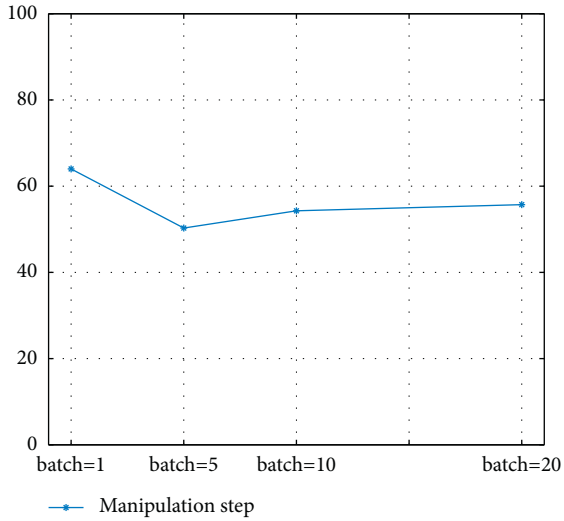
5.3. Cost and Performance Analysis. Figure 8 shows the changes in various indicators and three primary parameters (two types of decision steps, rehearsal frequency) in our decision-making method. Under different batch parameters, Figure 8(a) shows the variation of statistical standard deviation for each cluster's flight descent curve as the mission progresses. When the batch number is larger, the standard deviation firstly climbs up and then declines. Figure 8(b) shows that the average accumulated workload increases with the batch frequency. It proves that the more the decisions



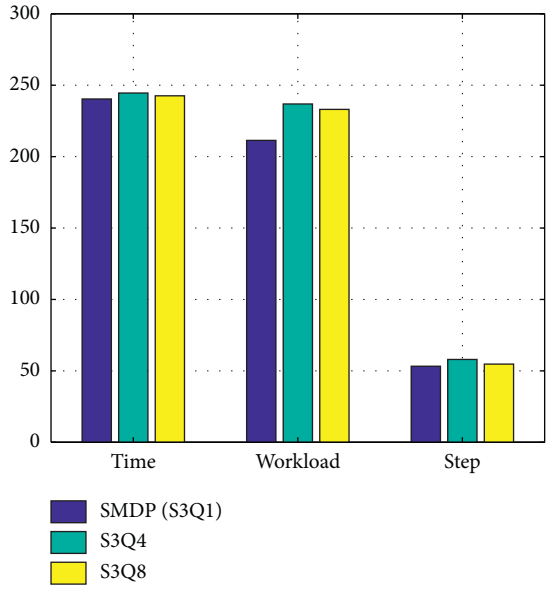
(a)



(b)



(c)



(d)

Figure 8: Continued.

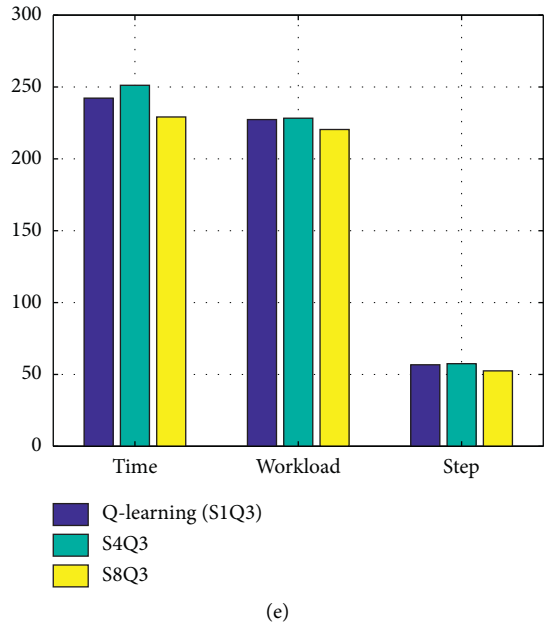


Figure 8: Analysis of flight altitude decline performance and reward index value. To compare the effects of different batches (mental rehearsal parameters), we calculate the overall mean values of flight height performance and reward indexes (time, accumulated workload, and decision manipulation steps) under different multistep transfer parameters. To compare the effect of multistep transition effect (the number of steps in historical consideration and the number of steps in future estimation) on the reward index, we calculate the overall mean value under different batch parameters. (a) Standard deviation of airplane performance for different batch frequencies. (b) Average cost index for different batch frequencies. (c) Average steps of manipulation for different batch frequencies. (d) Average steps of manipulation for different batch frequencies. (e) Average steps of manipulation for different batch frequencies.

people anticipate, the greater the workload caused by mental rehearsal. An inflection point exists in time index when batch = 5 shows that suitable mental rehearsal can decrease work time. Figure 8(c) shows that manipulation step drops down with batch increases, but its decline trend slows down. When the batch frequency is less (e.g., batch = 1), the manipulation step is more larger. The negative correlation can illustrate that non-optimal strategy step leads to generate more decision steps to revise the former. Figures 8(d) and 8(e) show the influence of different types of decision step. There is a peak in the histogram group under all of the different indicators showing that the appropriate number of decision steps can reduce the corresponding indicator's performance. Instead, the inappropriate number of decision step will increase the index value. Figure 9 analyzes the reward value presented in Figure 6. The combined reward index defined by flight performance and workload shows that batch 5 is the peak of these data, which shows that although the workload will increase, the overall value of batches 10 and 20 will decrease under the influence of the mission. Therefore, considering the three types of index data and reward values, the batch frequency between 1 and 5 is more appropriate.

5.4. Transition Probability and Searching Contraction Ratio Analysis. Figure 10 shows transition probability distribution varying in different types of methods. Transition probability, which is calculated from inner simulated estimation, reflects the dynamical selection from rules. The changing trend is more consistent with normal human decision-making behavior because inferring transition probability lies at the core of human sequence knowledge [49]. It demonstrates that cognitive interaction behavior constantly attempts to infer the time-varying matrix of transition probabilities when it receives the outer observed machine states. Therefore, dynamical transition probabilities are ensured by the bootstrap re-sampling controller in searching contraction method.

Additionally, Figure 10 shows transition probability with regard to control rules in Table 3 and parameters in Table 5. Transition probability is dynamically changeable during the flight manipulation stage. The transition probability value of pitch control rule (Rule 1) is higher than that of other rules on average. The transition probability of vertical-speed control rule (Rule 2) is less focused than the height control rule (Rule 3). Configuration control rule (Rule 4) is not used until the flight altitude attains the allowable range. At different flight stages, rule transition probability is verified by the specific tasks and its corresponding control rules. On the other hand, transition probability is prominently influenced by step tuple. The overall fluctuation of transition probabilities varies strengthening accompanied by the increase of estimated part in the decision step tuple. Fluctuation of probability variation in MDP and Q-learning methods is less than others.

On the other hand, searching contraction change happens in the continuously multistep decision HC^2I process. It denotes the damping of decision admissible exploitation dimension. The solution space in which human chooses

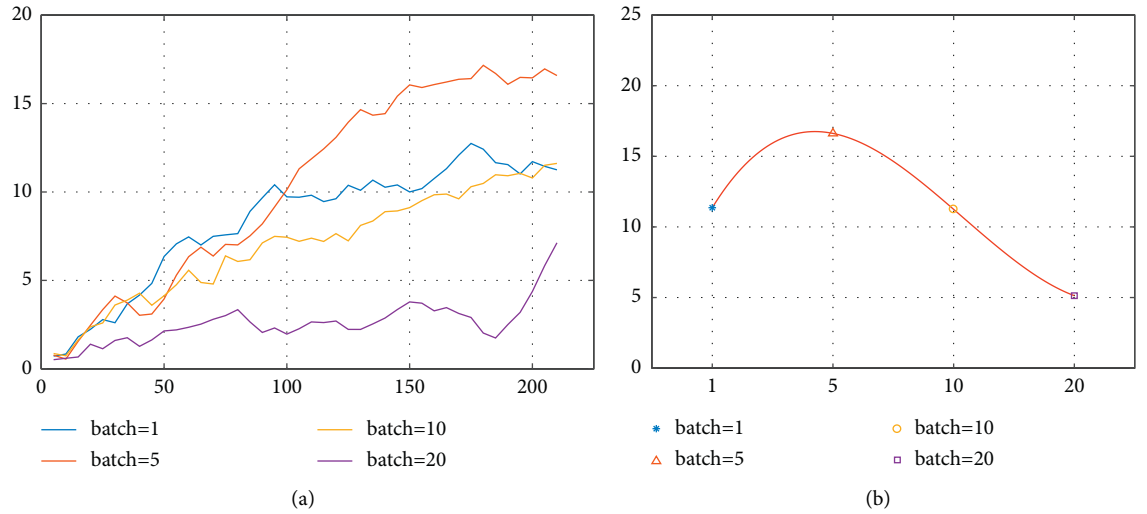


Figure 9: Analysis value of continuously accumulated reward. According to different batch parameters, we use the evolution curve of standard deviation to describe the overall fluctuation degree of each cluster curve. At the same time, a relatively stable standard deviation was selected between working time [150, 200], and a curve was fitted to reflect the change of the fluctuation degree of reward with the increase of the batch. (a) Standard deviation. (b) Fitting function.

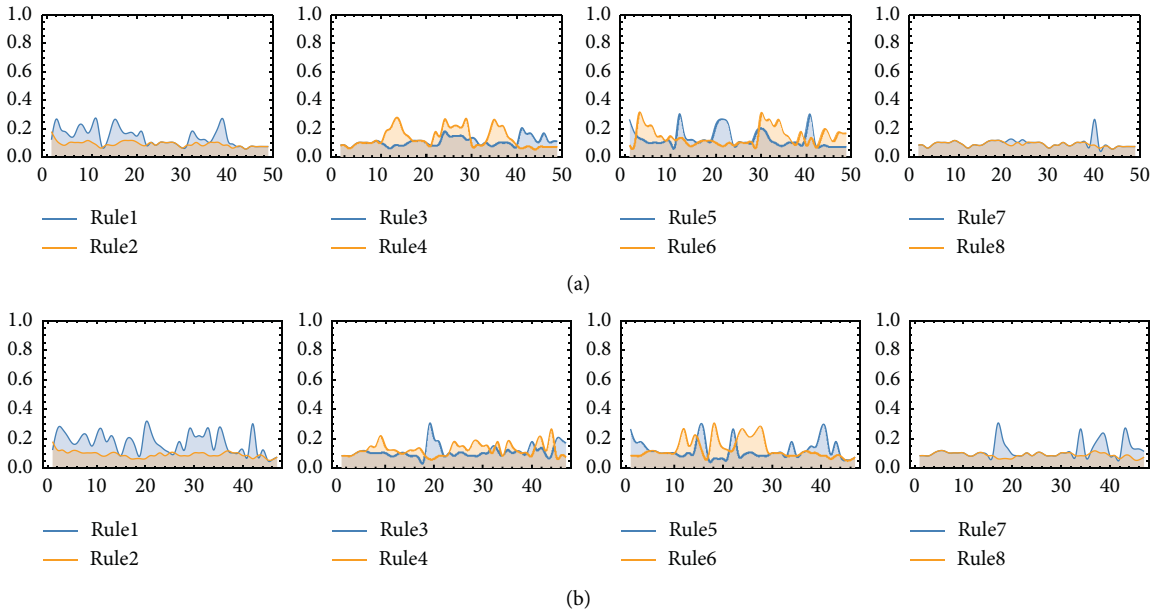


Figure 10: Transition probability evolving process for different control rules in (a) S4Q3, batch = 5, and (b) S4Q3, batch = 10. The horizontal axis represents the serial number of decision step, and the vertical axis represents the value of transition probability.

strategy rules contracts within the receding horizon controller K_Λ, after completing the inner inference by K_∇. Research about the brain prefontal also demonstrated this point that the existing high-level control area reprocesses the upcoming decision before it finally enters awareness [2]. Therefore, under the influence of interactive environment, the degree of searching contraction is a critical factor in cognitive decision.

We compare the searching contraction ratio from different aspects in Figure 11. First, stability of the dynamical dimension goes down as the batch frequency gradually goes up. Figure 11(a) compares the search contraction ratio in terms of the composition of multistep transfer steps and the frequency of mental previews. By calculating the average change of the overall contraction ratio under the specified batch parameters, the figure shows that under higher batch frequency situation, the contraction ratio fluctuation decreases when the contraction percentage increases. The intersection between the batch parameters 1–5 will provide better overall performance. The S4Q3 example in Figure 11(b) emphasizes that as the batch value increases, the amplitude of the contraction percentage fluctuation will

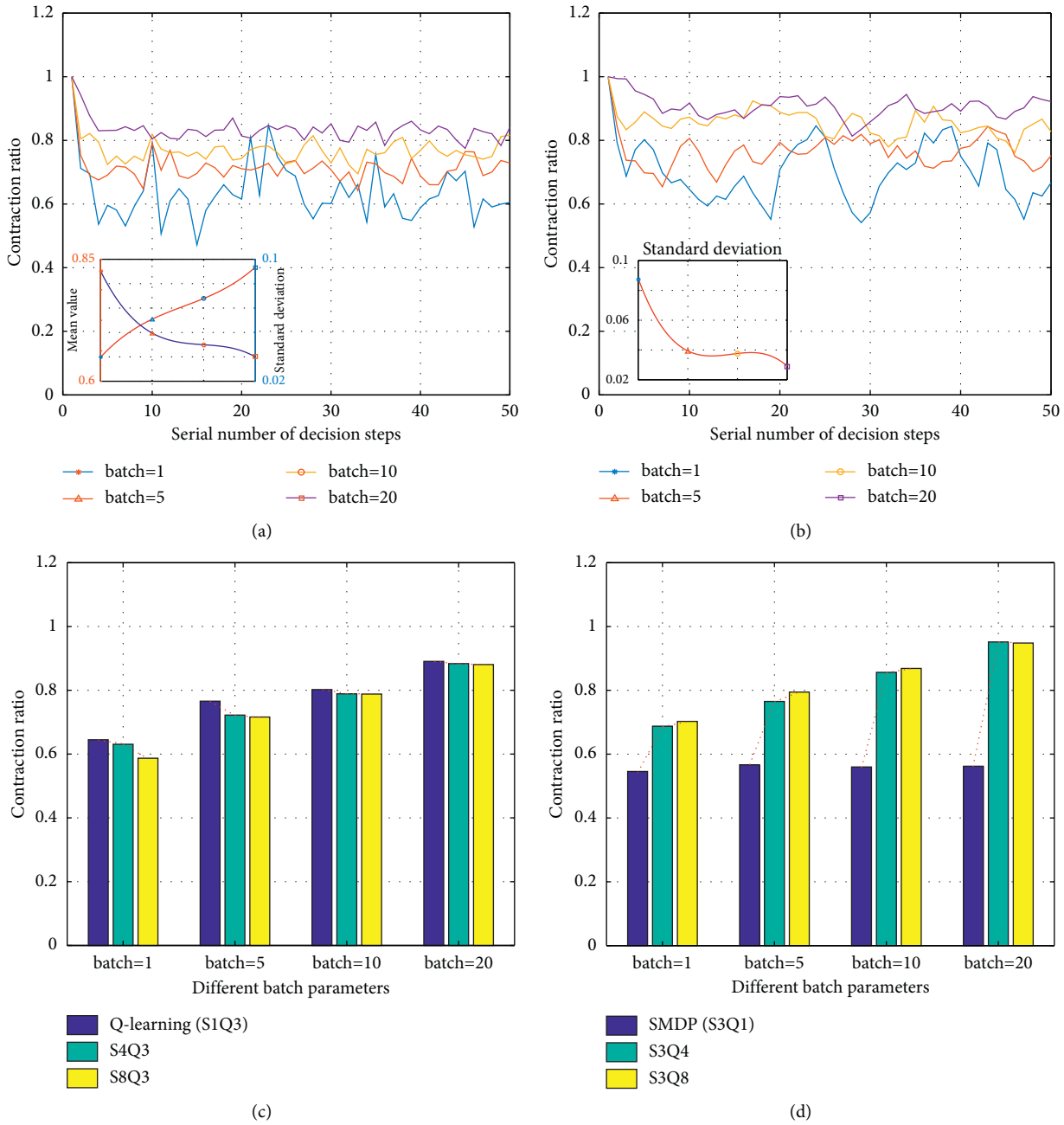


FIGURE 11: Sample comparison from four aspects. (a) Mean value comparison in different batch frequencies. (b) S4Q3 sample comparison in different batch frequencies. (c) History transition steps influence on searching contraction ratio. (d) Future estimation steps influence on searching contraction ratio. Fitted variation curves of standard deviation and mean value are also plotted. In (c) and (d), two variation trends of search contraction are plotted based on the different parameter values of the decision step number.

decrease, but when the number of batches is greater than 5, the standard deviation of the fluctuation amplitude tends to be flat. It demonstrates that the increase in batch frequency at this time has little effect on the increase of the fluctuation amplitude. Figures 11(c) and 11(d) compare the number's increase of history consideration steps (S_{step}) and future estimated step (Q_{step}) groups. The result in this figure states that during the HC^{2}I process, the contraction ratio caused by the history consideration steps and future estimated step in the experiment decreases and increases, respectively, with the number of steps increasing. But, they both indicate that too many steps will suppress the improvement of contraction ratio (e.g., the number of steps equals 8). And with the increase in the batch frequency, increment in the contraction ratio will gradually slow down. This result shows the fact that searching contraction is difficult for human brain under the situation of the excessive rehearsal numbers, such as batch = 20. The above data analysis explains the feasibility of searching contraction decision method proposed in this paper.

Table 6: Comparison of different methods.

Rand	Methods	Index: searching contraction		Index: time
		Mean value	Variation	
1	MDP	0.605	0.0250	256.48
2	SMDP	0.546	0.0091	246.45
3	Q-learning	0.645	0.0182	238.14
4	B5, S3Q4	0.765	0.0015	243.79
5	B5, S3Q8	0.794	0.0005	229.04
6	B10, S3Q1	0.549	0.0031	234.12

6. Discussion

Our experimental results illustrate that cognitive search contraction is a subconscious phenomenon that commonly occurs in the decision-making process. Therein, the associated multistep transition and mental rehearsal are two crucial factors. The multistep transition factor, which combines the influence of cumulative fragments and the jumpy transition interval, is the basis of interactive decision making. Under the credible admissible decision-making boundary, the mental rehearsal that covers the parallel execution of the decision fragments will screen the decision again until obtaining the optimal decision-making behavior.

The analysis of experimental results shows that different combinations of multistep transition step values and different rehearsal frequencies will affect the search contraction ratio and decision-making rewards. On the one hand, the number of historically considering decision steps will reduce the workload of brain (by increasing the searching contraction ratio). In contrast, the increase in the number of future estimated steps will increase the workload of brain (by decreasing the contraction ratio of searching). At the same time, too much rehearsal frequency makes the contraction efficiency of decision search in decision solution space decrease. This is consistent with the research result that cognitive channel is limited when human completes multitask [50].

Meanwhile, the increase in the rehearsal frequency will reduce the fluctuation of search contraction index. It shows that the if a person can get more information or experience before decision, the bias of result will lower. On the other hand, the decision reward of all different multistep transition step types presents a trend of increasing first and then decreasing. This reflects that the number of decision-making steps are not positively correlated with the decision reward. Moderate composition of multistep transition step can bring the optimal decision reward (interval time, workload, and number of decision steps). It is the same as the rehearsal frequency.

Here we define the time complexity as the sum of the worst-case running time for each operation (e.g., multiplication, division, and addition) required to process an output. The growth rate is then obtained by making the parameters of the worst-case complexity tend to infinity. Memory complexity is estimated as the number of 32 bit registers needed during the learning process to store variables. And also, only the required worst-case memory space is considered during the process phase. From the following complexity formulations, we can find that the time complexity grows quadratically with N and linearly with T, and its memory complexity of the algorithm grows linearly with T and N. When T is very large, the memory complexity will not exceed the resources available for the training process, avoiding overflow from internal system memory to disk storage.

(i) Time complexity: for a decision sequence with length T and ergodic Markov state N, the time complexity is composed of decision iteration process. We can get the time complexity as

$$\begin{aligned}\mathcal{O}(T) &= T * \left(S_{\text{step}} + \text{boots} * N * \log_m N\right) * Q_{\text{step}} \\ &\approx T * N * \log_m N.\end{aligned} \tag{24}$$

(iii) Memory complexity: the memory complexity relates to the decision space. In this paper, the decision space is composed of decision horizon space and bootstrap sampling space.

$$\begin{aligned}\mathcal{O}(S) &= T * \left(\text{boots} * \left(N_S + N_A + N_O\right)\right)^{Q_{\text{step}}} \\ &\approx T * N^{Q_{\text{step}}}.\end{aligned} \tag{25}$$

Research studies about decision tree considered the different algorithms to optimize the decision searching tree, such as the UCB1, UCT, and other non-greedy methods. However, the main index searching contraction ratio designed in our paper is not similar to those studies, where they verify their efficiency through the true/false ratio. On the other hand, the method in our paper used the Hamilton function to limit the admissible decision action while the other existing studies analyzed the decision state as a discrete classification problem. In Table 5, when parameters B (boots), S (S_{step}), and Q (Q_{step}) are different, our proposed method can be transformed into other existing methods, such as the standard Markov decision method (B1S1Q1), standard semi-Markov decision method (B1S3Q1), and standard Q-learning method (B1S1Q3). In Table 6, we add the comparison at the index searching contraction from its mean value and variation, and the index time is also listed. From the table, it can be shown that the stability and time efficiency of our proposed method are better than those of the previous studies.

7. Conclusion

In this paper, we propose a semi-Markov jump decision method to optimize the decision path and a searching contraction index to indicate cognitive searching optimization. The main difference of our work is using jumpy decision interval and multiple path estimation as deterministic features. Under the specific interaction target in decision boundary, this modification leads to optimizing cognitive interaction decision from the time dimension and depth dimension. Semi-Markov jump transition decision outperforms the traditional Markov method by strengthening the correlation from the dimension of decision time interval. The mental rehearsal improves the searching depth of decision solution space. The decision boundary filters out the infeasible human decision by the estimated admissible action boundary. Furthermore, numerical simulation shows the characteristic of searching contraction, and our decision method can be applied to evaluate a class of multiple element types' decision path. The reduction in searching contraction ratio proves that proper transition step length and mental rehearsal frequency can reduce and stabilize the searching space and reward of decision path in the HC^2I process. The future work will address the decision switch relation happening in the semi-Markov cognitive decision. To investigate the human fatigue influence on control accuracy and stationary, we will research the jumpy switch control according to the limited human behaviour rule. And the arbitrary number of historical decision steps in the decision-making is also deserved to be explored.

Data Availability

The numerical simulation data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] C. M. Wu, E. Schulz, M. Speekenbrink, J. D. Nelson, and B. Meder, "Generalization guides human exploration in vast decision spaces," *Nature human behaviour*, vol. 2, no. 12, pp. 915–924, 2018.

[2] C. S. Soon, M. Brass, H.-J. Heinze, and J.-D. Haynes, "Unconscious determinants of free decisions in the human brain," *Nature Neuroscience*, vol. 11, no. 5, pp. 543–545, 2008.

[3] H. G. Stassen, G. Johannsen, and N. Moray, "Internal representation, internal model, human performance model and mental workload," *Automatica*, vol. 26, no. 4, pp. 811–820, 1990.

[4] J. R. Larson Jr., "The performance feedback process: a preliminary model," *Organizational Behavior and Human Performance*, vol. 33, no. 1, pp. 42–76, 1984.

[5] T. L. Griffiths, F. Lieder, and N. D. Goodman, "Rational use of cognitive resources: levels of analysis between the computational and the algorithmic," *Topics in cognitive science*, vol. 7, no. 2, pp. 217–229, 2015.

[6] H. S. Chang, H.-G. Lee, M. C. Fu, and S. I. Marcus, "Evolutionary policy iteration for solving markov decision processes," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1804–1808, 2005.

[7] T. Jaakkola, S. P. Singh, and M. I. Jordan, "Reinforcement learning algorithm for partially observable markov decision problems," in *Advances in Neural Information Processing Systems*, pp. 345–352, Springer, Berlin, Germany, 1995.

[8] J. R. Busemeyer and T. J. Pleskac, "Theoretical tools for understanding and aiding dynamic decision making," *Journal of Mathematical Psychology*, vol. 53, no. 3, pp. 126–138, 2009.

[9] C. Buc Calderon, M. Dewulf, W. Gevers, and T. Verguts, "Continuous track paths reveal additive evidence integration in multistep decision making," *Proceedings of the National Academy of Sciences*, vol. 114, no. 40, pp. 10 618–10623, 2017.

[10] M. van Otterlo and M. Wiering, "Reinforcement learning and markov decision processes," in *Reinforcement Learning*, pp. 3–42, Springer, Berlin, Germany, 2012.

[11] Z. Xie and Y. Jin, "An extended reinforcement learning framework to model cognitive development with enactive pattern representation," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 3, pp. 738–750, 2018.

[12] J.-S. Song and X.-H. Chang, "H∞ controller design of networked control systems with a new quantization structure," *Applied Mathematics and Computation*, vol. 376, Article ID 125070, 2020.

[13] P. Dayan and N. D. Daw, "Decision theory, reinforcement learning, and the brain," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 8, no. 4, pp. 429–453, 2008.

[14] M. Lebreton, R. Abitbol, J. Daunizeau, and M. Pessiglione, "Automatic integration of confidence in the brain valuation signal," *Nature Neuroscience*, vol. 18, pp. 1159–1167, 2015.

[15] H. A. Yanco and J. Drury, "Classifying human-robot interaction: an updated taxonomy,"vol. 3, pp. 2841–2846, in *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 3, pp. 2841–2846, IEEE, Delft, The Netherlands, 2004.

[16] R. Moratz, K. Fischer, and T. Tenbrink, "Cognitive modeling of spatial reference for human-robot interaction," *International Journal on Artificial Intelligence Tools*, vol. 10, no. 4, pp. 589–611, 2001.

[17] B. A. Purcell and R. Kiani, "Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy," *Proceedings of the National Academy of Sciences*, vol. 113, no. 31, pp. E4531–E4540, 2016.

[18] B. Wu, L. Cui, and C. Fang, "Reliability analysis of semi-markov systems with restriction on transition times," *Reliability Engineering & System Safety*, vol. 190, Article ID 106516, 2019.

[19] L. Jones and G. Stuth, "The uses of mental imagery in athletics: an overview," *Applied and Preventive Psychology*, vol. 6, no. 2, pp. 101–115, 1997.

[20] R. Miranda, M. Weierich, V. Khait, J. Jurska, and S. M. Andersen, "Induced optimism as mental rehearsal to decrease depressive predictive certainty," *Behaviour Research and Therapy*, vol. 90, pp. 1–8, 2017.

[21] J. Ignacio, A. Scherpbier, D. Dolmans, J. J. Rethans, and S. Y. Liaw, "Mental rehearsal strategy for stress management and performance in simulations," *Clinical Simulation in Nursing*, vol. 13, no. 7, pp. 295–302, 2017.

[22] H. Su, W. Qi, Y. Hu, H. R. Karimi, G. Ferrigno, and E. De Momi, "An incremental learning framework for human-like redundancy optimization of anthropomorphic

manipulators," *IEEE Transactions on Industrial Informatics*, 2020.

[23] K. Oberauer, "Is rehearsal an effective maintenance strategy for working memory?" *Trends in Cognitive Sciences*, vol. 23, no. 9, 2019.

[24] K. Oberauer and S. Lewandowsky, "Modeling working memory: a computational implementation of the time-based resource-sharing theory," *Psychonomic Bulletin & Review*, vol. 18, no. 1, pp. 10–45, 2011.

[25] M. Mazher, A. A. Aziz, and A. S. Malik, "Evaluation of rehearsal effects of multimedia content based on EEG using machine learning algorithms," in *Proceedings of the 2016 6th International Conference on Intelligent and Advanced Systems (ICIAS)*, pp. 1–6, Kuala Lumpur, Malaysia, 2016.

[26] T. T. Hills, P. M. Todd, D. Lazer, A. D. Redish, I. D. Couzin, and C. S. R. Group, "Exploration versus exploitation in space, mind, and society," *Trends in Cognitive Sciences*, vol. 19, no. 1, pp. 46–54, 2015.

[27] X. Wang and H. Wang, "Evolutionary optimization with markov random field prior," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 6, pp. 567–579, 2004.

[28] A. Engel, M. Burke, K. Fiehler, S. Bien, and F. Rösler, "What activates the human mirror neuron system during observation of artificial movements: bottom-up visual features or top-down intentions?" *Neuropsychologia*, vol. 46, no. 7, pp. 2033–2042, 2008.

[29] R. A. Bjork and T. D. Wickens, "Memory, metamemory, and conditional statistics," *Behavioral and Brain Sciences*, vol. 19, no. 2, pp. 193-194, 1996.

[30] S. S. Basha, "Best proximity points: global optimal approximate solutions," *Journal of Global Optimization*, vol. 49, no. 1, pp. 15–21, 2011.

[31] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57, 2007.

[32] D. Jo, J. Han, K. Chung, and S. Lee, "Empathy between human and robot?" in *Proceedings of the 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 151-152, IEEE, Tokyo, Japan, 2013.

[33] N. A. Atiya, I. Rañó, G. Prasad, and K. Wong-Lin, "A neural circuit model of decision uncertainty and change-of-mind," *Nature Communications*, vol. 10, no. 1, p. 2287, 2019.

[34] R. D. Morey, "A Bayesian hierarchical model for the measurement of working memory capacity," *Journal of Mathematical Psychology*, vol. 55, no. 1, pp. 8–24, 2011.

[35] B. Jiang and H. R. Karimi, "Sliding mode control of semi-markovian jump systems," *Sliding Mode Control of Semi-Markovian Jump Systems*, pp. 87–113, 2021.

[36] K. L. Chung, "The general theory of markov processes according to doeblin," *Probability Theory and Related Fields*, vol. 2, no. 3, pp. 230–254, 1964.

[37] T. Yamashita, "On a support system for human decision making by the combination of fuzzy reasoning and fuzzy structural modeling," *Fuzzy Sets and Systems*, vol. 87, no. 3, pp. 257–263, 1997.

[38] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, Hoboken, NJ, USA, 2014.

[39] P.-L. Lions and J.-L. Menaldi, ""Optimal control of stochastic integrals and Hamilton–Jacobi–bellman equations. i," *SIAM Journal on Control and Optimization*, vol. 20, no. 1, pp. 58–81, 1982.

[40] B. Efron, "Bootstrap methods: another look at the jackknife," in *Breakthroughs in Statistics*, pp. 569–593, Springer, Berlin, Germany, 1992.

[41] H. Wan, X. Luan, H. R. Karimi, and F. Liu, "A resource-aware sliding mode control approach for markov jump systems," *ISA Transactions*, 2020.

[42] B. Tate, "Boeing 747 training developments and implementation," SAE, Warrendale, PA, USA, Technical Paper 710473, 1971.

[43] B. Ren, T. Yin, and S. Fu, "An approach analyzing cognitive process of human-machine interaction based on extended markov decision process," in *Proceedings of the 2019 Chinese Automation Congress (CAC)*, pp. 1306–1311, IEEE, Hangzhou, China, 2019.

[44] E. L. Wiener and D. C. Nagel, *Human Factors in Aviation*, Gulf Professional Publishing, Oxford, UK, 1988.

[45] D. E. Maurino, J. Reason, N. Johnston, and R. B. Lee, *Beyond Aviation Human Factors: Safety in High Technology Systems*, CRC Press, Boca Raton, FL, USA, 2017.

[46] K. Miller, "Timeline analysis program (tla-1), final report, boeing document D6-42377-5, prepared for National Aeronautics and Space Administration, Langley Research Center (NASA-CR-144942)," 1976.

[47] E. Galy, M. Cariou, and C. Mélan, "What is the relationship between mental workload factors and cognitive load types?" *International Journal of Psychophysiology*, vol. 83, no. 3, pp. 269–275, 2012.

[48] A. Fink and A. Neubauer, "Individual differences in time estimation related to cognitive ability, speed of information processing and working memory," *Intelligence*, vol. 33, no. 1, pp. 5–26, 2005.

[49] F. Meyniel, M. Maheu, and S. Dehaene, "Human inferences about sequences: a minimal transition probability model," *PLoS Computational Biology*, vol. 12, no. 12, 2016.

[50] S. E. Petersen and M. I. Posner, "The attention system of the human brain: 20 years after," *Annual Review of Neuroscience*, vol. 35, pp. 73–89, 2012.