Structural knowledge error, rather than reward insensitivity, explains the reduced metacontrol in aging

Zhaoyu Zuo[1,2], Lizhuang Yang[1,2,3] ✉, and Hai Li[1,2,3] ✉

[1]Anhui Province Key Laboratory of Medical Physics and Technology, Institute of Health and Medical Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China;
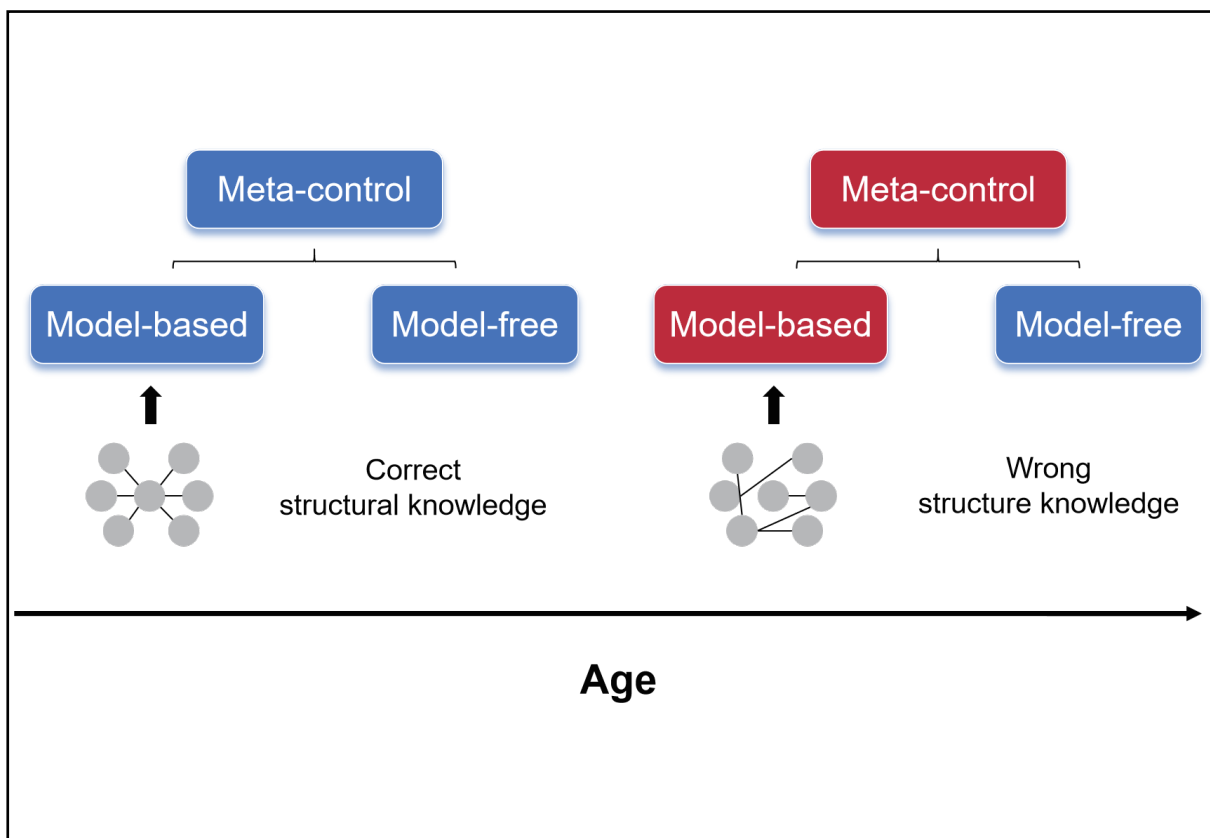*[2]University of Science and Technology of China, Hefei 230026, China;*
[3]Hefei Cancer Hospital, Chinese Academy of Sciences, Hefei 230031, China

✉Correspondence: Lizhuang Yang, E-mail: lzyang@cmpt.ac.cn; Hai Li, E-mail: hli@cmpt.ac.cn

Graphical abstract

The aging of metacontrol is due to structural knowledge errors.

Public summary

■ Pupillometry analysis proves that older adults are sensitive to the level of reward.

■ Older adults with intact structural knowledge show a comparable metacontrol efficiency to younger adults.

■ Computer simulations suggest that the aging of metacontrol is due to structural knowledge errors.

Structural knowledge error, rather than reward insensitivity, explains the reduced metacontrol in aging

Zhaoyu Zuo[1,2], Lizhuang Yang[1,2,3] ✉, and Hai Li[1,2,3] ✉

[1]*Anhui Province Key Laboratory of Medical Physics and Technology, Institute of Health and Medical Technology, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China;*
[2]*University of Science and Technology of China, Hefei 230026, China;*
[3]*Hefei Cancer Hospital, Chinese Academy of Sciences, Hefei 230031, China*

✉Correspondence: Lizhuang Yang, E-mail: lzyang@cmpt.ac.cn; Hai Li, E-mail: hli@cmpt.ac.cn

Abstract: Humans flexibly adjust their reliance on model-free (habitual) and model-based (goal-directed) strategies according to cost–benefit trade-offs, the ability of which is known as metacontrol. Recent studies have suggested that older adults show reduced flexibility in metacontrol. However, whether the metacontrol deficit in aging is due to cognitive or motivational factors remains ambiguous. The present study investigated this issue using pupillometry recording and a sequential decision-making task with varied task structures and reward stakes. Our results revealed that older adults performed less model-based control and less flexibility when the reward stake level changed, consistent with previous studies. However, pupillometry analysis indicated that older adults showed comparable sensitivity to the reward stake. Older adults varied in task structure knowledge according to their oral reports, and the subgroup with good structural knowledge exerted a similar pattern to younger adults. Computational simulation verified that poor structure knowledge representation impaired metacontrol. These results suggest that the inflexible metacontrol in the elderly population might not be due to motivational factors but rather poor structure knowledge.

Keywords: decision-making; metacontrol; reinforcement learning; aging

CLC number: B842   **Document code:** A

1 Introduction

Advance in decision-making research indicates that behavior is governed by at least two separable control systems[1,2]: System 1 is automatic and error prone, while System 2 is cognitively demanding and reliable[3]. Due to the complementary nature of these two systems in terms of cost and benefit (resulting from behavior), they apply to different task scenarios[4]. As a control mechanism for monitoring and adjusting cognitive control, metacontrol can dynamically regulate these two control systems based on a cost–benefit analysis of the current context[5]. Recent decision-making research has discovered a noticeable weakening of metacontrol in aging (i.e., changes in cost and benefit do not impact system-level control)[6]. However, the underlying mechanism behind metacognitive inefficiency remains elusive.

The hybrid reinforcement learning model provides a quantitative tool for evaluating metacontrol[4]. The computational model associates System 1 and System 2 with model-free and model-based reinforcement learning methods, in which the model-free method directly establishes the direct association between actions and rewards, while the model-based method plans based on learned structural knowledge (the mapping between actions and states)[7]. Assuming that the computational results of the two systems are mixed by linear weighting, the model-based weights in the hybrid reinforcement learning

model represent the relative contribution of the model-based system to the decision. Metacontrol studies usually set independent model-based weights in different task contexts and analyze the mechanism of metacontrol by observing the influence of experimental conditions on the weights[6].

Theoretical work describes the metacontrol process as follows: participants estimate the expected rewards of each system in the current situation, then weigh it against the cost of model-based control to find a balance point for control, and finally regulate the intensity of model-based control[2,8]. This description has received considerable empirical support. For example, when the stake (reward magnitude) is high, people tend to increase their cognitive investment in model-based control[4,9]. However, when planning complexity increases or the structure turns variable, people tend to decrease their model-based control[2].

Older adults exhibit weakened metacontrol compared to younger adults[6]. The inefficiency of metacontrol may be due to age-related differences in learning and maintaining task structural representation[10–12]. Obviously, metacontrol cannot work properly even if the structural knowledge is poorly represented. Alternatively, metacontrol inefficiency might stem from reward insensitivity or reduced willingness or motivation to perform cost–benefit analysis[6]. This view is consistent with recent research that suggests the existence of age-related deterioration of stimulus-reward processing[13].

However, there is also accumulating evidence that older people are just as sensitive to rewards as younger people. For instance, functional magnetic resonance imaging studies using incentive delay tasks suggest fewer age-related changes in reward network responses during anticipation and outcome processing[14]. In addition, it has been observed that participants exhibit more vigorous movement responses to obtain rewards when the average reward rate is higher, and this response is similar between older and younger groups[15].

The present study investigates the cognitive and motivational factors underlying metacontrol inefficiency by combining pupillometry and behavior experiments. Similar to Bolenz et al.[6], we used a sequential decision-making task that manipulated reward stake (low-stake vs. high-stake) and task structure variability (stable transition vs. variable transition). In addition, there were two methodological improvements. First, pupillometry recording was used to measure motivation to allocate cognitive effort[16,17]. Second, we used a structured questionnaire to ask participants' knowledge about the task background and task structure. The novel design of the experiment enabled us to systematically examine the motivational and cognitive differences between older and younger adults. Consistent with previous research[6], we predicted that older adults should exhibit weaker metacontrol than younger adults, manifested as the insignificant influence of stake and transition conditions on model-based weights in behavioral analysis. However, we predicted that older adults should be sensitive to stake levels and feedback but insensitive to structural needs in pupillometry analysis. In addition, we expected that the intactness of structure knowledge reflected in the subjective oral reports determined whether older adults could show comparable metacontrol efficiency as younger adults. Last, we used computational simulation to demonstrate that poor structure knowledge representation impaired metacontrol.

2 Materials and methods

2.1 Participants

We recruited 32 healthy younger and 55 healthy older adults. Before the appointment, participants were telephone screened for potential exclusion due to psychiatric or neurological illness. All subjects had normal or corrected-to-normal vision and abstained from caffeine in the hour immediately preceding their participation. Seven participants were excluded for failing to complete the task, leaving a final sample of 29 younger adults (age range 20–44 years, mean age 30.14 years, 14 female) and 51 older adults (age range 60–83 years, mean age 68.23 years, 26 female). The sample size was comparable to that in a previous study[18].

Pupil sizes were measured using a desktop eye-tracking platform (PupilEXT) with a sampling rate of 100 Hz[19]. Pupil diameter was determined by employing the Pure algorithm[20]. The stimulus materials utilized in the previous study[6] were employed. The Ethics Committee of Hefei Institutes of Physical Science, Chinese Academy of Sciences, approved the research protocol (No. YXLL-2023-32). All participants provided written informed consent and received monetary incentives after completion of the study.

2.2 Experimental procedure

The experimental procedure included a practice session, quiz A, a formal session and quiz B. In each trial, the participants viewed a cue followed by two-step choices and then feedback indicating how much reward they received. Three hundred twenty trials were equally divided by the combination of low- or high-stake and stable- or variable-transition.

Each trial began with one of two first-stage states, each containing two alternatives (Fig. 1a). All four first-stage choices deterministically led to one of the two second-stage states. In the second stage, participants had to press the space bar to earn scalar points generated by the random walk (ranging from 0 to 9).

The experiment comprised 320 trials organized into four blocks of 80 trials each: two stable transition blocks and two variable transition blocks. These blocks alternated in sequence. In half of the trials, a high-stake cue was randomly assigned, while the remaining half received a low-stake cue (Fig. 1b). At the onset of each trial, the stake cue provided information to the participant regarding the trial's stakes condition. In other words, these cues determined how rewards were converted into points. In low-stake trials, the obtained reward was multiplied by a factor of 1, whereas in high-stake trials, the reward was multiplied by a factor of 5. The transition structure remained the same throughout the block in stable transition blocks. In contrast, during variable transition blocks, changes in the trial structure would occur every 6 to 14 trials.

The whole process, including receiving instructions, performing practice, taking two quizzes and a test session, and the final payment, took between 80 and 110 min for each subject. The experimental stimulus material was partly drawn from a shared repository from previous studies[21], and the paradigm is similar to Bolenz et al[6]. The experiment was conducted using jsPsych[22] and executed on a standard PC. All participants controlled the experiment using a computer keyboard. The left and right spaceships were chosen using the F and J keys, respectively, while the space bar was used for the second-stage response. The response time limit for both
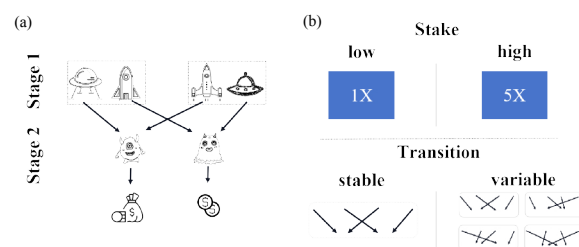

Fig. 1. Behavioral task. (a) State transition structure of the task. Each trial starts with a random first-stage state. Given the transition structure, each first-stage choice leads deterministically to the second-stage state. Each second-stage choice is associated with a drifting scalar reward. (b) The stake manipulation (top). One of the high-stake or low-stake tips is randomly presented at the beginning of the trial, which means that the actual benefit of the trial is several times the score shown in the feedback. Transition manipulation (bottom). The task transition did not change in stable-transition blocks, and the task transition structure changed irregularly in variable-transition blocks.

the first-stage and second-stage phases was set at 2 s.

In the practice session, participants were repeatedly instructed to choose one spaceship in the first stage, transition to one planet in the second stage, and eventually receive the random-walked reward. Participants were instructed extensively about the reward distribution, the transition structure, and the stakes manipulation. Participants must choose the correct path to a specified planet ten times in a row and answer the correct score they received under a specified stake condition ten times in a row. Participants also had to complete 40 practice trials that were divided into two blocks within stable and variable transitions.

Before the formal testing phase, we used quiz A to test participants' understanding of the task rules. Quiz A includes:

(i) "How many spaceships are in the task?" The options include "2" and "4". The correct answer is "4".

(ii) "How many planets are in the task?" The options include "2" and "4". The correct answer is "2".

(iii) "If you got 4 treasures on the purple planet in the last trial, what is the range of the number of treasures you are most likely to get on the purple planet in this trial?" The options include "0–2", "3–5" and "7–9". The correct answer is "3–5".

(iv) "The treasure yield on the red planet is increasing. What about the treasure yield on the purple planet?" The options include "increasing", "decreasing", "remain the same", and "All of the above could happen." The correct answer is "All of the above could happen."

(v) "If we saw spaceship 1 in the first stage, what would the other ship be?" The options include "spaceship 2", "spaceship 3", "spaceship 4", and "All of the above could happen." The correct answer is "spaceship 2".

(vi) "If spaceship 3 goes to the purple planet, what planet does spaceship 4 go to?" The options include "red planet", "purple planet", and "All of the above could happen." The correct answer is "red planet".

(vii) "If spaceship 1 always goes to the red planet, this turn suddenly brings you to the purple planet, what's the reason?" The options include "I pressed the wrong key." and "The transition structure has just changed." The correct answer is "The transition structure has just changed."

Participants who scored less than 5 points on quiz A dropped out of the experiment, and for the rest of the participants, the experimenter gave them an oral explanation of the incorrectly answered questions from a prewritten manuscript.

The left and right spaceships were chosen using the F and J keys, respectively, and the second-stage response was executed by pressing the space bar. If participants failed to respond within the 2-second time limit for both the first-stage and second-stage states, the experiment advanced to the next trial. If responses were made within the allotted time, the selected option remained highlighted for the remainder of the response window.

After the test session, the participants were asked to complete quiz B. The quiz contained the entire content of the previous quiz A and the following questions:

(viii) "What is the main basis for choosing the spaceship?" The options include "intuition", " reasoning", and "both".

(ix) "Will you be more serious when the price sign is five times higher?" The options include "yes" and "no".

(x) "Do you always remember that more gems can be exchanged for more cash?" The options include "yes" and "no".

(xi) "How long do you need to redefine the route when the route schedule changes?" The options include "Easy: 1–3 trials", "Medium: more than 4 trials", and "Hard: It is hard to figure it out."

2.3 Preprocessing of pupillometric data

Preprocessing of the pupil data was conducted based on previous studies[16] and involved several steps. First, eye blink artifacts were removed by excluding data from 100 ms before the blink to 150 ms after. Pupil sizes smaller than 1 mm were also excluded, and data points that exceeded the inner fences were removed. In cases of missing data, linear interpolation was applied up to 500 ms. To eliminate slow drift below 0.012 Hz, a high-pass Butterworth filter was utilized on the pupil diameter measurements.

Three distinct processing modes are subsequently distinguished. When analyzing the pupil response induced by the stake, the pupil diameter values were standardized (Z scored) within blocks to ensure comparability across different blocks. Following this, baseline correction was performed by subtracting the mean diameter of a 200 ms baseline period preceding the trial onset. Alternatively, when analyzing the pupil response induced by feedback, baseline correction was conducted by subtracting the mean diameter of a 200 ms baseline period before the feedback phase onset. When comparing the pupil baseline of different blocks, Z-scoring within the blocks was not applied. Due to the fixed distance between the camera and the headrest, the pupil diameters recorded in pixels by PupilEXT were converted into actual diameters using a fixed coefficient.

2.4 Computational model

We implemented an established hybrid reinforcement learning algorithm in our study[6,21,23]. This algorithm combines elements from both model-based learning methods and the model-free SARSA(λ) algorithm[24] to learn action values.

For each trial t, δ_1 and δ_2, the reward prediction errors of the first stage and second stage are calculated as:

$$\delta_1 = Q_{MF}(s_{2,t}, a_{2,t}) - Q_{MF}(s_{1,t}, a_{1,t}) \ , \tag{1}$$

$$\delta_2 = r_t - Q_{MF}(s_{2,t}, a_{2,t}) \ , \tag{2}$$

where $s_{1,t}$ and $s_{2,t}$ are the first stage and second stage at trial t, respectively. $a_{1,t}$ and $a_{2,t}$ are the actions in the first stage and second stage at trial t, respectively. $(s_{1,t}, a_{1,t})$ and $(s_{2,t}, a_{2,t})$ is the state-action pair in the first stage and second stage at trial t, respectively. r_t is the received immediate reward at trial t.

The update of the model-free state-action value Q_{MF} at the first stages and second stages is performed as:

$$Q_{MF}(s_{1,t}, a_{1,t}) = Q_{MF}(s_{1,t}, a_{1,t}) + \alpha\delta_1 + \alpha\lambda\delta_2 \ , \tag{3}$$

$$Q_{MF}(s_{2,t}, a_{2,t}) = Q_{MF}(s_{2,t}, a_{2,t}) + \alpha\delta_2 \ , \tag{4}$$

where α is the learning rate ($0 \leqslant \alpha \leqslant 1$), and λ is the

eligibility trace decay ($0 \leqslant \lambda \leqslant 1$) that modulates the extent of the reward prediction error in the second stage influencing the first-stage state-action value.

The model-based system learns a transition probability matrix $T(s_2 \mid s_{1,t}, a_{1,t})$ representing the probability of transitioning to the second-stage state s_2 after selecting an action a_1 in the first-stage state s_1. The transition probability matrix can be set to be an invariant matrix or updated by state prediction errors. Please refer to Daw et al.[23] and Bolenz et al.[6] for detailed model information.

The model-based state-action value at the first stage is given by:

$$Q_{\mathrm{MB}}(s_{1,t}, a_{1,t}) = \sum_{s_2} T(s_2 \mid s_{1,t}, a_{1,t}) Q_{\mathrm{MF}}(s_{2,t}, a_{2,t}) \ . \quad (5)$$

Outside of the model-based and model-free systems, the model-free and model-based state-action values are combined according to the model-based weight ω ($0 \leqslant \omega \leqslant 1$):

$$Q_{\mathrm{net}}(s_1, a_1) = \omega Q_{\mathrm{MB}}(s_1, a_1) + (1-\omega) Q_{\mathrm{MF}}(s_1, a_1) \ . \quad (6)$$

The choice probabilities were calculated as follows:

$$P(a_{1,t} = a_1 \mid s_{1,t}) = \frac{\exp(\beta [Q_{\mathrm{net}}(s_{1,t}, a_1) + \pi \cdot \mathrm{rep}(a_1) + \rho \cdot \mathrm{resp}(a_1)])}{\sum_{a_1'} \exp(\beta [Q_{\mathrm{net}}(s_{1,t}, a_1') + \pi \cdot \mathrm{rep}(a_1') + \rho \cdot \mathrm{resp}(a_1')])}, \quad (7)$$

where β is the inverse SoftMax temperature. The choice stickiness parameter π captures the stimulus preference, and the response stickiness ρ captures the key preference. $\mathrm{rep}(a_1)$ is equal to 1 when choosing the same stimulus as the last trial and 0 otherwise. $\mathrm{resp}(a_1)$ is equal to 1 when choosing the same key as the last trial and 0 otherwise.

2.5 Statistical analyses

Paired-sample Student's t tests (two-tailed) were used for comparisons. Hierarchical Bayesian regression analyses were performed in R Studio using the brms package[25]. Effect coding was employed for categorical variables (younger adults = −1, older adults = 1, low-stake = −1, high-stake = −1, stable-transitions = −1, variable-transitions = −1), allowing for the interpretation of regression coefficients as main effects. Each reported regression coefficient is characterized by the mean of its marginal posterior distribution and the 95% credible interval. The interval is computed as the [0.025,0.975] percentile range, and it can be understood as containing the parameter of interest with 95% probability.

3 Results

We conducted a metacontrol task (Fig. 1) in younger and older adults. The task involved presenting stake cues randomly at the onset of a single sequential decision-making trial, with the cues indicating either a high or low reward magnitude. Additionally, the task manipulated the level of structural variability at the block level by either keeping it stable or making it variable. For each participant, we estimated model-based weights in four different scenario demands. We compared the estimated model-based weights parameter with the hierarchical regression analysis in all participants and

discovered that aging had reduced model-based weights ($\beta_{\mathrm{age\ group}}$ = −0.06, 95% CI = [−0.08, −0.04]).

3.1 Reduced metacontrol in older adults

We separately compared the model-based weights with an effect-coded hierarchical regression analysis in younger and older adults (Fig. 2). For younger adults, the results indicated a main effect of stakes condition (β_{stake} = 0.04, 95% CI = [0.01, 0.07]), indicating an elevation in model-based control during high-stakes trials. Furthermore, the transition condition also exhibited a significant main effect ($\beta_{\mathrm{transition}}$ = −0.04, 95% CI = [−0.07, −0.01]), signifying that model-based control was less pronounced in variable-transition blocks than in stable-transition blocks.

In contrast, the model-based weights of older adults were not affected by task conditions (β_{stake} = 0.01, 95% CI = [−0.02, 0.04], $\beta_{\mathrm{transition}}$ = 0.01, 95% CI = [−0.01, 0.04]). These findings align with previous research[6], demonstrating that older adults performed weaker model-based control and exhibited weaker metacontrol than their younger counterparts.

3.2 Older adults are sensitive to stake cues and feedback

To investigate the participants' sensitivity to rewards, we analyzed their real-time pupil diameter during low- and high-stake trials. Fig. 3a and 3d present a significant difference between the two stake conditions, indicating that high-stake trials elicited larger pupil diameters than low-stake trials. Then, we examined the reaction time under low-stake and high-stake conditions. Consistent with the findings in the pupillary analysis, we observed that high-stake trials led to shorter reaction times than low-stake trials during stage 1 for both younger adults (paired t test: $t(28) = 2.69, p = 0.01$, Fig. 3b) and older adults (paired t test: $t(50) = 2.34, p = 0.02$, Fig. 3c). Additionally, this effect is replicated in stage 2 (younger adults, paired t test: $t(28) = 3.1, p = 0.004$, Fig. 3e; older adults, paired t test: $t(50) = 3.07, p = 0.003$, Fig. 3f).

These results show that cognitive effort can be adaptively allocated based on stake cues, supporting the notion that older adults exhibit similar sensitivity to reward magnitude as
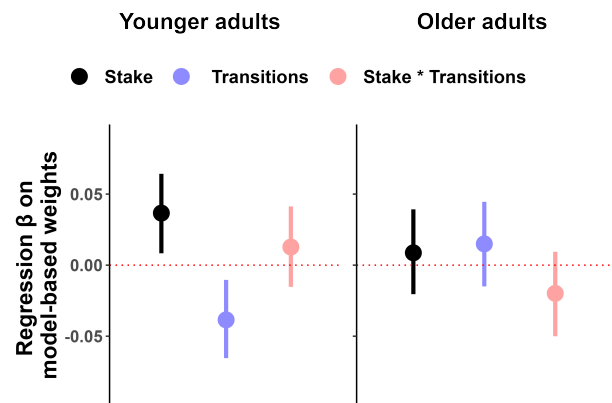

Fig. 2. The metacontrol effect. Logistic regression weights show the influence of stakes, transitions and their interaction effect on the model-based weights in the two groups. The vertical line represents the 95% confidence interval, and the dots represent the mean.
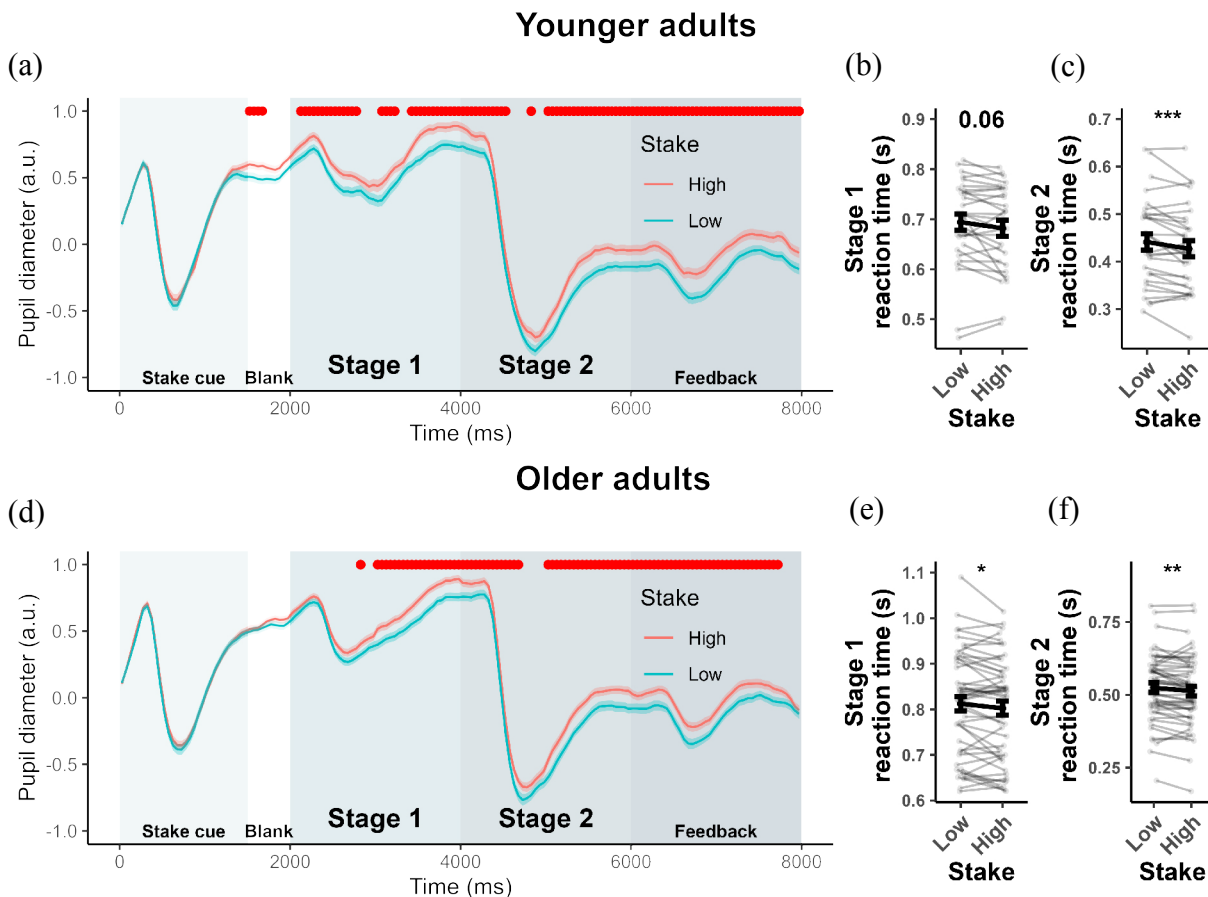
Fig. 3. Time series depicting the average pupil diameter of low-stake and high-stake trials (a, d) over the course of trials. Solid lines indicate the mean pupil diameter (baseline-corrected). Shaded areas indicate standard errors (SEs) of pupil diameter (baseline-corrected). Red lines on the top indicate time points of a reliable [$p < 0.05$] positive effect for younger and older adults. Reaction time in stage 1 (b, e) and stage 2 (c, f). The error bar represents the standard error of the mean. * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

younger adults.

We also analyzed whether participants could differentiate between different types of feedback by observing their pupil responses during the feedback phase. Our findings showed that both groups could distinguish between different feedback sizes based on the changes in pupil diameters (Fig. 4). This finding further supports the notion that older adults are sensitive to rewards.

3.3 Older adults are not sensitive to structural needs

To investigate whether participants were sensitive to structural needs, we analyzed their real-time pupil diameter during stable-transition and variable-transition blocks. Our findings revealed that in younger adults, the structure needs in the variable-transition block led to an increase in pupil diameter (Fig. 5a and 5d) and in reaction time (stage 1: paired t test: t (28) = −1.8, $p = 0.08$, Fig. 5b; stage 2: paired t test: t (28) = −2.79, $p = 0.0009$, Fig. 5c). However, these changes were not observed in older adults (stage 1, paired t test: t (50) = −0.04, $p = 0.96$, Fig. 5e; stage 2, paired t test: t (50) = 0.26, $p = 0.79$, Fig. 5f).

These results suggest that older adults have a diminished sensitivity to structural needs and renewal, implying that there may be defects in the representation of structural knowledge.

3.4 Subjective report recognizes older adults with metacontrol

Our experimental results suggest that older adults are not sensitive to structure. To confirm that differences in structural knowledge are an essential reason for the decline in metacontrol in older adults, we group participants using subjectively reported difficulty in updating structures. Twenty-eight, 1, and 0 of 29 younger participants reported easy/medium/hard, respectively, compared with 25, 15, and 11 of 51 older participants, respectively. We found significant metacontrol effects under stake conditions ($\beta_{stake} = 0.04$, 95% CI = [0.01, 0.07], Fig. 6) among older adults who perceived structural renewal as easy and moderate evidence for metacontrol effects under transition conditions ($\beta_{transition} = -0.02$, 95% CI = [−0.05, 0], with 92% of the posterior mass below zero), which suggests that older adults with good structural knowledge can exert metacontrol similar to younger adults.

3.5 Metacontrol can be masked by structural knowledge error

To further illustrate the relationship between structural knowledge error and metacontrol, we generated simulated data demonstrating that standard computational models have difficulty detecting metacontrol in participants with inaccurate
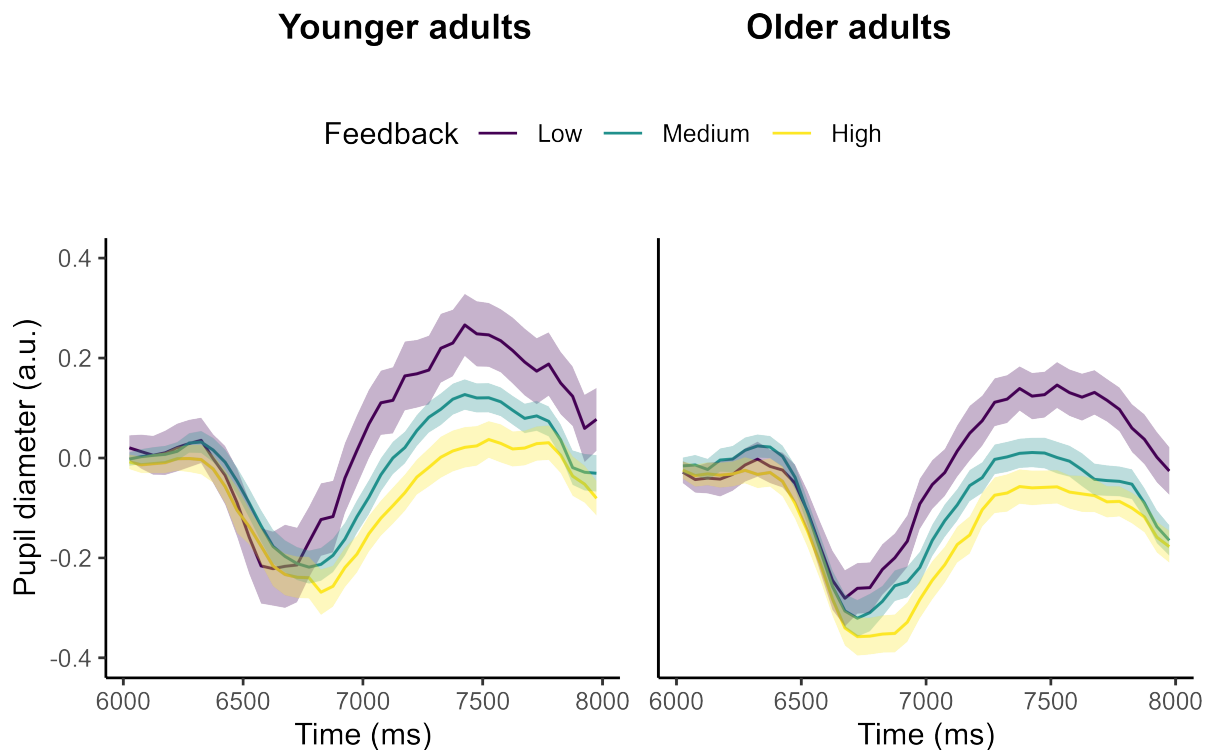
Fig. 4. Time series depicting the average pupil diameter of high (7–9), medium (3–6) and low (0–2) reward trials in the feedback phase. Solid lines indicate the mean pupil diameter (baseline-corrected). Shaded areas indicate SEs of pupil diameter (baseline-corrected).

structural representations. Our approach involves introducing different noise levels to the classical hybrid reinforcement learning model, accompanied by varying degrees of model-based weight adjustment. Subsequently, we use these agents for behavioral simulation and fit the simulated behavior with a model that assumes correct structural knowledge. As anticipated, our simulations reveal that higher levels of structural noise led to less differentiated estimates of model-based weight (Fig. 7). Older adults with inaccurate structural representations may encounter similar challenges.

4 Discussion

This study aimed to determine whether cognitive or motivational factors are primarily responsible for weakening metacontrol in older adults. In our metacontrol task, pupillary and behavioral evidence supports that older adults are as sensitive to stake cues and feedback as younger adults, refuting the assumption that older individuals are insensitive to rewards or do not engage in cost–benefit analysis in sequential decision-making. Then, we found that older adults are minimally affected by structural needs in comparison to younger adults, indicating cognitive deficits in the foundation of model-based decision-making among older individuals. Importantly, older adults with good structural knowledge exercised metacontrol as younger adults, underscoring the role of structural knowledge error in reduced metacontrol in aging. Finally, model simulation shows that if the structure is not expressed correctly, it is difficult for the computational model to detect even if metacontrol is performed.

The theory of the expected value of control[26] describes the computational principles and underlying processes of cognitive control allocation, including monitoring, specification (based on cost–benefit analysis), and regulation. Based on this theory, metacontrol in decision-making can be decomposed into monitoring the decision-making process, determining the intensity of model-based control to exert, and implementing the specified model-based control. In contrast to previous metacontrol research in aging[6], our study identifies the stages at which errors occur in metacontrol. Specifically, the difference in pupil response and reaction time following different stake cues suggests that cognitive control has been allocated and executed. Therefore, motivational factors may play a secondary role, and the observed reduced metacontrol in older adults can primarily be attributed to the failure of model-based control regulation. This conclusion is consistent with recent findings showing that metacontrol is mainly independent of intrinsic motivation[27].

An explicit structure is essential for model-based decisions[18,28]. However, this cognitive foundation is greatly affected by aging. Consistent with previous research that found that older adults may have age-related deficits in the representation of task transition structures or state spaces[10,11], our study did not find pupillometry evidence that older adults responded to structural needs. These results all indicate that older adults rarely maintain accurate structural knowledge. When the structure is unreliable, it is challenging to regulate model-based control that relies on structure precisely. Moreover, this viewpoint is supported by our model
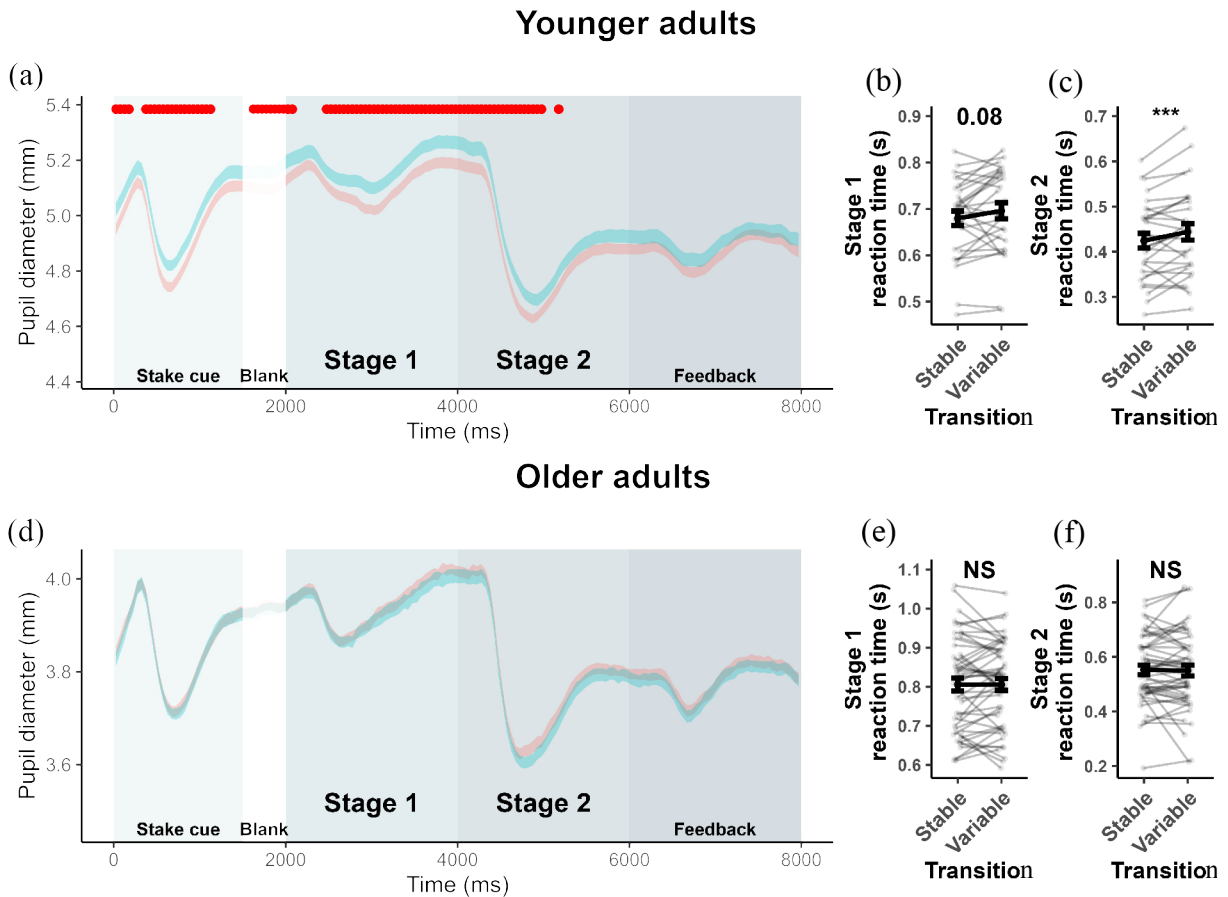
Younger adults

Older adults

Fig. 5. Time series depicting the average pupil diameter of stable-transition and variable-transition trials (a, d) over the course of trials. Solid lines indicate the mean pupil diameter. Shaded areas indicate SEs of pupil diameter (baseline-corrected). Red lines on the top indicate time points of a reliable [$p < 0.05$] positive effect for younger and older adults. Reaction time in stage 1 (b, e) and stage 2 (c, f). The error bar represents the standard error of the mean. *** indicates $p < 0.001$ and NS = not significant.
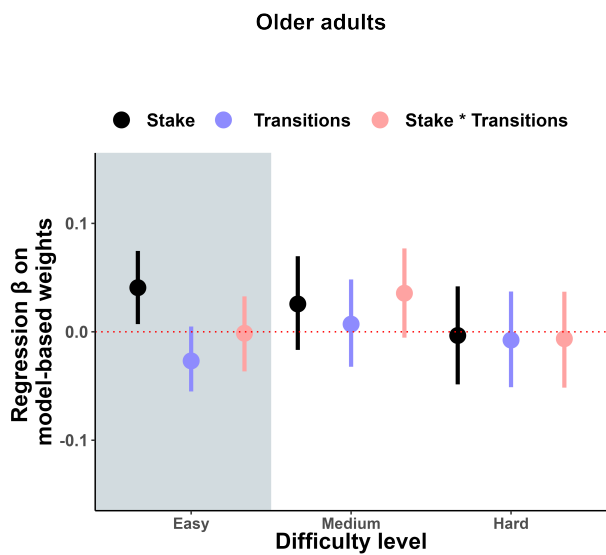

Fig. 6. The metacontrol effect after grouping. Logistic regression weights show the influence of stakes, transitions and their interaction effect on the model-based weights in older adults. Older adults were grouped according to subjective reports of different levels of structural update difficulty. The vertical line represents the 95% confidence interval, and the dots represent the mean.
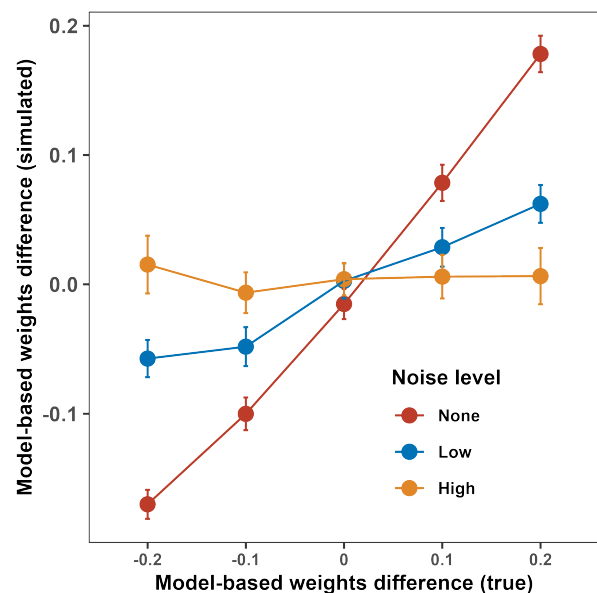
Fig. 7. Model prediction. Comparison of simulated vs. empirically observed model-based weight differences. Agents were divided into three groups based on structural noise levels. The error bar represents the standard error of the mean.

simulation results.

The second explanation for being unable to observe metacontrol in older adults is that they do not use model-based control. For example, previous research has shown that older adults are more prone to using heuristic methods than younger adults[29]. Specifically, an experiment found that older adults use the heuristic "win-stay, lose-shift" strategy instead of the more complex model-based strategy used by younger adults[30]. If older adults adopt heuristic strategies or discard model-based decision-making in sequential decision-making tasks, using the classical dual-system model may lead to a misunderstanding of the relative contributions of different decision strategies and, naturally, a misunderstanding of the changes in relative contributions of the two systems under different contextual demands. Future work should use advanced computational models and matched paradigms to solve this problem.

Our study has several limitations that need to be further optimized in future studies. First, our interpretation of the experimental phenomenon is limited. The current study found that younger adults exhibited earlier and longer pupil dilation after high-stake cues compared to older adults, which may reveal differences between younger and older adults in cognitive effort allocation or preplanning processes[31]. The underlying mechanisms go beyond this study's scope. Second, although the sample size of our study was comparable to that of previous studies, a larger sample size might better reveal the individual variability of metacontrol.

5   Conclusions

In summary, our results suggest that metacontrol inefficiency in older adults is primarily due to the inability to learn and maintain structural knowledge rather than reward insensitivity or reduced motivation. The present study sheds light on aging-related impairments in goal-directed behavior[32,33]. We advocate that future studies explore intervention strategies to help older adults flexibly construct task structure representations in daily life activities.

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Zhaoyu Zuo   received his master's degree in Pattern Recognition and Intelligent Systems from the University of Science and Technology of China. His research mainly focuses on the competitive cooperative relationship between working memory and reinforcement learning.

Lizhuang Yang   received his Ph.D. degree in Psychology from the Chinese University of Hong Kong. He is currently an Associate Professor at the Hefei Institutes of Physical Science, Chinese Academy of Sciences. His research interests include decision-making, metacognition, and individual difference.

Hai Li   received his Ph.D. degree in Automation from the Northwestern Polytechnical University. He is currently a Professor at the Hefei Institutes of Physical Science, Chinese Academy of Sciences. His research focuses on neuroimaging, cognitive science, and translational technologies.

References

[1] Collins A G E, Cockburn J. Beyond dichotomies in reinforcement learning. *Nature Reviews Neuroscience*, **2020**, *21* (10): 576–586.

[2] Kool W, Gershman S J, Cushman F A. Planning complexity registers as a cost in metacontrol. *Journal of Cognitive Neuroscience*, **2018**, *30* (10): 1391–1404.

[3] Gilovich T, Griffin D, Kahneman D. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge: Cambridge University Press, **2002**.

[4] Kool W, Gershman S J, Cushman F A. Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological Science*, **2017**, *28* (9): 1321–1333.

[5] Kool W, Cushman F A, Gershman S J. Competition and cooperation between multiple reinforcement learning systems. In: Morris R, Bornstein A, Shenhav A, editors. Goal-Directed Decision Making. New York: Academic Press, **2018**: 153–178.

[6] Bolenz F, Kool W, Reiter A M, et al. Metacontrol of decision-making strategies in human aging. *eLife*, **2019**, *8*: e49154.

[7] Gläscher J, Daw N, Dayan P, et al. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, **2010**, *66* (4): 585–595.

[8] Kool W, Botvinick M. Mental labour. *Nature Human Behaviour*, **2018**, *2* (12): 899–908.

[9] Smid C R, Ganesan K, Thompson A, et al. Neurocognitive basis of model-based decision making and its metacontrol in childhood. *Developmental Cognitive Neuroscience*, **2023**, *62*: 101269.

[10] Hämmerer D, Schwartenbeck P, Gallagher M, et al. Older adults fail to form stable task representations during model-based reversal inference. *Neurobiology of Aging*, **2019**, *74*: 90–100.

[11] Eppinger B, Heekeren H R, Li S C. Age-related prefrontal impairments implicate deficient prediction of future reward in older adults. *Neurobiology of Aging*, **2015**, *36* (8): 2380–2390.

[12] Ruel A, Bolenz F, Li S C, et al. Neural evidence for age-related deficits in the representation of state spaces. *Cerebral Cortex*, **2023**, *33* (5): 1768–1781.

[13] Vink M, Kleerekooper I, van den Wildenberg W P M, et al. Impact of aging on frontostriatal reward processing. *Human Brain Mapping*, **2015**, *36* (6): 2305–2317.

[14] Spaniol J, Bowen H J, Wegier P, et al. Neural responses to monetary incentives in younger and older adults. *Brain Research*, **2015**, *1612*: 70–82.

[15] Hird E J, Beierholm U, De Boer L, et al. Dopamine and reward-related vigor in younger and older adults. *Neurobiology of Aging*, **2022**, *118*: 34–43.

[16] da Silva Castanheira K, LoParco S, Otto A R. Task-evoked pupillary responses track effort exertion: Evidence from task-switching. *Cognitive, Affective, & Behavioral Neuroscience*, **2021**, *21* (3): 592–606.

[17] Rondeel E W M, van Steenbergen H, Holland R W, et al. A closer

look at cognitive control: differences in resource allocation during updating, inhibition and switching as revealed by pupillometry. *Frontiers in Human Neuroscience*, **2015**, *9*: 494.

[18] Feher da Silva C, Hare T A. Humans primarily use model-based inference in the two-stage task. *Nature Human Behaviour*, **2020**, *4* (10): 1053−1066.

[19] Zandi B, Lode M, Herzog A, et al. PupilEXT: flexible open-source platform for high-resolution pupillometry in vision research. *Frontiers in Neuroscience*, **2021**, *15*: 676220.

[20] Santini T, Fuhl W, Kasneci E. PuRe: Robust pupil detection for real-time pervasive eye tracking. *Computer Vision and Image Understanding*, **2018**, *170*: 40−50.

[21] Kool W, Cushman F A, Gershman S J. When does model-based control pay off. *PLoS Computational Biology*, **2016**, *12* (8): e1005090.

[22] de Leeuw J R. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, **2015**, *47*: 1−12.

[23] Daw N D, Gershman S J, Seymour B, et al. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, **2011**, *69* (6): 1204−1215.

[24] Rummery G A, Niranjan M. On-line Q-learning using connectionist systems. Cambridge: University of Cambridge, **1994**.

[25] Bürkner P C. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, **2017**, *80* (1): 1−28.

[26] Shenhav A, Botvinick M M, Cohen J D. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, **2013**, *79* (2): 217−240.

[27] Bolenz F, Profitt M F, Stechbarth F, et al. Need for cognition does not account for individual differences in metacontrol of decision making. *Scientific Reports*, **2022**, *12* (1): 8240.

[28] Castro-Rodrigues P, Akam T, Snorasson I, et al. Explicit knowledge of task structure is a primary determinant of human model-based action. *Nature Human Behaviour*, **2022**, *6* (8): 1126−1141.

[29] Eppinger B, Walter M, Heekeren H R, et al. Of goals and habits: age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience*, **2013**, *7*: 253.

[30] Walsh M M, Anderson J R. Navigating complex decision spaces: Problems and paradigms in sequential choice. *Psychological Bulletin*, **2014**, *140* (2): 466−486.

[31] Jablonska K, Stanczyk M, Piotrowska M, et al. Age as a moderator of the relationship between planning and temporal information processing. *Scientific Reports*, **2022**, *12* (1): 1548.

[32] de Wit S, van de Vijver I, Ridderinkhof K R. Impaired acquisition of goal-directed action in healthy aging. *Cognitive, Affective, & Behavioral Neuroscience*, **2014**, *14* (2): 647−658.

[33] Patzelt E H, Kool W, Millner A J, et al. Incentives boost model-based control across a range of severity on several psychiatric constructs. *Biological Psychiatry*, **2019**, *85* (5): 425−433.