Neural evidence for age-related deficits in the representation of state spaces

Alexa Ruel[1], Florian Bolenz[3,4,5], Shu-Chen Li[3,6], Adrian Fischer[7], Ben Eppinger[1,2.3]

[1] Department of Psychology, Concordia University, Montreal, H4B 1R6, Canada

[2] PERFORM center, Concordia University, Montreal, H4B 1R6, Canada

[3] Chair of Lifespan Developmental Neuroscience, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany

[4] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

[5] Cluster of Excellence "Science of Intelligence", Technische Universität Berlin, Berlin, Germany

[6] Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop", Technische Universität Dresden, Dresden, Germany

[7] Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

Running head: Adult age differences on model-based decision-making

Correspondence: alexa.ruel@mail.concordia.ca

Keywords: decision-making, reinforcement-learning, model-based, model-free, task demands, single-trial EEG, ERP

Conflict of Interest: The authors declare no conflict of interest.

Acknowledgements: This work was supported by a grant from the German Research Foundation (DFG) [SFB 940/2 B7] as well as the National Sciences and Engineering Research Council of Canada (NSERC), [funding reference number N01882] the DFG, German Research Foundation under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence" – Project ID 390523135 and the EXC 2050/1 Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" Project ID 390696704. Furthermore, this research was undertaken, in part, thanks to funding from the Canada Research Chairs program. The authors thank all participants for taking part in the study.

Abstract

When under high cognitive demand older adults tend to resort to simpler, model-free decision strategies. This age-related shift in decision behaviour has been attributed to deficits in the representation of the cognitive maps, or state spaces, necessary for more complex model-based decision-making. Yet, the neural mechanism behind this shift remains unclear. We analysed performance on a modified two-stage Markov task using a novel neurocomputational approach including computational modeling and single-trial EEG analyses to establish neural markers of age-related changes in goal-directed decision-making under different representational demands. Our results reveal that the shift to simpler decision strategies in older adults is due a) impairments in the representation of the transition structure of the task and b) a diminished signaling of the reward value associated with decision options. Consistent with the diminished state space hypothesis of human aging, our findings reveal that deficits in goal-directed, model-based behavior in older adults results from impairments in the representation of state spaces of cognitive tasks.

The ability to make goal-directed decisions that rely on mental models of the environment has been shown to decline with advancing age during adult development. However, the neural mechanisms underlying these age-related deficits in complex decision-making remain unclear. In the current study we administered a modified two-stage Markov task and used a novel neurocomputational approach which included computational modeling in combination with single-trial EEG analyses to examine the neural mechanism behind the age-related shift towards model-free behavior in older adults.

Rather than making decisions guided by a mental model of the task environment (i.e., model-based decision-making), older adults tend to engage in a simpler model-free strategy, which involves learning associations between choice actions and rewards (Bolenz, Kool, Reiter & Eppinger, 2019; Eppinger, Heekeren & Li, 2015). The shift towards model-free decision-making in older adults has been hypothesized to result from age-related difficulties in representing a mental model of the task environment (Eppinger, Kray, Mock & Mecklinger, 2008; Eppinger & Kray, 2011; Hämmerer &Eppinger, 2012), a critical prerequisite for model-based decision making. These difficulties could arise from the deterioration of prefrontal brain regions during aging (Eppinger, Walter, Heekeren & Li, 2013; Raz et al., 2005; Resnick, Pham, Kraut, Zonderman & Davatzikos, 2003) as these regions are thought contribute to representing a mental model of the task environment (Schuck et al., 2016; Wilson, Fern & Tadepalli, 2014, Vikbladh et al, 2019). Aging-related declines in dopamine modulation of the frontal-striatal network (see Li & Rieckmann, 2014 for review) may also contribute to less distinctive representations of state spaces. In line with this view, work by Wunderlich and colleagues showed that administration of L-DOPA in younger adults enhanced model-based over model-free control (Wunderlich, Smittenaar & Dolan, 2012).

In this study, our aim was twofold; (1) to determine if reducing the demands on representing a mental model of the state transitions may lead older adults to engage in greater model-based decision-making, and (2) to examine the neural dynamics underlying age-related deficits in goal-directed decision-making. To do so, we recorded electroencephalography (EEG) activity and adopted a single trial EEG regression approach (based on Fischer & Ullsperger, 2013; Fischer, Danielmeier, Villringer, Klein & Ullsperger, 2016).. This novel approach allowed us to directly examine previous hypotheses regarding the relationship between prediction errors and neural data for both age groups, and therefore provide unique insights into the neural mechanism underlying the age-related shift toward model-free learning in older adults.

To address our first aim, we had younger and older adults complete a modified two-stage Markov decision task which differed with respect to the predictability of the state transitions (see Figure 1). In the 80%-20% transition probability condition, due to the large difference between the common and rare transitions, the internal structure of the task should be simpler to represent, and thus the upcoming state should be easier to predict than in the 60%-40% condition. We hypothesized that both age groups would demonstrate a greater contribution of model-based decision-making in the low demand compared to the high demand condition. .To address the second aim, we investigate the neural dynamics underlying age-related differences in model-free and model-based decision making and focus on two components of event-related potential (ERP) that have been shown to reflect model-based and model-free decision processes (Eppinger, Walter & Li, 2017; Sambrook, Hardwick, Wills, & Goslin, 2018): the stimulus-locked P300 component and the feedback-related negativity (FRN) respectively. In line with previous work (Eppinger et al., 2017; Gläscher, Daw, Dayan & O'Doherthy, 2010), we assume that the P300 component covaries with the degree to which participants update their internal state and value representation on a trial-by-trial basis (Eppinger et al., 2017; Gläscher et al., 2010).

Next, to examine the neural mechanisms underlying feedback processing during decision-making, we focus on the FRN which has been suggested to be sensitive to negative prediction errors during reinforcement learning tasks (Holroyd & Coles, 2002; Nieuwenhuis et al., 2002; Walsh & Anderson, 2012) and more recently shown to reflect surprise (unsigned prediction errors; Cavanagh, Figueroa, Cohen & Frank, 2012; Talmi, Atkinson, El-Deredy, 2013). In contrast to the P300, we did not predict that the FRN would vary across conditions. However, as previous work has shown that the amplitude of FRN during probability reinforcement learning is attenuated in old age (Hämmerer, Li, Müller & Lindenberger, 2011), we did predicted an age group effect. We examined these hypotheses using single trial EEG regressions in order to determine if the FRN reflects trial-by-trial reward prediction errors and how this differs across age groups.

Method

Participants

Twenty-eight healthy young adults and 30 healthy older adults participated in the study. The participants were recruited through the participant pool of the Lifelab in the Chair of Lifespan Developmental Neuroscience. We excluded participants from analysis for whom 20% of reaction times were under 200 ms (1 younger adults, 3 older adults) or who showed insufficient understanding of the task during the practice trials, assessed by choice behavior (2 older adults).The final sample size therefore consisted of 26 younger adults (M_{age} = 23.73 years, *SD*= 3.08, 11 males), and 25 older adults (M_{age} = 72.32 years, *SD*= 3.36, 13 males). All participants gave informed written consent before participating. The ethics committee of the Technische Universität Dresden approved the study. Participants received a minimum payment of 5.00 euros and an additional amount (up to 8.00 euros) depending on their reward they obtained on the task.

Stimuli

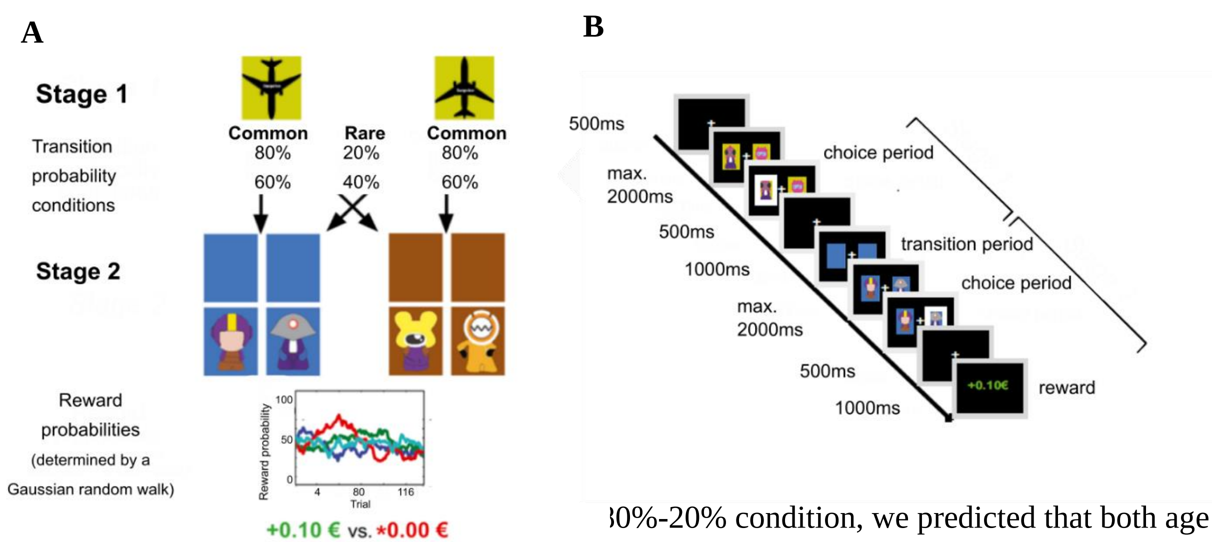
Stimuli for the first stage of the task were two airplanes: one pointing upwards and the other downwards, representing the two different choice options. Stimuli at the second stage was 8 pairs of colored figures (i.e., Gogos) which were created using free software. To avoid carryover effects, 2 new pairs of stimuli were used in each block of the task. All stimuli were further processed in Adobe Photoshop. Stimuli for the second stage of the task were on a blue or brown background, representing the two different states participants could transition to at the second stage. Feedback stimuli for each trial either indicated a reward of 10 euro cents in green or 00 euro cents in red (i.e., no reward). All stimuli were presented on a 19-inch CRT computer screen using the Eprime 2.0 software (PST Inc., Pittsburg, PA).

Task

A modified version of the two-stage Markov decision-task (cf. Daw, Gershman, Seymour, Dayan & Dolan, 2011) was administered to all participants (see Eppinger et al., 2013; Eppinger et al., 2017). All participants completed two 60%-40% transition probability blocks, and two 80%-20% transition probability blocks, corresponding to a high demand and low demand condition, respectively. Each block contained 116 trials. Block order alternated within participants and was counterbalanced across participants. Before the start of each condition block, participants were informed about the transition probability condition.

The task involved two decision stages with one decision at each stage. During the first stage, participants had to make a choice between two first-stage options (airplanes) (see Figure 1a). Depending on this choice, participants ended up in one of two possible second-stage states based on the transition probabilities: In the example illustrated in Figure 1, the left option is associated with a higher probability (i.e., common trials, 60% or 80%) of transitioning to the lower left second-stage state, and a lower probability (i.e., rare trials, 40% or 20%) of transitioning to the lower right second-stage state. The reverse is true for the right stimulus at stage one. Second-stages states were represented by two colored squares (with a separate color in each individual state) and upon these squares, the second-stage choice options (Gogo figures) were displayed between which participants made a choice. Finally, second-stage choices were either rewarded (10 euro cents) or unrewarded (0 euro cents) and the probability of receiving a reward followed from the chosen second-stage option. Consequently, at the second stage of the task, participants had to learn which of the 4 stimuli is currently associated with the highest probability of reward. However, to ensure participants continuously learn during the task, the reward probabilities fluctuated over time based on four independent Gaussian random walks with a standard deviation of 0.025 and reflecting boundaries of 0.25 and 0.75 (see Doll, Simon & Daw, 2012).

Choices options at both stages were randomly counterbalanced to appear on the left or right side of the screen. Responses at both stages were made by using the “f” key for the left option, and “j” key for the right option using a standard keyboard. Choices were presented for a response window of 2 s and after a choice was made, the chosen option was highlighted for the remaining time. If participants did not make a choice within the response window, the trial ended and the task proceeded with the next trial. The colored squares signaling the second-stages and the received reward at the end of the trial were each presented for 1 s. Prior to both decision stages and the reward, a fixation cross was displayed for 500 ms.To ensure all participants understood the task, a cover story was applied. The cover story was about a businessman who had to decide between two airline companies (i.e., two airplanes during the first stage choice of the task). Each plane would bring the businessman to one of two islands on which two populations of inhabitants could be found (i.e., the Gogos figures on the blue and brown background colors). However, the airlines are somewhat unreliable regarding their destinations, and thus end up on one island most of the time, but the other some of the time (i.e., reflecting the 80%-20% and 60%-40% transition probabilities). The task of the businessman is to make as much money as possible by tracking information about the reward probabilities of the options at the second stage of the task and the transition structure at the first stage. Importantly, participants are informed that the productivities (reward probabilities) of the populations change across time.



30%-20% condition, we predicted that both age groups should demonstrate greater model-based decision-making in this condition in contrast to the 60%-40% condition. **(b)** Trial procedure of the two-stage task.

Procedure

While the experimenter prepared the electroencephalogram (EEG), participants completed a demographic questionnaire as well as the BIS/BAS personality questionnaire (Carver & White, 1994), the identical pictures, spot a word and an N-back task. Before starting the experimental task, all participants completed a computer training session. During this training session, participants were familiarized with the reward probability structure of the task and had to perform 10 choices between options with a fixed reward probability of 60%. To help their understanding of the probabilistic structure of the task, the experimenter explained the reward probabilities in terms of absolute numbers (i.e., receiving a reward in approximately six of ten trials). Participants then completed 10 trials on which they had to find the option (out of two) with the highest reward probability. After the successful completion of these 10 training trials, the experimenter explained that the reward probabilities would change slowly across the experiment. Participants were shown two examples of the random walks (Fig 1a). In the next part of the training phase, participants were introduced to the transition probabilities connecting the first and the second stage of the task. They were explained that the task involved common and rare transitions and shown an illustration of these transitions (similar to Fig 1a). Following this explanation, participants completed ten trials in which they practiced transitioning from the first to the second stage of the task. At the end of the training session, participants completed 30 full practice trials of the experimental task which involved different stimuli than in the experimental task.

Data Analysis

Behavioral data were analysed using R (R Development Core Team, 2010), while the RL model (based on Eppinger et al. 2017) and the single trial regression analyses (based on Fischer & Ullsperger, 2013; Fischer, Danielmeier, Villringer, Klein & Ullsperger, 2016) were implemented using MATLAB (Mathworks Inc. Natick, MA). EEG data were processed and analysed in MATLAB using EEGLAB (Delorme & Makeig, 2004).

First Stage Choice Data

Stay-switch behavior was defined as the probability of repeating a choice at the first stage as a function of the transition (common, rare) and the feedback (reward, no reward) on the previous trial. Mean stay probabilities were analysed using a mixed effects logistic regression with the factors of age group (younger or older adults), condition (60%-40% or 80%-20%), previous transition (common or rare) and previous trial feedback (reward or no reward). All regressions were run using the lme4 package (Bates, Maechler, Bolker & Walker, 2013) in R (R Development Core Team, 2010).

Based on past results (Daw, Gershman, Seymour, Dayan & Dolan, 2011; Eppinger et al., 2013; Eppinger et al., 2017), a pure model-free decision strategy will be reflected by a main effect of reward (Figure 2). That is, that participants using a pure model-free strategy should be more likely to repeat choices that were rewarded on the previous trial and more likely to switch their first stage choice after non-rewarded trials. In contrast, a model-based decision strategy will be reflected by an interaction between the transition and the feedback received on the previous trial. This is the case because a pure model-based agent would be more likely to repeat choices after rewarded trials with a common transition and after non-rewarded trials with a rare transition and more likely to switch first-stage choices after rewarded trials with a rare transitions and after non-rewarded trials with a common transition.

Computational Modeling

As in previous studies (Daw, German, Seymour, Dayan & Dolan, 2011; Eppinger et al., 2013, Eppinger et al., 2017; Wunderlich et al,. 2012), each participant's choice behavior was fitted using a hybrid RL model. This model first acquires independent state-action values (Q values) for both a model-free and a model-based decision-making algorithm and then computes an integrated Q value as a weighted mean of model-based and model-free Q values. This weighting is controlled by a model-based weight which ranges from 0 to 1 and was held constant across trials for a given participant (but varies between conditions). A model-based weight closer to 0 represents a greater contribution of model-free behavior, while in contrast, a model-based weight closer to 1 reflects a greater contribution of model-based decision-making. Further, the integrated Q values were converted into choice probabilities by using a softmax function (see equation 6).

The model parameters were estimated for each condition separately and participant individually as maximum a posteriori estimates. These parameters were held constant across trials but were allowed to vary across participants and within each condition across states for the same participant. Trial by trial RPEs were extracted for each participant and were consequently used as predictors in the single trial EEG regression analyses.

Description of the computational model

In each trial *t* of the task, a participant visited two states $s_{1,t}$ and $s_{2,t}$ at the first and second stage, respectively, decided for two actions $a_{1,t}$ and $a_{2,t}$ and received a final reward r_t at state 2. We model the reward expectation held for each state-action pair as Q(s,a). Specifically, the reward expectations held by the model-free learning system are represented by $Q_{MF}(s,a)$ and the reward expectations held by the model-based learning system are represented by $Q_{MB}(s,a)$.

Model-free state-action values. Model-free state-action values were updated using SARSA (λ) temporal-difference learning (Rummery & Niranjan, 1994). After each trial, a reward prediction error δ was computed as the difference between the expected and the actual experienced reward for each of the two decision stages.

$$\delta_{1,t} = \mathrm{Q}_{\mathrm{MF}}\left(s_{2,t}, a_{2,t}\right) - Q_{MF}\left(s_{1,t}, a_{1,t}\right)$$

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

These RPEs were then used to update the state-action values according to the equation

$$Q_{MF}\ \left(s_{2,t}, a_{2,t},\right) = \ Q_{MF}\left(s_{2,t} a_{2,t}\right) + \ \alpha_2 \delta_2, \qquad (1)$$

$$Q_{MF}\left(s_{1,t}, a_{1,t}\right) = \ Q_{MF}\ \left(s_{1,t}, a_{1,t}\right) + \ \alpha_1 \delta_1 + \alpha_1 \lambda \delta_2 \qquad (2)$$

where α_i represented the learning rate at a given stage, and λ represented the eligibility trace decay.

Note that eligibility traces are not assumed to carry over from trial to trial because the task structure involved constantly changing reward probabilities (determined by the random walks) for each option across trials.

Model-based state-action values. Model-based state-action values are computed for each trial using Bellman's equation (Sutton & Barto, 1998) by taking the model-free state-action values from the second stage and the transition probabilities into account (Eq. 3 & 4):

$$Q_{MB}(s_{1,t}, a_{1,t}) = p_1 * \left[\max_a Q_{MF}(s_{2,t} = 1, a) \right] + p_2 * \left[\max_a Q_{MF}(s_{2,t} = 2, a) \right], \quad (3)$$

$$Q_{MB|S1}(a_2) = LowTran * \left[Q_{MF|S2}(a)\right] + HighTran * \left[Q_{MF|S2}(a)\right]. \quad (4)$$

The p_1 and p_2 in Eg. 3 refer to the probabilities of transitioning each of the two second-stage states after choosing action $a_{1,t}$ (80% or 60% for common transitions and 20% or 40% for rare transitions).

Finally, the action-value for the full hybrid model (Q_{net}) was calculated as the weighted sum of the model-based and model-free action-values:

$$Q_{Net} = \Omega * Q_{MB}(s_{1,t}, a_{1,t}) + (1 - \Omega) * Q_{MF}(s_{1,t}, a_{1,t}), \quad (5)$$

where Ω is the weighting parameter. At the second stage, the Q_{net} state-action value is equal to the model-free state-action value ($Q_{net|S2} = Q_{MF|S2}$).

Softmax rule. Choice probabilities at the first stage were calculated according to a softmax rule:

$$P_{S1}(a_1, t) = \frac{exp\,(\beta_1 * [Q_{net|S1}(a_1, t) + \pi * rep(a_1)])}{(expexp\,(\beta_1 * [Q_{net|S1}(a_1, t) + \pi * rep(a_1)])) + (expexp\,(\beta_1 * [Q_{net|S1}(a_2, t) + \pi * rep(a_2)]))} \quad (6)$$

where β_1 represents the inverse softmax temperature parameter which controls the distinctiveness of the choices within each stage. We allowed both learning parameters (α_1, α_2) as well as the softmax temperature parameters (β_1, β_2) to change between both stages of the task. The indicator function rep(a) is 1 if a is a top-stage action and is the same as was chosen on the previous trial and zero otherwise. Together, the rep(a) function and the parameter π capture the degree of perseveration ($\pi > 0$) or switching ($\pi < 0$) for the first stage (Lau & Glimcher, 2005).

Choice probabilities at the second stage were calculated similarly as:

$$P_{S2}(a_1, t) = \frac{exp\ (\beta_2 * Q_{net|S2}(a_1, t))}{exp\left(\beta_2 * Q_{net|S2}(a_1, t)\right) + exp\ (\beta_2 * Q_{net|S2}(a_2, t))} \tag{7}$$

The model therefore contained seven parameters (α_1, α_2, β_1, β_2, π, λ, Ω) and this set of parameters was estimated separately for each individual participant for each of the two conditions.

EEG recordings and Analysis

Two approaches were used to examine the EEG data based on our predictions; (1) ERP analyses and (2) Multiple Single-trial Robust Regressions. The ERP analysis allows for an overview of the P300 and FRN effects and how they were affected by the transition and feedback, respectively, across conditions. In contrast, the single trial analyses allow for exploration of the prediction that the P300 may be explained by the value prediction errors (VPEs) at the second stage of the task while variance in the FRN may be explained by participants' reward prediction errors (RPEs).

Pre-processing. EEG and electrooculography (EOG) were recorded continuously from 64 active Ag/AgCl electrodes embedded in an elastic plastic cap, using a BrainVision Recorder (Brain Products GmbH, Gilching, Germany). The electrodes were placed according to the international 10-10 system. During the recording, electrodes were referenced to the right mastoid, and re-referenced offline to the average of the left and right mastoids. The EEG signal was first filtered using a band pass filter in the range of 0.01 and 100Hz and were digitized with a sampling rate of 1000Hz. The ground electrode was placed above the forehead. Vertical and horizontal EOGs were recorded next to each eye and below the left eye. Electrode impedances were kept below 5kΩ.

For all statistical analyses, in addition to being re-sampled at 500Hz, a low-pass filter of 30Hz and a high-pass filter of 0.5Hz were applied to the EEG data. Next, bad channels were linearly interpolated and artifacts were rejected by visual inspection of the continuous data. Visual rejection resulted in 6.38% of trials being rejected for younger adults, and 5.83% of trials for older adult groups. Each data set was then epoched (-1 to 3s) surrounding the second stage choice and run through independent component analysis (ICA) to allow for further artifact rejection. Blinks, eye movements and muscle components determined using the ICA were marked and were rejected from further data analysis. All trials for which participants did not provide a first or a second stage response were removed from both the EEG and behavioral data.

ERP Analysis. The larger 4s epochs were then re-epoched (-200ms to 800ms) around the first stage choices, second stage stimuli and feedback phases of the task, where 0 corresponds to stimuli/feedback onset. These epochs were then baseline corrected by subtracting the average of the first 200ms pre-stimulus activity from the entire epoch (-200 to 800ms). To match trial numbers across transition probability conditions (i.e., 80%, 60%, 40% and 20%) we randomly drew a subset of trials according to the number of trials in the least frequency condition (20%) in each condition to calculate the individual participant ERP averages, resulting in 32 trials for the second stage stimulus-locked ERPs and 29 trials for the feedback-related ERPs.

Stimulus-locked ERPs at the first stage: Choice Period. We analysed two ERPs at the first stage choice period: the N200 and P300 components at Pz. The N200 was measured as the mean amplitude in a 220-320ms time window after the onset of the first stage stimuli for younger adults, and as the mean amplitude in the 230-330ms time window for older adults. The P300 was measured as the mean amplitude in a 300-400ms time window after the onset of the first stage stimuli for younger adults, and as the mean amplitude in the 350-450ms time window for older adults. These times were determined separately for both age groups by building a 100ms window around the peaks which were determined by visual inspection. We then used a repeated measures ANOVA with the between subject's factor age group (younger and older adults) and the within subject's factors condition (80%-20% and 60%-40%) and transition type (common, rare).

Stimulus-locked ERPs at the second stage: Transition Phase. The P300 component at the transition phase of the task was measured as the mean amplitude in a 300-400ms time window after the onset of the color patches during the transition phase for younger adults, and as the mean amplitude in the 410-510ms time window after stimulus onset for older adults. We also examined the transition phase effects at the frontal electrode FCz as the mean amplitude in a 300-400ms time window for younger adults, and as the mean amplitude in the 410-510ms time window after stimulus onset for older adults. These times were determined separately for both age groups by building a 100ms window around the peaks which were determined by visual inspection. To examine these components, we used a repeated measures ANOVA with the between subject's factor age group (younger and older adults) and the within subject's factors condition (80%-20% and 60%-40%) and transition type (common, rare).

Stimulus-locked ERPs at the second stage: Second stage Stimuli. The P300 component at the second stage of the task was measured as the mean amplitude in a 330- to 430-ms time window after stimulus onset (appearance of the two second stage choice stimuli on top of the colored background) for younger adults, and as the mean amplitude in the 430- to 530-ms time window after stimulus onset for older adults. These times were determined separately for both age groups by building a 100ms window around the peaks which were determined by visual inspection. To examine the mean N200 and P300 amplitudes, we used a repeated measures ANOVA with the between subject's factor age group (younger and older adults) and the within subject's factors condition (80%-20% and 60%-40%) and transition type (common, rare).

Feedback-locked ERPs. The FRN was measured as the mean amplitude in the 210- to 310-ms time window after stimulus onset for younger adults, and in the 250- to 350-ms time window for older adults. These times were determined separately for both age groups by building a 100ms window around the peaks which were determined by visual inspection. For analyzing differences in the FRN amplitude, we used a repeated measures ANOVA with the between subject's factor age group (younger and older adults) and the within subject factors condition (80%-20% and 60%-40%) and feedback (reward, no reward).

Multiple Single-Trial Robust Regressions. To examine the relationship between variables extracted from the computational model (VPE, RPE) and EEG signals we used single-trial robust regressions in a multi-level approach. We first used general linear models (GLM) to regress single-trial EEG activity at each electrode and time point against the variables extracted from the computational model. The resulting regression weights from these analyses were standardized and averaged across their respective time windows. We then used these values in set of two-way

ANOVAs to examine the effects of age group and condition on the association between computational parameters and single trial EEG signals.

Multiple Single-Trial Robust Regression of stimulus-locked ERPs at the second stage. To examine if VPEs explained significant variance in the P300 component elicited by second stage stimuli, we regressed trial-by-trial VPEs on the second stage stimulus-locked EEG data in each condition (60%-40% and 80%-20%) (based on Fischer & Ullsperger, 2013; Fisher et al., 2016). Regressions were run separately by condition to allow for the effect of condition to be examined at the second stage (described below). Both regressions (one per condition) were run at the subject level, resulting in a set of *b* values (and their associated p-values) per subject per condition. These *b* values therefore represent the average effect of the VPEs at each electrode in 10ms windows from -200 to 600ms. These first stage regressions revealed maximum amplitudes of positive going EEG activity at electrode Pz approximately 350ms after stimulus onset in the 60%-40% condition, and around 410ms after stimulus-onset in the 80%-20% condition which are consistent with the P300 component.

From the first level regressions, we extracted *b* values for each subject in each condition for the time window of interest based on the maximal effects at the first level. In order to capture the maximum amplitudes found at the first stage, we used a time window of 330-430ms. These b values from the first level regressions were standardized by their SDs before being averaged within each subject to ensure comparability between subjects. These new β weights representing the effect of VPE on the EEG signal were then used as the dependent variable in a set of two-way ANOVAs using condition and age group as predictors. This second level of analysis allowed for the examination of main effects of condition and age on the VPE β weights as well as the interaction effect between condition and age group on the VPE β weights. The main effects are therefore interpreted as an interaction between VPE and the main effect predictor on the P300 component, while a significant interaction is interpreted as a 3-way interaction.

In a separate analysis, we regressed the current trial's transition (common vs. rare) on the second stage stimulus-locked EEG data to examine the effect of transition on the signal in each condition. Analyses examining the effect of VPE and transition on the P300 component were run separately due to the collinearity between these predictors. All steps of analysis were identical to those described for the VPE regression and ANOVA. See *Supplemental Material* for details.

Multiple Single-Trial Robust Regression of Feedback-locked ERPs. To examine if RPEs explained significant variance in the FRN component elicited by feedback, we regressed trial-by-trial signed and unsigned RPEs on the feedback-locked EEG data in each condition. All steps of analysis were identical to those described for the VPE regressions. Using signed RPEs as predictor, first stage regressions revealed maximum amplitudes of positive going EEG activity at electrode Pz approximately 270ms after stimulus onset in the 60%-40% condition, and around 270ms after feedback-onset in the 80%-20% condition, consistent with the FRN component. We therefore used a 220-320ms time window for the second level analyses. In contrast, using unsigned RPEs as predictor, the first stage regressions revealed a significant age group effect, but no significant effects of condition. We therefore did not run second level analyses using unsigned RPEs.

Results

Choice Behavior

In line with previous studies (Daw et al., 2011; Eppinger et al., 2013; Eppinger et al., 2017), we examined choice behavior (stay/switch) at the first stage was analysed using condition (80%-20%, 60%-40%), previous trial transition (common, rare), the previous trial feedback (reward, no reward), as well as age group (younger adults, older adults) as predictors in a mixed-effects logistic regression (see table 1). This analysis revealed significant main effects of age group (β = -0.400, *SE* = 0.107, $p < 0.001$) and feedback (β = 0.280, *SE* = 0.046, $p < 0.001$). As shown in Figure 1a, older adults had a greater tendency to repeat their choices at the first stage than younger adults and across age groups participants were more likely to stick to their choice after rewarded compared to unrewarded trials. As expected based on previous studies, the analysis releaved a set of signficant two and three way interaction between the factors (see table 1). Most importantly we also obtained the predicted four-way interaction between the factors age group, condition, transition and feedback (β = -0.060, *SE* = 0.020, p = 0.001).

Table 1. Mixed-effects logistic regression on stay probabilities for first-stage choices.

Predictor	**β**	***SE***	***p***
(Intercept)	1.538	0.107	< 2e-16
Age group	-0.400	0.107	< 0.000
Condition	-0.110	0.043	0.010
Transition	0.054	0.029	0.060
Feedback	0.280	0.046	1.43e-09
Age group x Condition	-0.096	0.043	0.023
Age group x Transition	0.073	0.028	0.010

Condition x Transition	-0.033	0.021	0.107
Age group x Feedback	-0.061	0.046	0.186
Condition x Feedback	-0.001	0.021	0.967
Transition x Feedback	0.304	0.021	< 2e-16
Age group x Condition x Transition	-0.030	0.020	0.144
Age group x Condition x Feedback	0.020	0.021	0.330
Age group x Transition x Feedback	0.177	0.021	< 2e-16
Condition x Transition x Feedback	-0.148	0.020	3.66e-13
Age group x Condition x Transition x Feedback	-0.066	0.020	0.001

To further examine this interaction, we performed separate regression analyses for each age group. These analyses revealed significant three-way interactions between condition, transition and feedback in both age groups (*younger adults*: $\beta = -0.215$, $SE = 0.026$, $p < 0.001$, *older adults*: $\beta = -0.081$, $SE = 0.032$, $p = 0.011$; see *Supplementary Table S1* for regression by group). To further analyse these interactions, we performed analyses separately for the two age groups and conditions.

For younger adults we obtained significant transition by feedback interactions in both conditions (60%-40%: $\beta = 0.266$, $SE = 0.031$, $p < 0.001$, 80%-20%: $\beta = 0.706$, $SE = 0.042$, $p < 0.001$). In contrast, for older adults we only obtained a significant interaction between transition and feedback in the 80%-20% condition ($\beta = 0.220$, $SE = 0.050$, $p < 0.001$), but not in the 60%-40% condition ($\beta = 0.047$, $SE = 0.041$, $p = 0.247$).

Together, these results suggest that younger adults show model-based behavior in both conditions yet relied more heavily on this strategy in the 80%-20% compared to the 60%-40%

condition (β of 0.71 vs. 0.27). Older adults, in contrast, demonstrated no evidence of model-based decision-making in the 60%-40% condition but did engage in the model-based decision strategy in the 80%-20% condition (see Figure 2).

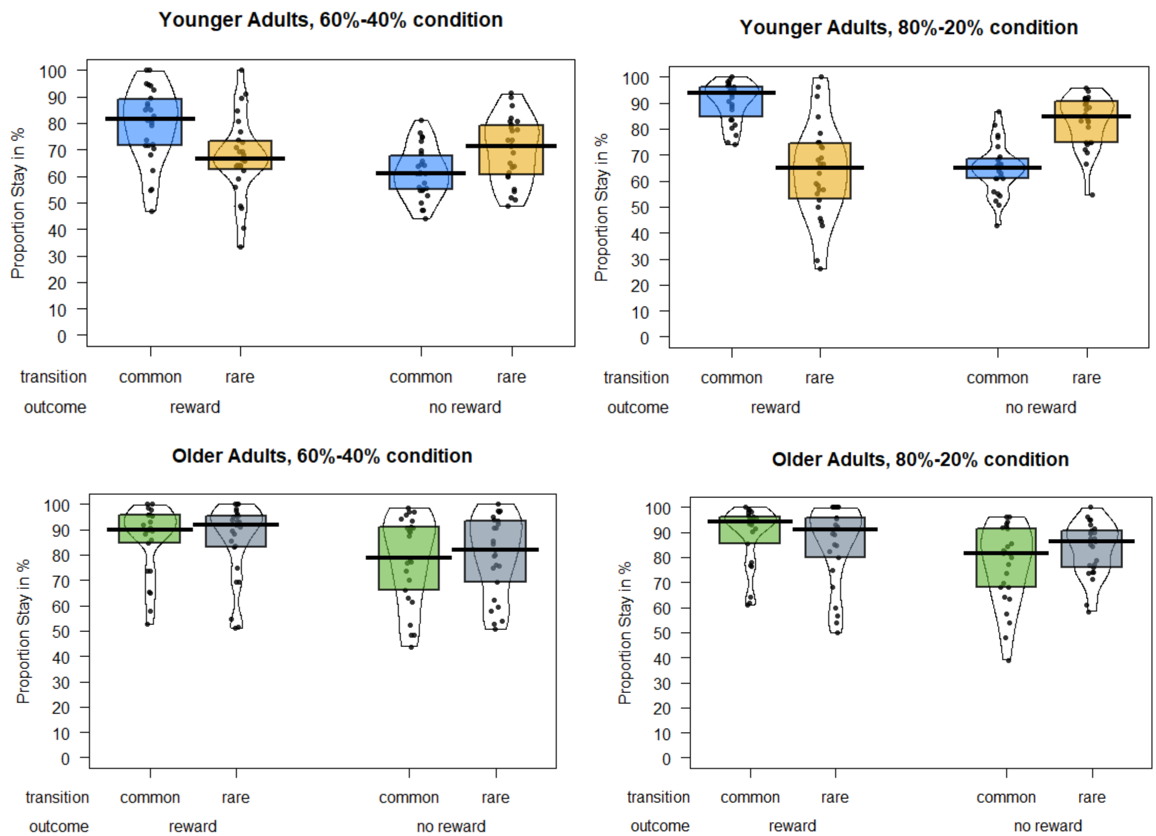


Figure 2. Probability of repeating first choice (stay behavior) as a function of the transition on the previous trial (common or rare transition) and the feedback on the previous trial (reward or no reward). Stay probabilities are displayed separately for each condition (60%-40% and 80%-20%) across both age groups (younger and older adults). Vertical black lines represent the median, while boxes represent the inter-quartile range. Black dots represent individual participants' data, and the black outline represents the overall distribution.

Computational Modeling Results

To analyse age group and condition differences in the model parameters we performed a repeated measures ANOVA. As estimated values of some of the parameters violated normality, all parameters were log transformed first. Our analyses revealed a main effect of age group for the inverse temperature parameter at the first stage (β_1 ; $F(1,49)=15.285$, $p<0.001$) and second stage (β_2; $F(1,49)=13.445$, $p < 0.001$), the learning rate at the second stage (α_2;$F(1,49)=7.391$, $p=0.009$), as well as the model-based weight (Ω;$F(1,49)=9.425$, $p=0.003$) and the choice stickiness parameter($\pi; F(1,49) = 4.398, p < 0.001$). As shown in Table 2, $\beta_1$and β_2 were higher for the younger adults than the older adults in both conditions (β_1: $F(1,49)=15.285$, $p<0.001$), β_2: $F(1,49)=13.445$, $p < 0.001$), suggesting that younger adults were better at distinguishing between the choice options at both the first and second stage. Further, younger adults demonstrated a higher second-stage learning rate than older adults (α_2; $F(1,49)=7.391$, $p=0.009$). This indicates that younger adults attributed a greater weight to the most recent trials relative to previous ones more so than older adults did. A main effect of condition was only found for the eligibility trace parameter (λ; $F(1,49) = 4.126$, $p=0.048$), revealing that the VPEs informed choices at the first stage to a greater degree in the 80%-20% condition as compared to the 60%-40% condition. Contrary to our predictions, we did not observe significant interactions between age group and condition for any of the estimated model parameters (see *Supplemental Material Table S2*).

Table 2. Model parameters in each condition for both age groups.

		α_1	α_2	β_1	β_2	π	λ	Ω

Younger Adults	**60%-40%**							
	25th percentile	0.20	0.52	4.15	2.39	0.05	0.29	0.41
	Median	0.50	0.61	6.68	3.18	0.10	0.45	0.59
	75th percentile	0.61	0.76	8.42	3.89	0.18	0.71	0.72
	80%-20%							
	25th percentile	0.20	0.49	4.44	3.21	0.09	0.49	0.42
	Median	0.37	0.63	5.83	3.93	0.13	0.64	0.60
	75th percentile	0.63	0.75	9.53	4.95	0.20	0.76	0.71
Older Adults	**60%-40%**							
	25th percentile	0.38	0.17	3.18	1.54	0.25	0.40	0.35
	Median	0.49	0.43	3.77	2.09	0.43	0.50	0.45
	75th percentile	0.60	0.59	5.14	2.53	0.63	0.65	0.61
	80%-20%							
	25th percentile	0.34	0.27	3.09	1.69	0.22	0.37	0.30
	Median	0.43	0.54	3.96	2.09	0.38	0.57	0.34
	75th percentile	0.65	0.81	5.48	2.96	0.57	0.76	0.48

Interestingly, there is an apparent mismatch between the results of the descriptive (regression) analysis and the computational modeling results. Whereas the regression analysis points to greater model-based behavior in the 80%-20% compared to the 60%-40% condition in younger and older adults with younger adults also showing model-based learning in the 60%-40% condition, the computational modeling analysis shows no age group by condition differences in the model-based weight or any other parameters. Findings by Feher da Silva & Hare (2018) show that seemingly small alterations to the two-step task can result in a discrepancy between the analysis of stay probabilities and of parameters of the computational model. Following these suggestions, one interpretation of our findings could be that the age by condition interaction that we observed in the regression analysis is due to differences in the predictability of the two task environments rather than underlying differences in model-based control. To test this prediction, we simulated behavior for three different task environments (60%-40%, 70%-30% and 80%-20%) using the same model, while keeping parameter constellations constant (for details see *Supplemental Material*). The results of the simulations are shown in *Supplemental Figure S2*. Simulation results revealed that while the model-free difference values ((common reward + rare reward) – (common no reward + rare no reward)) decrease as a function of transition probabilities ($F(2,1000) = 8.312$, $p<0.001$), consistent with our descriptive findings, the model-based values increase as a function of the transition probabilities ($F(2, 1000) = 231.8$, $p<0.001$). These findings suggest that the model-based difference values increase when the task structure becomes more predictable. However, this effect is independent of the degree of model-based control. With respect to our choice data, these simulations suggest that both younger and older adults engage in more model-based decision-making in a more predictable task environment, yet to different degrees. As previous work has shown that modifying the original two step task can lead to unexpected deviations in model-free and model-based behavior (Feher da Silva & Hare, 2018). If choices made at one second stage state are more often rewarded than choices made at the other second stage state, the first stage state choice that commonly leads to the most frequently rewarded second state will also be rewarded more often than the other first stage state. In this case, one first stage choice becomes more rewarding than the other, leading model-free agents to choose that action more often than the other, leading the stay probability for that action will be on average higher than the stay probability for the other action. This results in a model-free agent to appear model-based by with first stage choices that reveal an interaction between reward and transition. A two-way ANOVA confirmed that our task produced no significant difference in the rewards obtained across second stage state, speaking against the possibility that modifying the original two step task may have led model-free behavior to appear model-based in our descriptive analyses (see *Supplemental Material* for this analysis).

ERP Results

Stimulus-locked ERPs at the first stage: Choice Period

In line with previous work (Eppinger et al., 2017), the analysis of the first stage stimulus-locked ERPs revealed a main effect of age group for the N200 component, ($F(1,49) = 5.898$, $p = 0.020$). In addition, we found a main effect of age group ($F(1,49) = 10.343$, $p = 0.003$), and a main effect of condition ($F(1,49) = 5.582$, $p = 0.022$) for the P300 component. However, no significant interaction effects were for either of the components. These effects reflect that both the N200 and P300 were larger for younger adults as compared to older adults. Further, the P300 was larger in the 80%-20% condition for both age groups (see *Supplemental Figure S4*).

Stimulus-locked ERPs at the second stage: Transition Phase

The analysis of the transition phase ERPs at Pz revealed only a main effect of condition ($F(1,49) = 18.6239$, $p < 0.001$) showing that both age groups revealed a larger component (more positive going) in the 60%-40% condition (*M*= 2.730, *SD* = 5.924) as compared to the 80%-20% condition (*M* = 1.391, *SD* = 4.540). Transition phase ERPs at the FCz revealed a main affect of transition ($F(1,49)= 5.422$, $p = 0.024$) as well as an interaction between transition and condition ($F(1,49) = 5.045$, $p = 0.029$) (see *Supplemental Figure S5*).

Stimulus-locked ERPs at the second stage: Second stage Stimuli

The analysis of the stimulus locked ERPs at the second stage revealed a main effect of transition, indicating that larger P300 components were elicited following common compared to rare transitions ($F(1, 49) = 38.008$, $p < 0.000$). Further, we found significant interactions between transition and age group ($F(2,47) = 26.377$, $p < 0.000$) we well as between transition and transition condition ($F(2,47) = 4.296$, $p = 0.044$). To further examine the P300 effects, we ran the ANOVAs separately for the two age groups.

For the younger adults, the analysis revealed a main effect of transition ($F(1,24) = 50.073$, $p < 0.001$), as well as a significant interaction between transition and condition ($F(1, 50) = 4.3087$, $p= 0.042$). For older adults, the same analysis revealed no significant main effects or interactions (p's > 0.265) (see Fig. 3). Based on past work demonstrating that the parietal P300 component decreases with advancing age, while the frontal P300 may not (van Dinteren, Arns, Jongsma & Kessels, 2014), we ran a post-hoc ERP analysis at electrode Fz. The results revealed no significant transition by condition interaction at electrode Fz for older adults, $F(1, 50) = 0.5564$, $p= 0.459$ indicting that the absence of transition effects is not due to an overall reduced P300 response at parietal sights.

Taken together, the results show a greater P300 component in the 80%-20% condition compared to the 60%-40% condition in younger adults whereas no effect of transition probability was observed in older adults. In line with the findings from the simulated behavioral data, we interpret the larger P300 component seen in the 80%-20% condition for younger adults as evidence for an increased ability to represent the task structure in this condition. The lack of a condition effect in the older adults reveal that a more predictable environment did not help them represent the task structure.

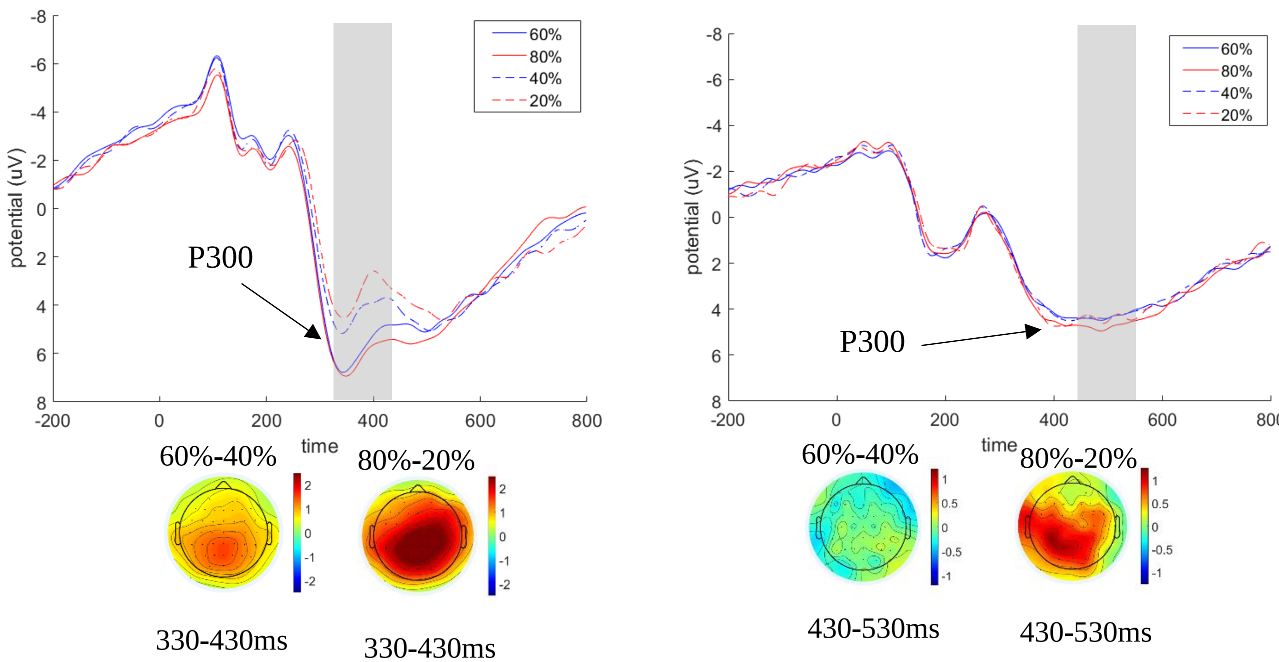


Figure 3. *Top:* ERPs elicited by second-stage stimuli at electrode Pz, displayed separately for the 80%-20% condition (red) and the 60%-40% condition (blue), as well as the common transitions (solid lines), and rare transitions (dashed lines) for both **a)** younger adults, and **b)** older adults. *Bottom:* The topographical map displays of the difference between common and

rare transitions for the 60%-40% condition (left) and the 80%-20% condition (right) for **a)** young adults and **b)** old adults.

Feedback-locked ERPs

The analysis of the feedback-related negativity (FRN) revealed main effects of feedback ($F(1,49) = 227.538$, $p< 0.000$), condition ($F(1,49) = 4.670$, $p = 0.036$) and age group ($F(1,49)= 5.561$, $p= 0.023$). Moreover, we obtained significant interactions between feedback and age group ($F(1,49)= 57.339$, $p< 0.000$) as well as between condition and age group ($F(1,49)= 8.144$, $p = 0.007$). To further analyse these effects, we examined each age group separately.

For younger adults, this analysis revealed a main effect of feedback ($F(1,24)= 193.478$, $p<0.000$) as well as a main effect of condition ($F(1,24)= 9.084$, $p = 0.006$). These main effects demonstrate that the FRN component was larger (more negative going) for no reward feedback as compared to reward feedback as well as for the 80%-20% condition as compared to the 60%-40% condition. For older adults, this analysis revealed only a main effect of feedback ($F(1,23)= 37.211$, $p< 0.001$) which reflects a greater (more negative) FRN on no reward compared to reward trials (see Fig 4).

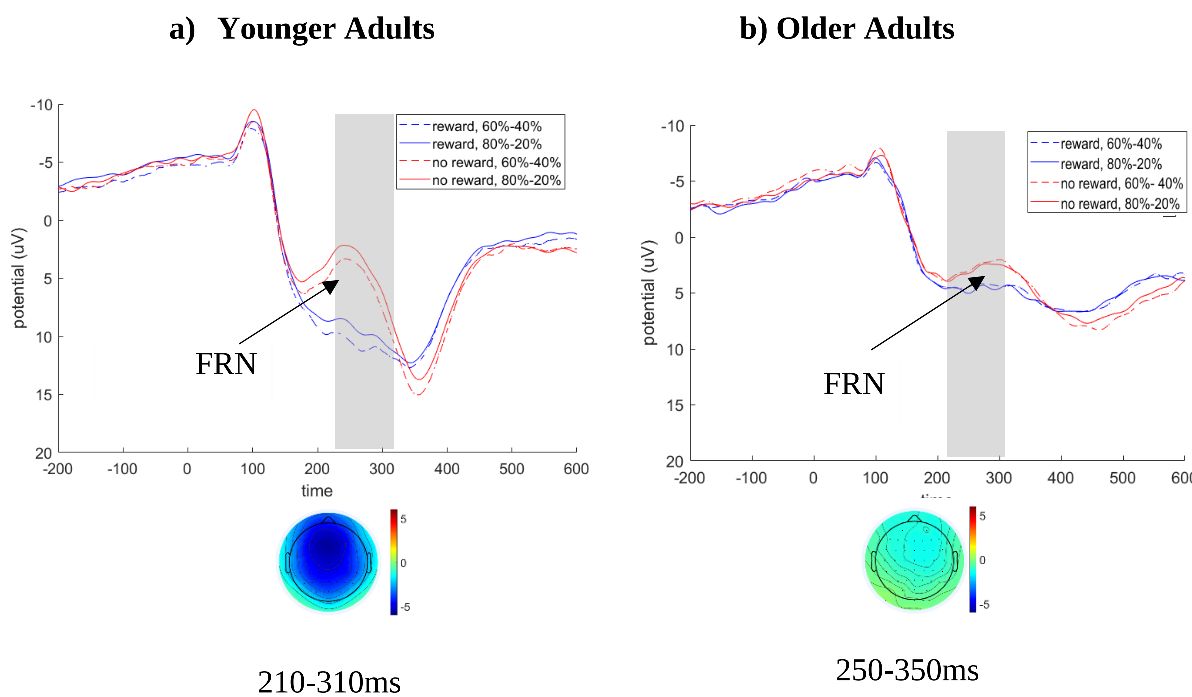


Figure 4. Top: Feedback locked ERPs at electrode FCz for rewards (blue) and no rewards (red) for **a)** younger adults, and **b)** older adults. *Bottom:* The topographical map displays of the differences between no reward and reward feedback across for **a)** younger adults, and **b)** older adults.

Single Trial EEG Regression Results

Following work by Fischer and Ullsperger (2013) and Fischer and colleagues (2016) we ran Multiple Single-Trial Robust Regression analysis on the EEG data to examine the impact of state and reward prediction errors on the corresponding event-related potentials. We then used standardized β weights from first the level analysis as the dependent variable in two-way

ANOVAs involving the within-subject factor transition condition and the between-subjects factor age group (for details see methods section).

VPE effects on the second stage P300 component. The analysis of the effects of VPE' s on the P300 component revealed a significant main effect of age group ($F(1, 49) = 10.770$, $p = 0.001$) and a significant interaction between age group and transition condition ($F(1, 49) = 5.361$, $p = 0.023$). To further examine the age group x transition condition interaction effect, we analysed each group separately.

This analysis revealed a significant main effect of transition condition for younger adults ($F(1,49)$ – 4.405, $p = 0.041$), but not older adults ($F(1,49) = 1.188$, $p = 0.281$). As shown in Figure 5, in younger adults VPEs explained significantly more variance in the P300 in the 80%-20% condition (M= 0.477, SD = 1.006) compared to the 60%-40% condition (M= -0.037, SD = 0.739). No such effect was observed in older adults.

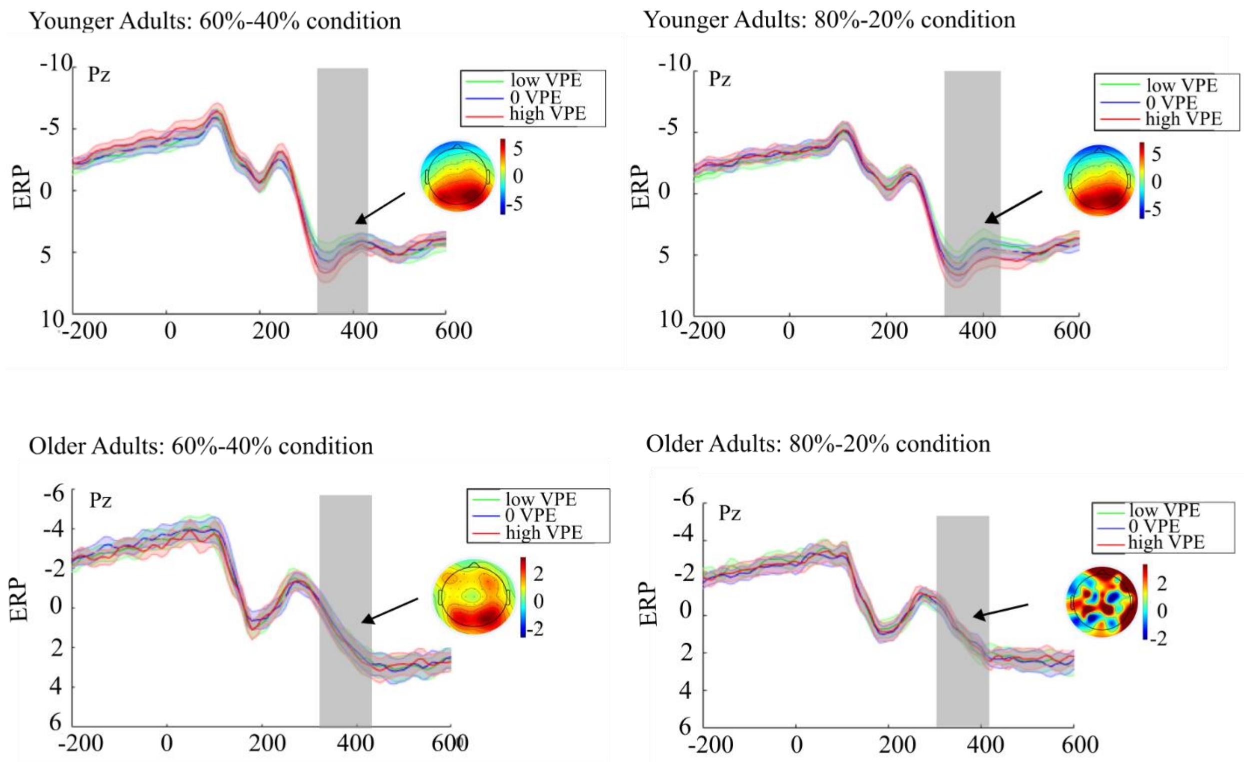


Figure 5. ERPs elicited by second-stage stimuli at electrode Pz of VPE variance at the second stage choice period displayed separately for young and older adults and 60%-40% and 80%-20% condition. VPEs were averaged in to high (red), neutral (blue) and low (green) values according to a tertiary split. The shaded areas within each plot represents the window of analysis (330-430ms). Topographies of maximum amplitudes are overlaid on each plot (350ms in the 6040 condition and 410ms in the 8020 condition).

Together, these results demonstrate that, for younger adults, the P300 component reflects VPEs which was greater in the 80%-20% condition. In contrast, no such effects were observed for the older adults.

Analysis of feedback-locked EEG data. The second level analysis of the FRN component revealed only a significant main effect of age group ($F(1,49) = 31.685$, $p < 0.001$), indicating that the RPE explained significantly more variance in the FRN for younger adults (*M*= 0.432, *SD* = 1.069) than older adults (*M*= -0.479, *SD*= 0.234). As shown in Figure 6, these findings suggest that the FRN is more tightly linked to RPE's in younger compared to older adults.

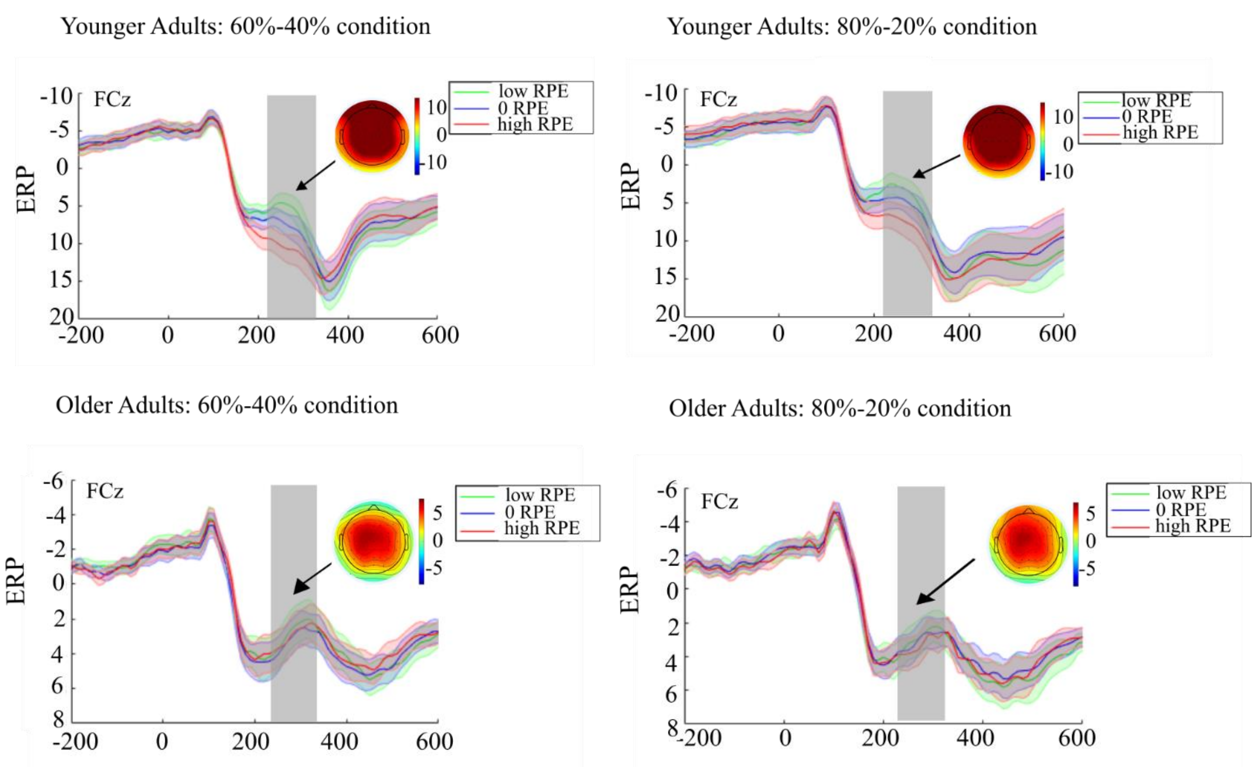


Figure 6. Feedback-locked ERPs at electrode FCz of RPE variance at the feedback stage displayed separately for young and older adults and 60%-40% and 80%-20% condition. RPEs were averaged in to high (red), neutral (blue) and low (green) values according to a tertiary split. The shaded areas within each plot represents the window of analysis (330-430ms). Topographies of maximum amplitudes are overlaid on each plot (370ms in the 6040 condition and 380ms in the 8020 condition).

Discussion

The ability to make goal-directed decisions that rely on mental models of the environment (model-based decision-making) has been shown to decline with advancing age (Bolenz et al., 2019; Eppinger et al., 2013; Eppinger et al., 2015). However, the neural dynamics underlying these age-related deficits in complex decision-making remain unclear. In the current study we used a sequential decision-making task and novel neurocomputational approach which included computational modeling in combination with ERP and single-trial EEG analyses to (1) determine whether age differences in decision-making depend on the ability to internally represent the structure of the environment (the state space of the task) and (2) establish neural markers of age-related differences in goal-directed decision-making.

Younger and older adults completed a modified version of a two-stage sequential decision task (cf. Daw et al., 2011; Eppinger et al., 2013, 2017) in which we manipulated the demands on the representation of the state space in two conditions; a high representational demand condition in which the transition probabilities between first and second stage states were less differentiated (60%/40%) and a low representational demand condition in which the transition probabilities between these states were more differentiated (80%/20%) (see Figure 1). To investigate the neural dynamics underlying age-related differences in model-free and model-based decision making, we used two different analyses: a standard ERP analysis and multiple single-trial robust regression analyses (see Fischer & Ullsperger, 2013; Fischer et al., 2016). To our knowledge, no prior study has examined age-related changes in decision-making strategies using single trial EEG regression analyses. This method allows us to directly examine previous hypotheses regarding the relationship between prediction errors and neural data for both age groups, and therefore provide unique insights into the neural mechanism underlying the age-related shift toward model-free learning in older adults.

For both analyses, we focused on two ERP components that have been shown to reflect model-based and model-free decision processes (Eppinger et al., 2017; Sambrook et al., 2018 more citations): the second-stage stimulus-locked P300 component and the feedback-related negativity (FRN). Previous work suggests that the P300 covaries with the state prediction error (SPE) which is thought to reflect the degree to which participants update their internal state space representation (Eppinger et al., 2017; Gläscher et al., 2010; , Shahnazian, Ribas-Fernandes & Holroyd 2019; Wurm, Ernst & Steinhauser, 2020). Based on these findings we examined if age-related differences in the P300 may help to explain the shift from model-based to model-free learning in older adults. Specifically, we were able to directly examine the relationship between the P300 and Value Prediction Errors (VPEs) across age groups. Next, based on previous work demonstrating the relationship between the FRN and feedback processing (Eppinger et al., 2008; Eppinger & Kray, 2011; Holroyd & Coles, 2002; Nieuwenhuis et al., 2002; Walsh & Anderson, 2012), we examined age-related differences in the relationship between the FRN component and reward prediction errors (RPEs) during learning using single trial EEG regression analyses.

In line with previous studies (Eppinger et al., 2017), an analysis of the stimulus-locked ERPs at the second stage of the task showed a larger P300 component in response to common transitions as opposed to rare transitions in younger adults. Further, in younger adults the difference between common and rare transitions was larger in the 80%-20% condition compared to the 60%-40% condition. This finding points to a more differentiated representation of the task transition structure in the 80%-20% condition in the young adults, which is in line with previous work suggesting that the P300 covaries with the degree to which participants update their cognitive map of the state space of the task (Eppinger et al., 2017). The results in older adults showed a larger P300 component for common compared to rare transitions, but no significant effect of representational demand. This absence of condition effects is consistent with recent findings, suggesting that older adults may have deficits in the adjustment of internal models of the environment (Bolenz et al., 2019; Hämmerer et al., 2019). The single-trial analyses support this view by showing that in the younger adults, trial-by-trial changes in transition type predicts changes in the P300 amplitude, whereas this is not the case in older adults (see *Supplemental Figure S1*).

In addition to age differences in state transition effects, we also studied whether older adults differed from younger adults in their ability to predict the value of the state they ended up in during the second stage of the task. We tested this with the reward prediction error elicited when transitioning to the second-stage state as derived from the computational reinforcement learning model. As shown in Figure 5, this analysis revealed that in younger adults the P300 component covaries with the VPE and that this effect is more pronounced in the 80%-20% compared to the 60%-40% condition. Thus, in the condition with a more differentiable state transition structure younger adults were better able to predict the value of the upcoming state. In older adults, we observed no such relationship between VPEs and the P300 which, consistent with the findings on the transition effects, suggests that the older adults had difficulties in representing the task structure and, as a consequence, were unable to predict the value of the upcoming states.

Next, we analysed feedback-locked ERPs to examine FRN effects across both age groups and transition conditions. Younger as well as older adults showed a larger (more negative going) component in response to unrewarded compared to rewarded feedback (see Figure 4). Further, younger adults showed larger components in the 80%-20% condition as compared to the 60%-40% condition whereas the older adults showed no condition effect.

To examine age difference in the trial-by-trial dynamics of feedback processing we performed single-trial analyses of the feedback-locked EEG. The first analysis used trial-by-trial signed reward prediction errors (RPEs) as the independent variable whereas the second analysis used unsigned RPEs. Examining both signed and unsigned RPEs allowed us to determine if the FRN covaries with the degree to which participants recognised an feedback as better or worse than predicted (signed RPE; see Ullsperger, Fischer, Nigbur & Endrass, 2014) or the degree of surprise following feedback (unsigned RPE; Cavanagh et al., 2012; Talmi et al., 2013). Results revealed significant age group effects for both signed and unsigned RPEs, a result that is consistent with several previous findings on age differences in feedback processing (see Eppinger et al., 2008; Eppinger et al.,2013). Interestingly, trial-by-trial signed RPEs but not unsigned RPEs explained significant variance in the FRN for both younger and older adults across conditions. Thus, our findings seem in line with previous work on the FRN (Fischer & Ullsperger, 2013; Talmi et al., 2012; Walsh & Anderson, 2012), showing its involvement in feedback processing and model-free decision-making. As only the signed RPEs explained variance in the FRN, this component appears to reflect participants' predictions regarding the feedback being worse (negative RPEs) or better (positive RPEs) than they predicted (Ullsperger et al., 2014).

Together, our EEG results corroborate past research suggesting that (1) the P300 component reflects the updating of the internal task representation and that (2) the FRN reflects signed prediction error signals involved in reinforcement learning. Further, we demonstrate that more differentiable state transitions (i.e, 80%-20% condition) make the decision-making environment more predictable and therefore support younger adults in building an internal representation of the task structure, as reflected in the P300 component. In contrast, older adults show a reduced neural representation of the state transition structure as well as the value of the upcoming decision options. Together, these findings provide empirical evidence in line with previous suggestions that older adults may have age-related deficits in the representation of the transition structure or state space of the task (see Bolenz et al., 2019).

Although the results of the ERP analyses provide a compelling and consistent set of findings, several open questions, particularly with respect to the behavioral results, remain. The descriptive (regression) analyses of the choice data are largely consistent with our EEG results. As shown in Figure 2, lower demands on the representation of the transition structure in the 80%-20% condition may have led to greater model-based behavior as compared to the high demand condition (60%-40%) for both age groups. Further, while younger adults demonstrated a significant contribution of model-based learning to learning in the 60%-40% condition, older adults showed no significant model-based learning in this condition. Thus, according to the descriptive analyses it appears that the 80%-20% condition led to a greater contribution of model-based decision-making for both age groups. Based on these results one could predict a greater model-based weight in the low compared to the high representational demand condition. Yet, the computational modeling results reveal a different picture.

As in previous studies, we used a hybrid RL model (see Daw et al., 2011; Eppinger et al., 2013, Eppinger et al., 2017) which was fit to task behavior. Consistent with previous work, our results demonstrate that older adults demonstrated less model-based decision-making than younger adults. Yet, in contrast to our predictions and the results of our descriptive analyses, we found no age group by condition interaction effect for the model-based weight or any other model parameter, indicating that the 80%-20% condition did not lead to an increase in model-based decision-making for either age group.

To better understand the relationship between our descriptive results and computational modeling findings, we ran a simulation across three transition probabilities (60%-40%, 70%-30% and 80%-20%) using fixed computational modeling parameter values. We also ran a posterior predictive check on our descriptive data to see if our model can recover the effects seen in the descriptive data. Together, as reflected in the descriptive analyses, model-based behavior did increase as the transition probabilities became more distinct, regardless of the degree of model-based control (see *Supplemental Figure S1*). This is in line with findings by Feher da Silva & Hare (2018) showing that seemingly small changes to the two-stage task can lead to a discrepancy between the analysis of stay probabilities and the analysis of computational model parameters. In light of the P300 analyses and the descriptive results, we propose that the increase in model-based behavior seen in the descriptive results for the 80%-20% condition as compared to the 60%-40% condition is a consequence of the greater predictability of the environment rather than a change in decision strategies per se.

One possible explanation for this result could be that participants may have noticed that the greater effort required to engage in a goal-directed strategy does not lead to more reward in the 80%-20% compared to the 60%-40% condition. Previous work has shown that the willingness to engage in model-based behavior depends on whether this (more effortful) behavior leads to greater overall payouts in the task (Kool, Cushman & Gershman, 2016; Bolenz et al., 2019). Further, when the task requires greater cognitive effort, younger adults have been shown to reduce the degree to which they engage in model-based learning (Bolenz et al., 2019; Kool, Gershman & Cushman, 2017). Consequently, it seems plausible that particularly the younger participants were not enticed to engage in more effortful model-based behavior in the low demand condition, Whether greater incentives in combination with strongly reduced demands on the representation of the task structure would lead to greater model-based behavior in older adults is unclear. Recent findings from our group speak against this view. The task applied in the study by Bolenz et al. (2019) is a strongly simplified version of the current paradigm with deterministic transitions between states. Still, older adults show reduced overall model-based behavior and no adjustment of learning strategies as a function of incentives. This finding suggests that it may not be the probabilistic nature of the transitions between states but the complexity of the state space (the number of states and possible transition) that determines age-related deficits. This and the question about age differences in the neural systems involved in model-based decision-making need to be answered in future studies.

To summarize, making the state transition structure of the task environment more predictable seems to support goal-directed behavior in both age groups, albeit to different degrees. However, this change in decision behaviour does not coincide with a shift in decision strategies. Rather, the more predictable environment allowed participants to build a more differentiated representation of the of the state transition structure. The results of the single-trial EEG analyses are consistent with the behavioral findings, suggesting that (1) P300 component reflects the updating of the internal task representation and that (2) the FRN reflects signed prediction error signals involved in reinforcement learning. Further, through the relationship between VPEs and the P300 component, we provide empirical support for the idea that more differentiable state transitions help younger adults in building up an internal representation of the task structure that they can use to predict the value of the upcoming states. In contrast, the absence of such an association in older adults suggests that they have deficits in the representation of the state transition as well as in the representation of the value of the upcoming decision options. To conclude, our findings point to age-related deficits in the representation of the state space of the task that lead to diminished value predictions in older adults (see Bolenz et al., 2019).

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bolenz, F., Kool, W., Reiter, A. M., & Eppinger, B. (2019). Metacontrol of decision-making strategies in human aging. *Elife*, *8*, e49154.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of personality and social psychology*, *67*(2), 319.

Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2012). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral cortex*, *22*(11), 2575-2586.

Feher da Silva, C., & Hare, T. A. (2018). A note on the analysis of two-stage task results: How changes in task structure affect what model-free and model-based strategies predict about the effects of reward and transition on the stay probability. *PloS one*, *13*(4), e0195328.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204-1215.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current opinion in neurobiology*, *22*(6), 1075-1081.

Eppinger, B., & Kray, J. (2011). To choose or to avoid: age differences in learning from positive and negative feedback. *Journal of Cognitive Neuroscience*, *23*(1), 41-52.

Eppinger, B., Kray, J., Mock, B., & Mecklinger, A. (2008). Better or worse than expected? Aging, learning, and the ERN. *Neuropsychologia*, *46*(2), 521-539.

Eppinger, B., Walter, M., Heekeren, H. R., & Li, S. C. (2013). Of goals and habits: age-related and individual differences in goal-directed decision-making. *Frontiers in neuroscience*, *7*, 253.

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

Eppinger, B., Walter, M., & Li, S. C. (2017). Electrophysiological correlates reflect the integration of model-based and model-free decision information. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(2), 406-421.

Fischer, A. G., & Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, *79*(6), 1243-1255.

Fischer, A. G., Danielmeier, C., Villringer, A., Klein, T. A., & Ullsperger, M. (2016). Gender influences on brain responses to errors and post-error adjustments. *Scientific reports*, *6*(1), 1-11.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585-595.

Hämmerer, D., & Eppinger, B. (2012). Dopaminergic and prefrontal contributions to reward-based learning and outcome monitoring during child development and aging. *Developmental psychology*, *48*(3), 862.

Hämmerer D., Li, S.-C., Müller, V., & Lindenberger, U. (2011). Life span differences in electrophysiological correlates of monitoring gains and losses during probabilistic reinforcement learning. *Journal of Cognitive Neuroscience*, *23*, 579-592.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, *109*(4), 679.

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological science*, *28*(9), 1321-1333.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control pay off?. *PLoS computational biology*, *12*(8), e1005090.

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the experimental analysis of behavior*, *84*(3), 555-579.

Li, S.-C. & Rieckmann, A. (2014). Neuromodulation and aging: implications of aging neuronal gain control. *Current Opinion in Neurobiology*, *28*, 148-158.

Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G., Holroyd, C. B., Kok, A., & Van der Molen, M. W. (2002). A computational account of altered error processing in older age: dopamine and the error-related negativity. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(1), 19-36.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., ... & Acker, J. D. (2005). Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cerebral cortex*, *15*(11), 1676-1689.

Resnick, S. M., Pham, D. L., Kraut, M. A., Zonderman, A. B., & Davatzikos, C. (2003). Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *Journal of Neuroscience*, *23*(8), 3295-3301.

Rummery, G. A., & Niranjan, M. (1994). *On-line Q-learning using connectionist systems* (Vol. 37, p. 20). Cambridge, UK: University of Cambridge, Department of Engineering.

Sambrook, T. D., Hardwick, B., Wills, A. J., & Goslin, J. (2018). Model-free and model-based reward prediction errors in EEG. *NeuroImage*, *178*, 162-171.

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human orbitofrontal cortex represents a cognitive map of state space. *Neuron*, *91*(6), 1402-1412.

Shahnazian, D., Ribas-Fernandes, J. J., & Holroyd, C. B. (2019). Electrophysiological correlates of state transition prediction errors. *bioRxiv*, 544551.

Smittenaar, P., FitzGerald, T. H., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron*, *80*(4), 914-919.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (Vol. 135). Cambridge: MIT press.

Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *Journal of Neuroscience*, *33*(19), 8264-8269.

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in cognitive sciences*, *18*(5), 259-267.

van Dinteren, R., Arns, M., Jongsma, M. L., & Kessels, R. P. (2014). Combined frontal and parietal P300 amplitudes indicate compensated cognitive processing across the lifespan. *Frontiers in aging neuroscience*, *6*, 294.

Vikbladh, O. M., Meager, M. R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., ... & Daw, N. D. (2019). Hippocampal contributions to model-based planning and spatial memory. *Neuron*, *102*(3), 683-693.

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews*, *36*(8), 1870-1884.

Wilson, A., Fern, A., & Tadepalli, P. (2014). Using trajectory data to improve bayesian optimization for reinforcement learning. *The Journal of Machine Learning Research*, *15*(1), 253-282.

Wunderlich, K., Smittenaar, P., & Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron*, *75*(3), 418-424.

Wurm, F., Ernst, B., & Steinhauser, M. (2020). The influence of internal models on feedback-related brain activity. *Cognitive, Affective, & Behavioral Neuroscience*, *20*(5), 1070-1089.