Computational Brain & Behavior
https://doi.org/10.1007/s42113-024-00198-5

ORIGINAL PAPER

Externally Provided Rewards Increase Internal Preference, but Not as Much as Preferred Ones Without Extrinsic Rewards

Jianhong Zhu[1] · Kentaro Katahira[2] · Makoto Hirakawa[1] · Takashi Nakao[1]

Accepted: 28 January 2024

Abstract
It is well known that preferences are formed through choices, known as choice-induced preference change (CIPC). However, whether value learned through externally provided rewards influences the preferences formed through CIPC remains unclear. To address this issue, we used tasks for decision-making guided by reward provided by the external environment (externally guided decision-making; EDM) and for decision-making guided by one's internal preference (internally guided decision-making; IDM). In the IDM task, we presented stimuli with learned value in the EDM and novel stimuli to examine whether the value in the EDM affects preferences. Stimuli reinforced by rewards given in the EDM were reflected in the IDM's initial preference and further increased through CIPC in the IDM. However, such stimuli were not as strongly preferred as the most preferred novel stimulus in the IDM (superiority of intrinsically learned values; SIV), suggesting that the values learned by the EDM and IDM differ. The underlying process of this phenomenon is discussed in terms of the fundamental self-hypothesis.

Keywords Reinforcement learning (RL) · Choice-based learning (CBL) · Computational modeling · Reward learning task · Self-prioritization effect

Introduction

We make many decisions every day. Depending on the situation, decisions are made based on criteria given by the external environment (such as a monetary reward), known as externally guided decision-making (EDM), or on one's own internal criteria (such as a sense of value, beliefs, and preferences), referred to as internally guided decision-making (IDM). For years, EDM and IDM have been studied as separate research areas. Although comparative studies on EDM and IDM have been reported in recent years, these have primarily focused on their differences (Nakao et al., 2012, 2013, 2016, 2019; Ugazio et al., 2021; Wolff et al., 2019). The EDM and IDM differ in their conceptual definition, experimental operation, and neural bases (Nakao et al., 2012, 2013, 2016, 2019; Wolff et al., 2019). However, they are similar in that the option's values are learned through decision-making (Akaishi et al., 2014; Lee & Daunizeau, 2020; Nakao et al., 2016, 2019; Zhu et al., 2021).

In EDM, the value of an option is considered to be updated based on the predicted value and actual feedback given after the decision-making (Behrens et al., 2007; Biele et al., 2011; Gluth et al., 2014; Hauser et al., 2015; Katahira et al., 2011; Lindström et al., 2014; O'Doherty et al., 2007). For example, in a reward learning task where one item is chosen from two items and each is rewarded with a certain probability, the item's value increases when rewarded feedback is received (Behrens et al., 2007; Hauser et al., 2015).

In IDM tasks, such as a preference judgment task, no externally delivered feedback indicates a correct answer. Even in such a case, an item's value (preference) changes with the choice itself, not based on the feedback (Brehm, 1956; Colosio et al., 2017; Miyagi et al., 2017). More specifically, in the preference judgment task of choosing a preferred item from two presented items, the value of the chosen item increases, while the value of the rejected item

✉ Jianhong Zhu
zhujianhong1995@gmail.com

[1] Graduate School of Humanities and Social Sciences, Hiroshima University, 1-1-1 Kagamiyama, Higashihiroshima, Hiroshima 739-8524, Japan

[2] Human Informatics and Interaction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST) Tsukuba, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan

Published online: 21 February 2024

decreases (choice-induced preference change, CIPC (Brehm, 1956; Zhu et al., 2021)).

The decision-making processes of the EDM (Daw & Doya, 2006; Gläscher et al., 2010; Schönberg et al., 2007; Sugawara & Katahira, 2021) and IDM (Akaishi et al., 2014; Lee & Daunizeau, 2020; Nakao et al., 2016, 2019) are, therefore, similar in that the values are updated based on decisions and their consequences. Value learning processes in decision-making have been mathematically modeled using a reinforcement learning (RL) model for EDM and a choice-based learning (CBL) model based on the RL model for IDM.

In the standard RL model for EDM, the choice behavior is guided by the expected value (e.g., \$0.80 is the expected value of \$1.00 dollar being compensated with an 80% probability) associated with the option. The expected value is updated according to the prediction error (i.e., the difference between the provided reward and the expected value) (Daw & Doya, 2006; Dayan & Abbott, 2001; Dayan & Balleine, 2002; Sutton & Barto, 1998). The suitability of the computational model has been tested by fitting the model to the trial-by-trial choice of behavioral data. The model-based analysis can be used to estimate model parameters, such as learning rate (the degree of value update), and latent variables, such as the expected value and prediction error of each trial. Furthermore, these estimated parameters and latent variables have been used in neuroscience to explain the neural substrates of EDM (Brehm, 1956; Hampton et al., 2008; Niv et al., 2012).

For IDM, although CIPC has been examined for many years using changes in subjective preference ratings as an index (Izuma et al., 2010; Koster et al., 2015; Miyagi et al., 2017; Nakamura & Kawabata, 2013), in recent years, CBL models have been proposed as the computational model for the value learning process behind the CIPC phenomenon in IDM (Akaishi et al., 2014; Lee & Daunizeau, 2020; Nakao et al., 2016, 2019). CBL models are based on the RL model, but unlike RL, they use choice behavior itself instead of feedback from the external environment. In addition, unlike the typical RL, a CBL model updates the value of both chosen and rejected items (Zhu et al., 2021). Thus, in the CBL model, chosen items are treated as correct answers, while rejected items are treated as incorrect answers to update their value. Although the EDM and IDM differ in the kinds of choice results they have, they update values based on feedback from the external environment or their own choice. Therefore, they are similar in that they update values based on the difference between expected values and the results of choice.

Not only do EDM and IDM have similar computational models of value learning, but it has been reported that reward-related neural activities observed in EDM (Bechara et al., 1997, 2005; Marco-Pallarés et al., 2008, 2015; Yacubian et al., 2006) are also involved in the value learning processes in IDM (Aridan et al., 2019; Camille et al., 2011; Fellows & Farah, 2007; Izuma et al., 2010; Miyagi et al., 2017; Nakao et al., 2016). This suggests that similar reward-related neural activity contributes to value learning in the EDM and IDM. This means that the EDM and IDM are not completely independent decisions, and the values learned in these two types of decisions may not be distinguishable. Furthermore, it is possible that the values learned in the EDM are used in the IDM and vice versa. However, these previous studies examined the EDM and IDM independently, and the similarity of the neural basis did not provide sufficient evidence for the mutual influencing of values between the EDM and IDM. Human decision-making processes are complex. In everyday life, decisions are likely to be made with reference to both external and internal criteria. Understanding the relationship between values in external and internal criteria is the first step in understanding integrated decision-making processes.

This study investigated whether, how, and to what extent the value learned in EDM affects the value in IDM. We used simple EDM and IDM tasks with novel contour shapes; the IDM task followed the EDM task. In the IDM task, the same stimuli used with the EDM task were used for the four (two high and two low reward probability) stimuli in addition to eleven novel stimuli.

We first tested whether the values learned in EDM affect IDM from classical model-free behavioral data analysis using the chosen frequency of items in IDM (Nakao et al., 2013, 2019). If the value learned in the EDM affects the IDM, it is predicted that stimuli with higher values learned in the EDM will be chosen more frequently than novel stimuli in the IDM. It is also expected that stimuli with lower values learned in the EDM will be chosen less frequently than novel stimuli in the IDM.

To examine how the values in EDM reflected IDM, we applied four computational models (see Table 1 in the "Methods" section) to the IDM data, compared the models, and investigated the estimated initial values of each stimulus

Table 1 The settings of initial values of each stimulus type in different CBL models

Model	Novel	HP	LP
Model 1	η		
Model 2	0.5	η_{HP}	0.5
Model 3	0.5	0.5	η_{LP}
Model 4	0.5	η_{HP}	η_{LP}

η denotes the free parameter for the initial value in IDM. HP and LP denote high-probability reward stimuli (90%, 80%) and low probability reward stimuli (20%, 10%) in the EDM task, respectively

in the IDM. These four models are based on differences in the initial values of the three types of stimuli in the IDM, which can confirm whether the value learned in the EDM is reflected in the initial values of any type of stimulus. Model 1, in which the initial values of all stimuli were the same, indicated that the values learned in the EDM did not affect the IDM. Model 2, in which only the initial value of the high-reward probability stimulus differed from the other stimuli, indicated that the high value learned in the EDM affected the IDM. Model 3, in which only the initial value of the low-reward probability stimuli differed from the other stimuli, shows that the low value learned in the EDM affects the IDM. Model 4, in which the initial values of all types of stimuli were different, illustrates that high and low values learned in the EDM affect the IDM. If the value in the EDM is reflected in the initial value in the IDM, we predict that other models will fit behavioral data better than Model 1. In addition, if both high and low values in the EDM were reflected in the IDM, then Model 4 would be the best fit for this behavior. For the computational model, simulations were conducted to confirm that the model parameters could be adequately estimated (parameter recovery) and that an appropriate model could be selected (model recovery) before the behavioral data analyses. In addition, the best-fit model was tested to ensure its accuracy in describing the behavioral data as a posterior predictive check (Gelman et al., 1996).

To further investigate how the value learned in EDM affects IDM, we explored whether the values learned in EDM affect the degree of value change in IDM. To examine the value changes from initial to final, we used values estimated by models that best fitted the behavioral data. Stimuli that reflect high or low values in the EDM are expected to be chosen or rejected more frequently in the IDM, respectively, thereby further changing their values. As all four models described above with different initial value settings assumed value changes in the IDM based on a previous study (Zhu et al., 2021), an additional computational model analysis was performed to rule out the possibility that the value did not change in the IDM.

Finally, by synthesizing the effects of the EDM on the initial value and the value change in the IDM, the effect of the EDM on the IDM was examined. To this end, we compared the final values of the IDM between the EDM stimuli and all the novel stimuli. If the influence of the value in the EDM is dominant, then it is expected that stimuli with high values in the EDM will ultimately be the highest-value stimuli in the IDM. Similarly, a stimulus with a low value in the EDM is expected to have the lowest final value in the IDM. To confirm the validity of the final values estimated by the computational model, we examined the consistency between the final values estimated by the model and the chosen frequency of each stimulus in the IDM or the subjective ratings of each stimulus after the IDM task.

We confirmed whether the participants successfully learned the value of each stimulus in the EDM from the correct response rate. Therefore, in this study, a computational model analysis of the EDM behavioral data is not necessary. For reference, we report the results of the RL model analysis for EDM in the supplementary materials.

Results

Correct Response Rate in EDM Task

We confirmed that participants learned the value of stimuli through EDM by investigating the results of the correct response rate (Fig. 1a). The correct response rate of the 80% vs. 20% stimulus pair was 0.698 (*SD* [*Standard Deviation*] = 0.093), and the 90% vs. 10% stimulus pair was 0.830 (*SD* = 0.092), both of which were higher than chance level (0.5) ($ts(37) > 13.124$, $ps < 0.001$, 95% CI = 0.667, 0.728 in 80% vs. 20% stimulus pair, 95% CI = 0.800, 0.861 in 90% vs. 10% stimulus pair). In addition, there was also a significant difference between these two conditions ($t(37) = 8.189$, $p < 0.001$, $d = 1.434$, 95% CI = 0.100, 0.165).

Chosen Frequency in IDM Task

To examine whether the values learned in EDM affect IDM, we examined the chosen frequency of each stimulus type in IDM (Fig. 1b). In the IDM task, there was no difference in the frequency of choice between the 90% (chosen frequency = 0.675) and 80% (chosen frequency = 0.645) stimuli presented in EDM ($t(37) = 0.552$, Holm-adjusted $p = 1.000$, $d = 0.110$, 95% CI = −0.080, 0.140), or between the 10% (chosen frequency = 0.445) and 20% (chosen frequency = 0.485) stimuli presented in EDM ($t(37) = -0.601$, Holm-adjusted $p = 1.000$, $d = -0.130$, 95% CI = −0.173, 0.094). Therefore, the 90% and 80% stimuli were grouped together as high-probability (HP) stimuli, and the 10% and 20% stimuli as low-probability (LP) stimuli.

The mean chosen frequency and standard deviation (*SD*) of the three types of stimuli were 0.660 (0.213), 0.465 (0.222), and 0.477 (0.064) for HP, LP, and novel stimuli, respectively. Participants preferred to choose HP stimuli as the preferred item than novel stimuli and LP stimuli ($ts(37) > 4.211$, Holm-adjusted $ps < 0.001$, $ds > 0.886$). However, there was no significant difference between their preference for the LP and novel stimuli ($t(37) = -0.268$, Holm-adjusted $p = 0.790$, $d = -0.073$, 95% CI = −0.518, 0.372). These results showed that only the high value of stimuli learned in the EDM task influenced the IDM task.

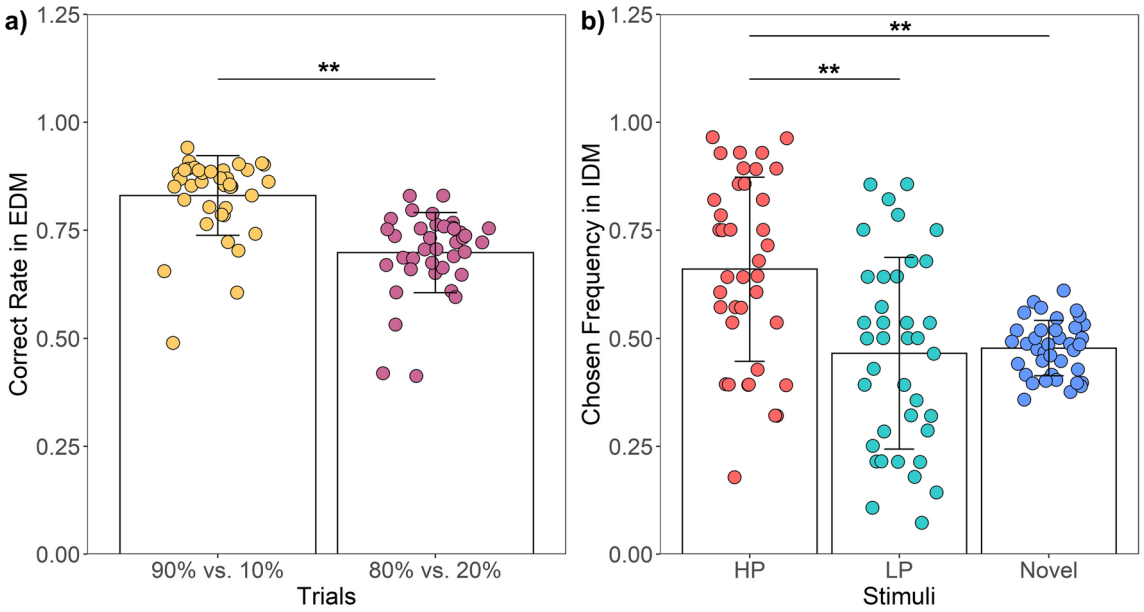

Fig. 1 Results of model-free behavioral data indicators. **a** Mean correct response rate in the EDM task. The easier condition was stimulus pairs with a 90% vs. 10% probability of obtaining a reward, and the harder condition was stimulus pairs with an 80% vs. 20% probability of obtaining a reward. **b** Mean chosen frequency of each stimulus type in the IDM task. The error bars and colored dots of all figures indicated *SD* and each participant's data, respectively. ** $p < .001$

Simulation 1 (Parameter Recovery)

Figure 2 shows the results of parameter recovery for each of the four CBL models (Table 1). We confirmed strong consistency between the set parameter values (simulated) to generate artificial behavioral data and the estimated values (fitted) by fitting the model generating the data in all CBL models ($rs > 0.664$).

As shown in Fig. 2, the range of model parameters used to generate the artificial data and the range of parameters estimated by fitting the model to the generated artificial data were generally consistent. Specifically, the ranges of model parameters generated to generate the artificial data were 0.010 to 0.577 for α, 0.010 to 20 for β, and 0.010 to 1 for η, except for Model 4. The ranges of α of Model 4 were 0.010 to 0.704; the ranges of β and two ηs were the same as the other models. The range of parameters estimated by fitting the models to the generated artificial data is as follows: in Model 1, α was 0.012 to 0.497, β was 0.516 to 15.169, and η was 0.075 to 0.884; in Model 2, α was 0.008 to 0.514, β was 0.363 to 17.908, and η was 0.200 to 0.812; in Model 3, α was 0.005 to 0.501, β was 0.285 to 15.385, and η was 0.230 to 0.836; in Model 4, α was 0.003 to 0.633, β was 0.157 to 19.442, and two η s were 0.251 to 0.707 and 0.067 to 0.921.

Simulation 2 (Model Recovery)

Table 2 presents the widely applicable Bayesian information criterion (WBIC) (Watanabe, 2013) as an index of the relative goodness of fit estimated when data generated by the individual models were fitted to all models. In the comparison of the WBICs, the smaller the value of the WBIC, the higher the fit of the generated data to the model. As shown in Table 2, all models could complete model recovery well. Whichever model generated the data had the highest match to the same model.

We calculated the Bayes factor (BF) to compare which model had the higher probability of generating data. When Model *X* was the true model (i.e., the model used to generate artificial data), the result of the BF between Models *X* and *Y* was shown by BF_{XY}. The marginal likelihood of Model *X* and *Y* was the numerator and denominator, respectively. When Model 1 was the true model, the evidence was not as strong for Model 1 when compared to Model 3 ($BF_{13} = 2.740$), indicating that the probability of data generated by Model 1 was 2.740 times that of Model 3. However, there was very strong evidence for Model 1 when compared to Models 2 and 4 ($BF_{12} = 9000.181$, $BF_{14} = 4.386 \times 10^6$). When Models 2 and 3 were the true models, there were very strong supports for each model compared to the other models ($BF_{21} = 4.254 \times 10^7$, $BF_{23} = 6.045 \times 10^{10}$, $BF_{24} = 5.505 \times 10^4$, $BF_{31} = 2.409 \times 10^{10}$, $BF_{32} = 4.664 \times 10^{12}$, $BF_{34} = 2.114 \times 10^4$). When Model 4 was the true model, strong evidence was found for comparison with Model 2 ($BF_{42} = 51.213$), and very strong evidence for comparison with Models 1 and 3 ($BF_{41} = 2.244 \times 10^{13}$, $BF_{43} = 4.706 \times 10^7$).

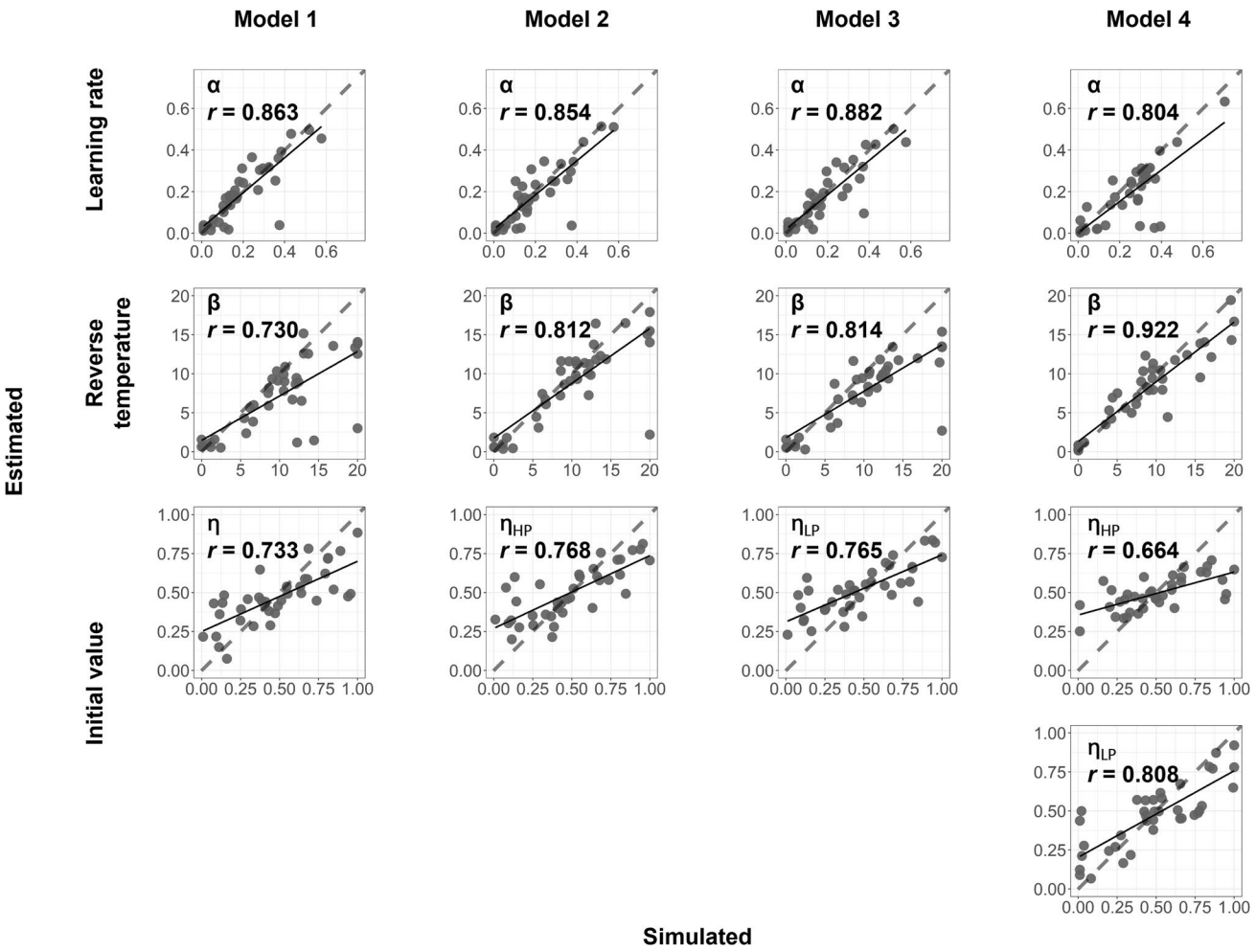

Fig. 2 Results of parameter recovery simulation. This simulation was conducted to confirm whether each model could be well estimated as the set value of each parameter. The correlation coefficient between simulated and fitted was shown as parameter recovery indices

Table 2 Results of WBIC for model recovery for Models 1–4 shown in Table 1

Simulated	Fitted			
	Model 1	Model 2	Model 3	Model 4
Model 1	**2135.055**	2144.160**	**2136.063**	2150.349**
Model 2	2100.200**	**2082.634**	2107.459**	2093.550**
Model 3	2113.086**	2118.352**	**2089.181**	2099.140**
Model 4	2103.990**	2077.184*	2090.915**	**2073.248**

Bold numbers in the table represent the WBIC values of the models that best fit the artificial data. Asterisk (*) is the BF value used for model comparison ($20 < {}^{*}BF < 150$, $150 < {}^{**}BF$). The larger the BF value, the greater the difference between the models

Table 3 WBIC results of IDM behavioral data fit with Models 1–4 shown in Table 1

Model	WBIC
Model 1	2189.52**
Model 2	**2174.51**
Model 3	2194.00**
Model 4	2185.79**

Bold numbers in the table represent the WBIC values of the model that best fit the behavioral data. Asterisk (*) is the BF value used for model comparison ($150 < {}^{**}BF$). The larger the BF value, the greater the difference between the models

Model Fit to Behavioral Data

To examine how the EDM values reflect the IDM, we fit the four models to actual IDM behavioral data. The results shown in Table 3 indicated that Model 2 had a better fit than other models. Model 2 was a model in which only the value of the HP stimuli in EDM was reflected in the initial value in IDM. This result was consistent with the result of

the chosen frequency in Fig. 1b. We calculated BF for inter-model comparison between Model 2 and other models. The BF was calculated using the marginal likelihood of Model 2 as the numerator. The results showed strong evidence supporting Model 2 ($BF_{21} = 3.302 \times 10^6$, $BF_{23} = 2.913 \times 10^8$, and $BF_{24} = 7.922 \times 10^4$). To evaluate the best-fit model's descriptive ability regarding actual behavioral data, we compared the posterior predictive distribution with the actual behavior data distribution as a posterior predictive check (Gelman et al., 1996). Figure 3 shows that, as the best-fit model, Model 2 can predict the patterns of observed actual behavioral data, i.e., the chosen frequency of the stimulus.

The range of parameters estimated when fitting Model 2 to the actual behavioral data resulted in α ranging from 0.064 to 0.078, β ranging from 1.544 to 14.400, and η ranging from 0.448 to 0.704. These results showed that the estimation range of β and η was similar to that of parameter recovery, but the estimation range of α was smaller than that of parameter recovery. When we conducted the simulation, although the generation range of α was determined by the results estimated from the behavioral data in the previous studies (Zhu et al., 2021), Model 2 included the η, which was not included in Zhu et al.'s (2021) study. Therefore, in the model of Zhu et al.'s (2021) study, the difference in stimulus value was explained by adjusting α, the degree of value learning, while in Model 2, it is explained as a function of the initial value of η in addition to α. Thus, it is likely that the estimated range of the actual behavioral data was smaller than that of the simulation.

In addition, to estimate the extent to which Model 2 explained behavior, we compared it to a model with random choice that did not predict behavior at all (see the supplementary materials for more details). The BF results showed that Model 2 was 4.130×10^{256} times more likely to explain behavior than the random model.

Confirmation of Estimated Initial Values in IDM

The mean initial value of the HP stimuli in the IDM estimated using the CBL model (Model 2) was 0.575 (Fig. 4a).

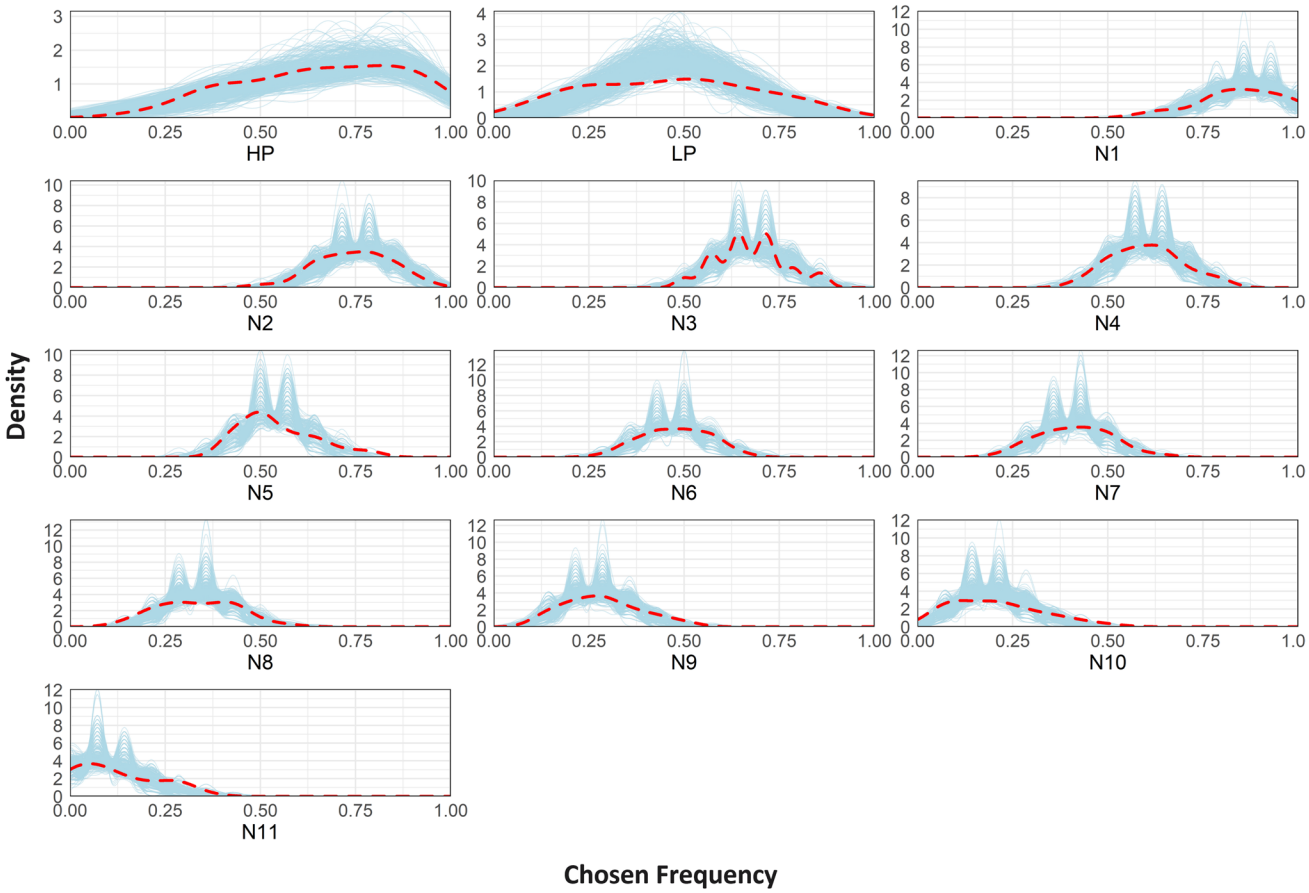

Fig. 3 Comparison of the chosen frequencies' distributions for various stimuli as a posterior predictive check. We examined whether Model 2 can accurately describe the actual behavioral data by comparing the posterior predictive distribution with the actual data distribution. The range of the blue line represents 500 posterior predictive distributions, while the range of the red line represents the actual behavioral data distribution

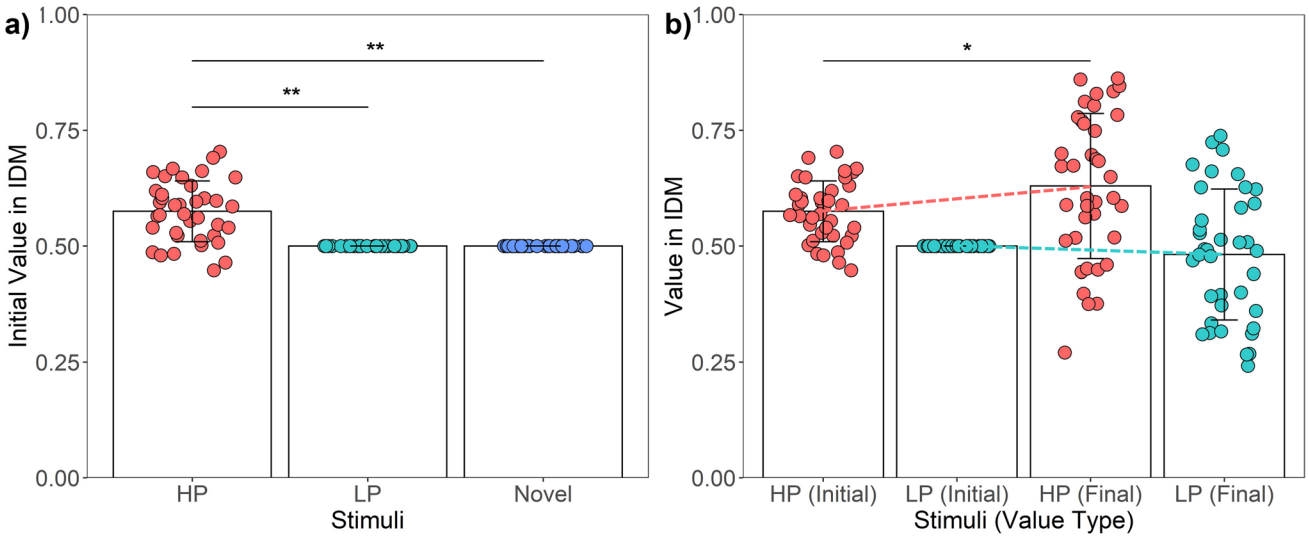

Fig. 4 Values for stimuli estimated by the computational models. **a** Mean initial value of each stimulus type estimated by Model 2 in the IDM task. **b** Mean initial and final values for HP and LP stimuli in the IDM task. Dotted lines represent the value change trend from the initial to the final of each stimulus type. The red and green represent the change trend of HP and LP stimuli, respectively. The error bars and colored dots of all figures indicate *SD* and each participant's data, respectively. $^{*}p < .05$, $^{**}p < .001$

To confirm the values learned in EDM were reflected in the initial values in IDM, we compared the estimated initial value for HP, fixed initial values (0.5) for LP, and novel stimuli. We found that HP stimuli had a higher initial value than LP stimuli and novel stimuli ($t(37) = 7.057$, $p < 0.001$, 95% CI = 0.554, 0.597). These results showed that the high value learned in EDM was reflected in the initial value in IDM.

Estimated Value Change from Initial to Final Values of HP and LP Stimuli in IDM

To examine whether the value learned in EDM affected the degree of value change in IDM, we compared the changes in value between HP and LP stimuli (Fig. 4b). We performed a two-factor repeated measures ANOVA for the stimuli type (HP stimuli and LP stimuli) and the value type (initial value and final value), which found no significant main effect in the value type ($F(1, 37) = 1.307$, $p = 0.260$, partial $\eta^2 = 0.034$) but a significant main effect for the stimuli type ($F(1, 37) = 37.072$, $p < 0.001$, partial $\eta^2 = 0.500$), and the interaction ($F(1, 37) = 10.500$, $p = 0.003$, partial $\eta^2 = 0.221$). We compared the initial value and the final value in each stimulus type and found that the final values were higher than the initial values for HP stimuli ($t(37) = -2.408$, Holm-adjusted $p = 0.021$, $d = -0.582$, 95% CI = $-0.101, -0.009$), whereas no difference was found for LP stimuli ($t(37) = 0.791$, Holm-adjusted $p = 0.434$, $d = 0.142$, 95% CI = $-0.040, 0.070$).

Thus, results showed that the HP stimuli had a higher initial value in IDM due to the influence of EDM and therefore was chosen frequently in IDM, making it increase in value in IDM.

Additional Model Comparison Between the Model 2 and the Models Without Value Updates in IDM

Although we observed that HP stimuli were further valued in IDM based on computational model analysis, Model 2 used in the analysis assumes that the values change. The validity of this assumption was not examined in the comparisons among the four models with different initial value settings (Table 1). Therefore, an additional computational model analysis was performed to investigate whether the value did not change in the IDM. Additional models without value changes were constructed based on Model 2. Specifically, we created four additional models (Table 4): Model A, in which only the value of the HP stimuli was not updated; Model B, in which both the HP and LP stimuli were not updated, and

Table 4 The settings of value updating of each stimulus type in the added models (A–D) without value updates in IDM

Model	Novel	HP	LP
Model A	T	F	T
Model B	T	F	F
Model C	F	T	T
Model D	F	F	F

All added models were constructed based on Model 2. "T" indicates that the value of the stimuli in the CBL model was updated after decision-making (i.e., the value increases when chosen and decreases when rejected), while "F" indicates that the value of the stimuli in the CBL model was not updated

only the value of the novel stimuli was updated; Model C, in which only the value of the novel stimuli was not updated; and Model D, in which the values of all the stimuli were not updated. All new models were analyzed through simulations and the fitting of behavioral data.

The results of the parameter recovery for all added models are shown in Fig. 5. Excluding Model D, there was consistency between the set parameter values (simulated) and the estimated values (fitted) for the remaining three added models ($rs > 0.522$).

Model recovery was confirmed for all models except Model D (see Table 5). When the same true model (used to generate artificial data) was used for the analysis, the fit of the analytical model to the artificial data was the best. When Model 2 and Models A–C were the true models, strong evidence was found for comparison with

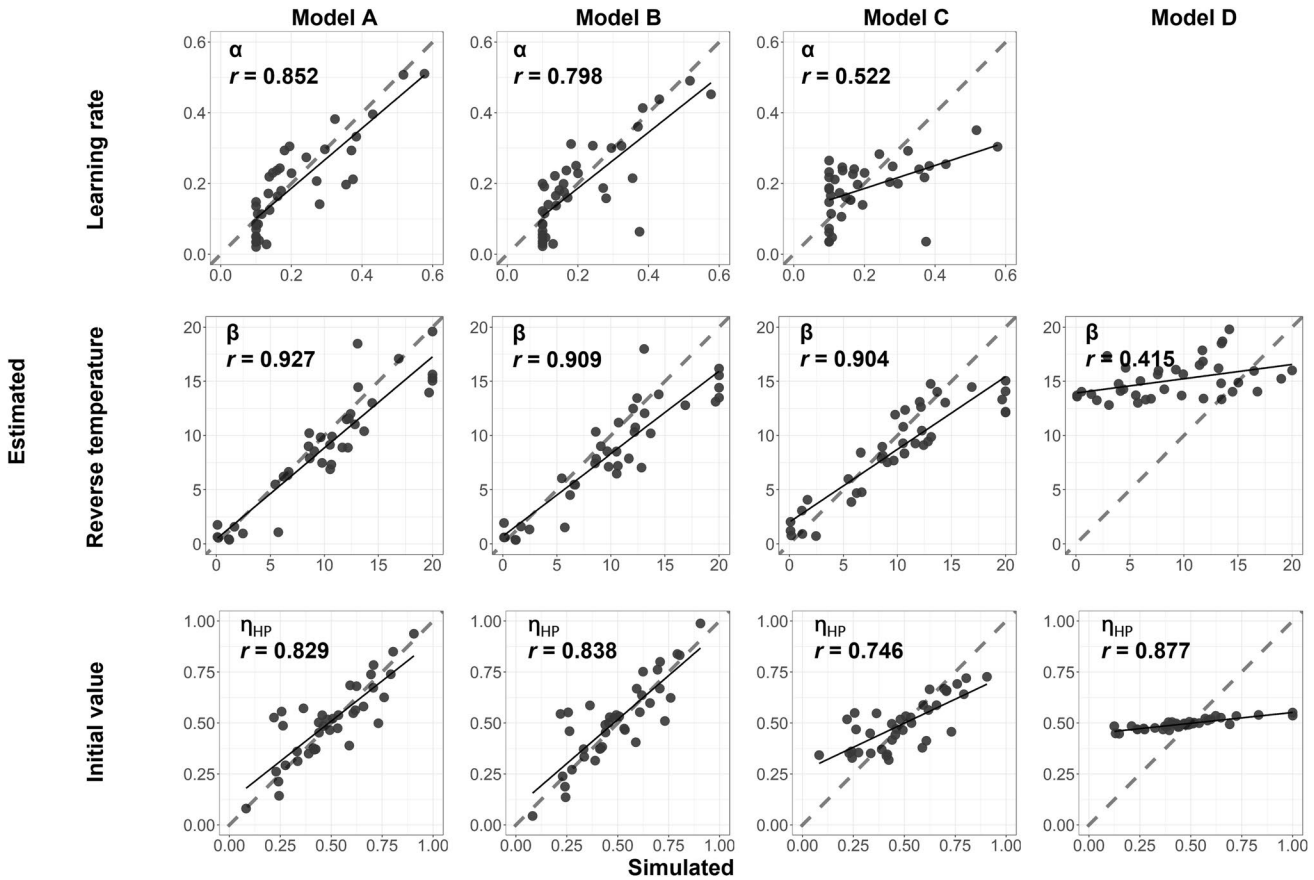

Fig. 5 Results of parameter recovery simulation for additional comparison between the Model 2 and the added models (A–D) without value updates in IDM. This simulation was conducted to confirm whether each newly added model (A–D) could be well estimated as the set value of each parameter. The correlation coefficient between simulated and fitted was shown as parameter recovery indices

Table 5 Results of WBIC for model recovery of additional comparison between Model 2 and the added models (A–D) shown in Table 4

Simulated	Fitted				
	Model 2	Model A	Model B	Model C	Model D
Model 2	**2342.541**	2430.301**	2525.013**	2654.265**	2767.034**
Model A	2427.446**	**2406.561**	2487.599**	2739.019**	2771.429**
Model B	2486.038**	2472.889**	**2445.276**	2771.629**	2771.826**
Model C	2691.428**	2749.732**	2774.443**	**2580.376**	2763.266**
Model D	2772.947^{+}	2772.818^{+}	2771.583^{+}	**2768.434**	**2770.404**

Bold numbers in the table represent the WBIC values of the models that best fit the artificial data. Asterisk (*) is the BF value used for model comparison ($3 <^{+} BF < 20$, $150 <^{**} BF$). The larger the BF value, the greater the difference between the models

other models ($BF_{2A} = 1.299 \times 10^{38}$, $BF_{2B} = 1.764 \times 10^{79}$, $BF_{2C} = 2.399 \times 10^{135}$, $BF_{2D} = 2.264 \times 10^{184}$, $BF_{A2} = 1.176 \times 10^{9}$, $BF_{AB} = 1.564 \times 10^{35}$, $BF_{AC} = 2.425 \times 10^{144}$, $BF_{AD} = 2.885 \times 10^{158}$, $BF_{B2} = 5.043 \times 10^{17}$, $BF_{BA} = 9.821 \times 10^{11}$, $BF_{BC} = 5.411 \times 10^{141}$, $BF_{BD} = 6.590 \times 10^{141}$, $BF_{C2} = 1.695 \times 10^{48}$, $BF_{CA} = 3.551 \times 10^{73}$, $BF_{CB} = 1.915 \times 10^{84}$, $BF_{CD} = 2.680 \times 10^{79}$). When Model D was the true model, although positive evidence was found for comparison with Models 2, A, and B ($BF_{D2} = 12.718$, $BF_{DA} = 11.179$, and $BF_{DB} = 3.251$), no evidence was found for comparison with Model C ($BF_{DC} = 0.139$).

Although Model D did not adequately complete the simulations, we compared all the models by fitting them to the behavioral data. The results indicate that Model 2 had a better fit than the other models. (Table 6; $BF_{2A} = 2.424 \times 10^{47}$, $BF_{2B} = 1.106 \times 10^{107}$, $BF_{2C} = 2.433 \times 10^{147}$, $BF_{2D} = 1.345 \times 10^{232}$).

These results confirmed that the values of all stimuli including HP in the IDM were updated after the choice.

Estimated Final Values in IDM

To examine the effect of the value learned in the EDM on the IDM, we compared the final value of the HP and LP with each novel stimulus in IDM. We used the best-fit model (Model 2) to estimate the values. We ranked all novel stimuli in the order of their final value from high to low within each participant (those labeled as N1 to N11) and then compared them with HP or LP stimuli (Fig. 6a). The results showed that HP stimuli were lower than N1 ($t(37) = -2.883$, Holm-adjusted $p = 0.026$, $d = -0.744$, 95% CI = −0.157, −0.027) and higher than N5 to N11 ($ts(37) > 3.133$, Holm-adjusted $ps < 0.05$, $ds > 0.854$). LP stimuli were lower than N1 to N4 ($ts(37) < -2.768$, Holm-adjusted $ps < 0.05$, $ds < -0.742$) and higher than N8 to N11 ($ts(37) > 2.650$, Holm-adjusted $ps < 0.05$, $ds > 0.749$).

Table 6 WBIC results of IDM behavioral data fit with Model 2 and added models (A–D) shown in Table 4

Model	WBIC
Model 2	**2170.18**
Model A	2279.29**
Model B	2416.66**
Model C	2509.55**
Model D	2704.68**

Bold numbers in the table represent the WBIC values of the model that best fit the behavioral data. Asterisk (*) is the BF value used for model comparison ($150 < {}^{**}BF$). The larger the BF value, the greater the difference between the models

Moreover, we compared the HP and LP with the mean value of all novel stimuli. Results showed that the final value of HP stimuli was higher than that of novel stimuli ($t(37) = 4.708$, Holm-adjusted $p < 0.001$, $d = 1.254$, 95% CI = 0.083, 0.207). In contrast, LP stimuli did not differ from novel stimuli ($t(37) = -0.091$, Holm-adjusted $p = 0.928$, $d = -0.025$, 95% CI = −0.060, 0.055).

These results showed that HP stimuli maintained their high value through the end of IDM, while their value was lower than the most preferred novel stimuli. That is, although the IDM is affected by the EDM value, the superiority of intrinsically learned values (SIV) was concurrently observed in the IDM. Regarding LP stimuli, there was no significant effect on the IDM final value as in the results for initial values, and LP stimuli were the average position of the novel stimuli.

Consistency Between the Final Value Estimated by Model 2 and the Chosen Frequency or Subjective Preference Ratings

To confirm the validity of the model-estimated final values, we examined the consistency between the model-estimated final values and the chosen frequency of each stimulus in the IDM or the subjective ratings of each stimulus after the IDM task. Comparisons were made between novel stimuli ranked by the final value (N1–N11), HP and LP stimuli for each chosen frequency, and subjective preference ratings. Figure 6b shows the consistency between the final value and the chosen frequency. The results demonstrated that the chosen frequency of the HP stimuli was lower than that of the N1 stimulus ($t(37) = -4.267$, Holm-adjusted $p < 0.001$, $d = -1.083$, 95% CI = −1.561, −0.605) but higher than that of the N6–N11 stimuli ($ts(37) > 3.905$, Holm-adjusted $ps < 0.01$, $ds > 1.083$). The LP stimuli were lower than the N1–N4 stimuli ($ts(37) < -2.783$, Holm-adjusted $ps < 0.05$, $ds < -0.747$) but higher than the N9–N11 stimuli ($ts(37) > 3.772$, Holm-adjusted $ps < 0.01$, $ds > 1.025$).

Figure 6c depicts the consistency between the final value and the subjective preference ratings. The results showed that the subjective preference for the N1 stimulus with the highest final value was higher than that for the N4–N11 stimuli ($ts(37) > 4.197$, Holm-adjusted $ps < 0.01$, $ds > 0.779$). The subjective preference of the HP stimuli with the second highest final value was higher than that of the N5–N11 stimuli ($ts(37) > 3.634$, Holm-adjusted $ps < 0.05$, $ds > 0.880$), while the subjective preference of the LP stimulus was lower than that of the N1–N2 ($ts(37) < -4.026$, Holm-adjusted

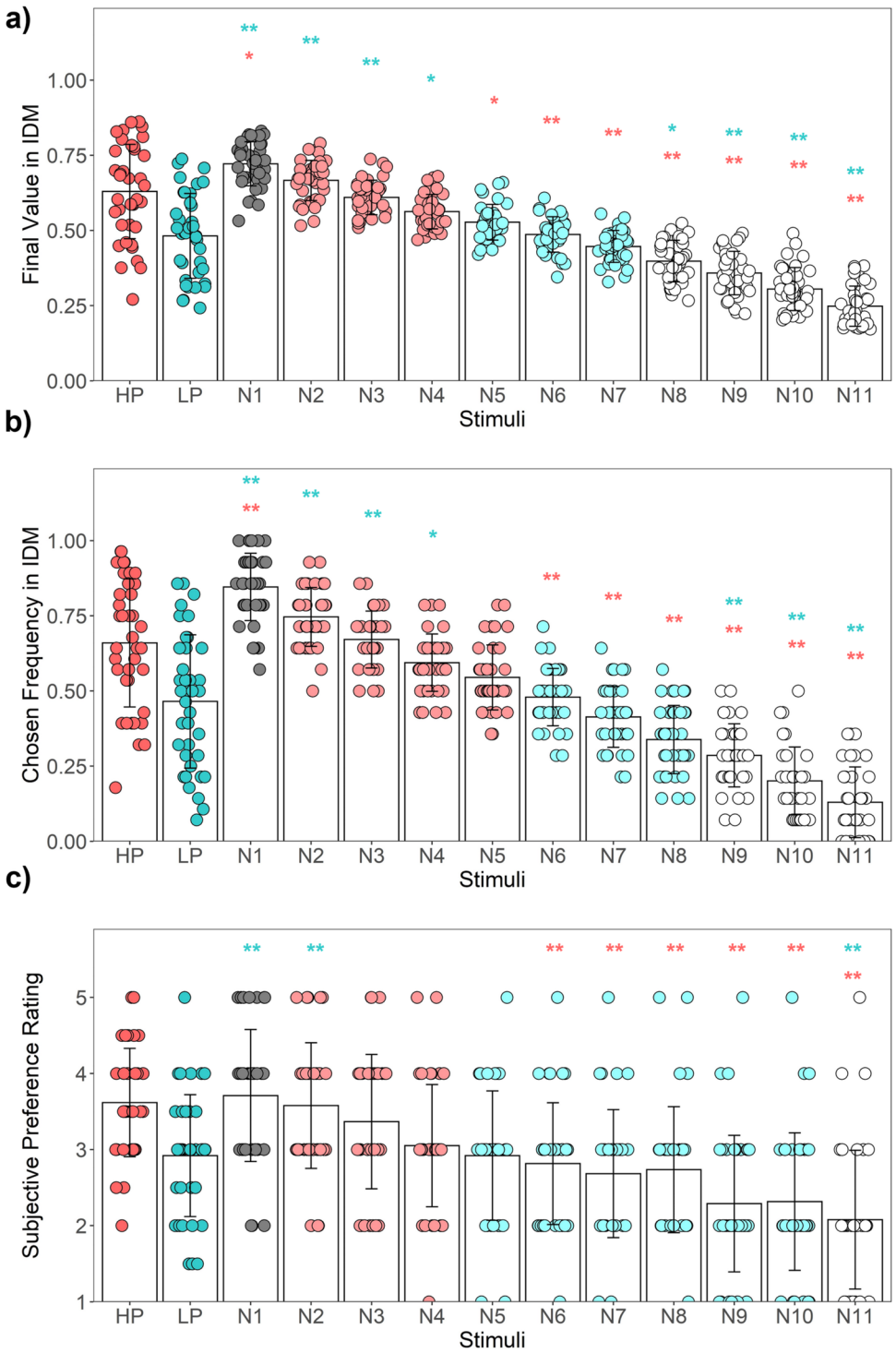

Fig. 6 Comparison of the final value estimated by the Model 2 among stimulus type (**a**) and the correspondence between the estimated final value and chosen frequency (**b**) or the subjective preference ratings (**c**). **a** Comparison of the mean final value of HP and LP stimuli in IDM task with all novel stimuli. N1–N11 denote the rank of the novel stimuli in the final value. N1 was the most favorite novel stimulus, and N11 was the least favorite novel stimulus for each participant. The red asterisk (*) represents the comparison result with HP, and the green asterisk (*) represents the comparison result with LP. There was no difference between N2–N4 and HP and between N5–N7 and LP. The error bars and colored dots indicate *SD* and each participant's data, respectively. $*p < .05$, $**p < .001$. **b** Comparison of the mean chosen frequency in IDM for all types of stimuli. N1–N11 denote the rank of the novel stimuli in the final value. The error bars and colored dots indicate SD and each participant's data, respectively. $*p < .05$, $**p < .001$. **c** Comparison of the subjective preferences ratings rated on a 5-point Likert scale (1 = Extremely Dislike, 5 = Extremely Like). N1–N11 denote the rank of the novel stimuli in the final value. The error bars and colored dots indicate *SD* and each participant's data, respectively. $*p < .05$, $**p < .001$

$ps < 0.01$, $ds < -1.263$) but higher than that of the N11 ($t(37) = 4.295$, Holm-adjusted $p < 0.01$, $d = 0.971$, 95% CI = 0.499, 1.443).

Overall, the results validated the final value estimated by Model 2, indicating that stimuli with high values in the IDM had a higher chosen frequency and were subjectively preferred.

Discussion

The goal of this study was to determine whether, how, and to what extent the EDM value affects the IDM.

Whether the Value Learned in EDM Affects IDM

To examine whether the EDM value affects the IDM, we compared the chosen frequency of the three types of stimuli (novel, HP, LP) in the IDM task. The chosen frequency for HP stimuli was higher than LP stimuli and novel stimuli (Fig. 1b). These results indicated that the high values learned in EDM affect the IDM.

It could be argued that our finding that stimuli that were highly valued in the EDM were also chosen in the IDM reflects previous results from the extinction procedure. In conditioning studies, subjects continued to choose the reinforced option even after the reward was removed (Bouton & Moody, 2004; Dickinson & Balleine, 1995; Rescorla & Wagner, 1972; Stevenson & Clayton, 1970; Thorndike, 1898). However, in such cases, unlike the present study, the subjects were not explicitly told that the decision task or situation had changed. Therefore, they were placed in a position in which they expected that choosing the reinforced option would eventually reward them (Dickinson & Balleine, 1995; Thorndike, 1898). In contrast, participants in the present study were clearly instructed on the difference between the EDM and the IDM, and they were aware that they would not be rewarded in the IDM. That is, the present study operationally eliminated the participants' choice of reinforced items based on the expectation of an externally derived reward and asked them to choose a preferred shape according to their own preference criteria in the IDM. Furthermore, in the extinction procedure, the value of the stimuli or behavior decreases when external rewards are not provided after choice, and the option is gradually not chosen (Bouton & Moody, 2004; Rescorla & Wagner, 1972; Stevenson & Clayton, 1970). In contrast, in the present study, HP stimuli were more frequently chosen in the IDM, even though no reward feedback was presented, and their value further increased (Fig. 4b). This difference suggests that, in the extinction procedure, the decision was based on the expectation of rewards from the external environment, whereas in the IDM in the present study, CIPC occurred because the HP stimuli were chosen based on their own preference. Therefore, the results of the present study cannot be explained by the sustained choice of highly rewarding stimuli reported in conditioning studies (Bouton & Moody, 2004; Rescorla & Wagner, 1972; Stevenson & Clayton, 1970).

How the Value Learned in EDM Affects IDM

To examine how the EDM values reflect the IDM, we conducted a computational model analysis. The model comparison revealed that Model 2, in which the initial values of the HP stimuli in the IDM were different from that of the other stimuli, best fit the data (Table 3). By comparing the initial values of each stimulus estimated by this model, we confirmed that the EDM values of the HP stimuli were reflected in the initial IDM values (Fig. 4a). This result was consistent with the results of the chosen frequency (Fig. 1b). Collectively, the high values learned in EDM (reward learning task) were reflected in the initial values of IDM (preference judgment). This suggests a close relationship between the values obtained using the EDM and IDM.

To further examine how the value learned in EDM affects IDM, we examined whether the value learned in EDM affects the degree of value change in IDM. The high value of HP at the end of IDM was not simply a result of the maintenance of the high initial value reflected as the effect of EDM, but it was also shown that the value of HP was further increased by selection in IDM (Fig. 4b). Although we also examined the possibility that the value of HP stimuli was not updated in the IDM (Table 4), such a model did not fit well with the behavioral data (Table 6). Those results demonstrate that what is learned to be of high value according to externally derived criteria will subsequently be further valued within that individual through their own choices and the following CIPC. This result may indicate part of the internalization process of the value of the external environment if the values in the external and internal criteria differ.

It remains unclear how EDM values were reflected as initial values in the IDM. The effect of the EDM on the IDM was inferred from the notion that activity in reward-related brain regions has been reported in both the EDM (Bechara et al., 1997, 2005; Marco-Pallarés et al., 2008, 2015; Yacubian et al., 2006) and IDM(Akaishi et al., 2014; Lee & Daunizeau, 2020; Nakao et al., 2016, 2019). It is possible that the shared neural basis of value representation between the EDM and IDM underlies the results of this study; however, further research with brain activity measurements is needed to verify this point.

Another possible explanation at the cognitive level is that the values learned in the EDM are more likely to be used as cues for preference judgments regarding novel contour shapes with less clear preferences. Because the IDM in this experiment was a preference decision, the observed CIPC as an increased HP value (Fig. 4) can be interpreted as the result of choosing the preferred HP in the IDM. However, CIPC does not necessarily have to be preference-based at the decision stage but can also occur when it is interpreted as a preference-based decision after the decision (Johansson

et al., 2014). Therefore, it is also possible that at least at the stage when the HP stimuli were first presented in the IDM, the choice was not solely based on preference, but the choice was made based on the value in the EDM. This is a possible process of influence of the EDM on the IDM that we aimed to examine in this study, but it means that the value in the EDM may not have been internalized as a value in the IDM from the beginning. However, even if participants did not choose HP stimuli based solely on preference in the early stages of IDM, it is likely that they interpreted their decision as preference-based after the choice because the value of HP subsequently increased (Fig. 4). Given that such preference judgments based on the values learned in the EDM are likely to occur when stimulus preferences are not formed at the onset of the IDM, it is possible that the effect of the EDM on the IDM was more easily observed in this study, which used novel contour shapes. In the future, it would be desirable to investigate whether stimuli with clearer preferences can be used to influence the EDM on IDM.

Is it possible to interpret the results of this study based on familiarity? It is thought that the stimuli presented in the EDM (i.e., HP and LP) are processed as more familiar stimuli in the IDM than as novel stimuli. Although familiarity effects cannot be completely ruled out, they alone cannot explain the results of this study. If familiarity could explain the behavioral data in the IDM, we would expect to observe the same differences in the chosen frequency between the LP and novel stimuli as between the HP and novel stimuli. Additionally, if familiarity is an important factor, we would expect a model like Model 4 (where HP and LP are different from novel stimuli) to fit the behavioral data better. However, Model 2 was adopted, in which the initial value of only the HP stimuli was different from the other stimuli. This suggests that it is more plausible to assume the influence of value rather than that of familiarity. Conversely, a caveat is that the influence of familiarity cannot be completely ruled out, and the value of the LP was possibly higher because of the influence of familiarity than it would have been without it. To examine the influence of familiarity, it would be necessary to compare the LP with a stimulus without feedback that was presented for the same duration as the LP before the IDM.

What Extent the EDM Value Affects the IDM

To determine the extent to which the values learned in the EDM affected the IDM, we compared the model-estimated final values for all stimuli in the IDM. Interestingly, we found the superiority of intrinsically learned value (SIV) in the IDM, in which the most preferred novel stimulus learned in IDM (N1 in Fig. 6a) was preferred over HP stimuli. If the values in the EDM and IDM were the same, then it would be expected that the stimuli with a higher value in the EDM would also have the highest value in HP in the final value in the IDM. This is because it was learned as highly valued in the EDM and subsequently chosen (Fig. 1b) and valued in the IDM (Fig. 4b). However, the SIV of novel stimuli is observed in the IDM (Fig. 6a), indicating that our preferences are strongly influenced not only by externally given rewards but also by increased preferences on our own choices. Although the EDM value affects the IDM value, it is unlikely that the EDM and IDM values are identical.

However, the mechanisms underlying SIV in the IDM remain unclear. The recently proposed fundamental self-hypothesis (Humphreys & Sui, 2016; Northoff, 2016; Northoff et al., 2022; Qin et al., 2020; Sui & Gu, 2017; Sui & Humphreys, 2015; Zhang et al., 2022) is considered relevant. This postulates that the self is a fundamental brain function that precedes and controls cognitive functions, such as perception, emotion, and reward, which has been proposed in studies of spontaneous brain activity (Northoff, 2016; Northoff et al., 2022; Qin et al., 2020) and the self-prioritization effect (SPE) (Humphreys & Sui, 2016; Sui & Gu, 2017; Sui & Humphreys, 2015; Zhang et al., 2022). In this hypothesis, the self is embedded in spontaneous brain activity, and when a stimulus appears, the default mode network (DMN), which is responsible for processing self-associated stimuli, interacts with a task-related network to influence cognitive processing. A meta-analysis of the neural basis of IDM and EDM also confirmed that IDM differs from EDM in that the DMN is its primary neural substrate (Nakao et al., 2012). In addition to the conceptual and operational differences between the EDM and IDM, there is a difference in task demands (i.e., whether decisions are made based on value criteria given by the environment or based on one's own value criteria), that is, an essential difference in self-involvement. The continuous choice of stimuli as one's own favorite shape, rather than because it has previously been rewarded, is likely to increase the self-relatedness of the item. As self-related stimuli are known to induce reward-related brain activity (de Greck et al., 2008; Enzi et al., 2009), an increase in self-relevance may trigger an internal reward response (Aridan et al., 2019; Camille et al., 2011; Fellows & Farah, 2007; Izuma et al., 2010; Miyagi et al., 2017; Nakao et al., 2016), leading to an increase in value. As a result, the most preferred novel stimulus learned in the IDM might have a higher value than HP stimuli in the IDM.

Although the model-free measure of chosen frequency also confirmed this SIV (Fig. 6b), which was not reflected in the subjective preference ratings after IDM, subjective preference showed no significant difference between HP and N1 (Fig. 6c). In the IDM task, 15 stimuli were presented in different combinations in each of the 105 trials. It is possible that this complex task setting prevented the subjective recognition of which stimuli were most favorable or chosen. It is necessary to reexamine whether this SIV can be observed in

subjective preference ratings by experimenting with simpler task settings.

The low values learned in EDM did not affect IDM. In the EDM task, when an incorrect answer was chosen, feedback was displayed as 0 (simply not presented with a reward) rather than presented with a punishment. There was a possibility that the non-reward of LP stimuli in EDM did not affect participants' preference for those stimuli, and therefore the initial value of LP stimuli in IDM was the same as for novel stimuli. People tend to choose options that can earn rewards and avoid punishment (Guitart-Masip et al., 2014). Depending on whether the value is learned through reward acquisition or punishment avoidance in EDM, it is possible that IDM will reflect either high or low value in EDM. It is likely that a lower value learned based on punishment feedback in the EDM can be perceived as a stimulus with a lower value than a novel stimulus in the IDM. Additionally, different levels of reward and punishment in the EDM task may have different effects on the IDM. High-rewarding HP stimuli will learn higher values in the EDM task, reflect higher initial values in the IDM task, and be chosen more frequently, whereas LP stimuli with higher penalties will learn lower values in the EDM task, reflect lower initial values in the IDM task, and be rejected more frequently. Future studies should investigate these possibilities.

Limitations and Further Directions

This study showed for the first time the value of EDM affects IDM and the SIV in IDM. Nevertheless, the present studies have several main limitations. First, the study showed that high value in EDM was reflected in IDM and low value was not. However, we should note that the results did not lead to any general conclusions about the relationship between EDM and IDM values but were a conclusion that depended on the task settings of this study. In this study, EDM used relatively easy reward probability settings such as 90% vs. 10% and 80% vs. 20%, where learning of value was easily established, to examine whether the value of EDM was reflected in IDM. Therefore, there were few opportunities to receive incorrect feedback after choosing LP stimuli and decreasing their values: the mean proportion of trials in which incorrect feedback was given after choosing the LP stimulus was 0.103, with *SD* of 0.113 (for comparison, the mean proportion of trials in which correct feedback was given after choosing HP stimuli was 0.734, with *SD* of 0.150). As a result, participants likely learned that LP stimuli were of relatively low value but did not come to the realization that they had to actively avoid LP stimuli. Therefore, it is assumed that LP stimuli in IDM were treated as having the same initial value as novel stimuli. When using an EDM task where the participant actively decides not to choose an item to avoid losses, the low value in EDM may affect IDM. Therefore, there is room for further study on this point.

Second, this study did not use a fixed interval between the EDM and IDM tasks but instead allowed participants to choose when to begin the IDM task after the EDM task. We predicted that the degree to which the value learned in the EDM affects the IDM will decrease with increasing time intervals. Values can decrease with prolonged periods of unexposure (e.g., Ito & Doya, 2009; Katahira et al., 2017). A longer time interval will result in less value learned in EDM being reflected in the initial value of the IDM, whereas a shorter time interval will allow participants without clear preferences to receive more value learned in EDM. In this study, although participants had shorter task intervals, with the longest being 27.08 s, there was no significant correlation between the time interval and the initial value of the HP stimuli ($r = 0.009$, $p = 0.956$). However, we cannot rule out the impact of time intervals on the value relationship between the EDM and IDM. These findings need to be validated in future studies.

Third, the possibility of explanations using other types of models has not yet been explored. For example, the cognitive dissonance theory has been used to explain the phenomenon of CIPC in IDM (Festinger, 1957). In this theory, CIPC is explained by which choosing one item from two items with the same subjective preference rating, causing dissonant feelings (i.e., cognitive dissonance). Subsequently, they adjust their preferences for the chosen and rejected items in order to reinforce that their choice is reasonable. A difference between the cognitive dissonance theory and the CBL model is that the former assumes that preferences change in situations where one of the pairs of equal liking is chosen. The latter assumes that preferences change for all stimulus pairs independently of the equality of the preferences of the two options. While a computational model that represents cognitive dissonance has also been constructed (Vinckier et al., 2019), the model was not used for the IDM task as in the present study. It remains possible that if we build a CBL model incorporating cognitive dissonance and the model fits the behavior well, we may get different results from the present results. Furthermore, it goes beyond the integration of cognitive dissonance and the CBL model. There may be room in the future to consider the integration of RL and CBL models, which have similar formulas. This makes it possible to model complex decisions in which the EDM and IDM processes interact.

Finally, there was a possibility that preferences were formed to some extent by the first impression in IDM. Although we used novel contour shapes by following the previous study (Zhu et al., 2021) to minimize the impact of initial preferential differences, we cannot rule out the possibility that value can be formed by first impression. For individuals whose preferences were formed by first

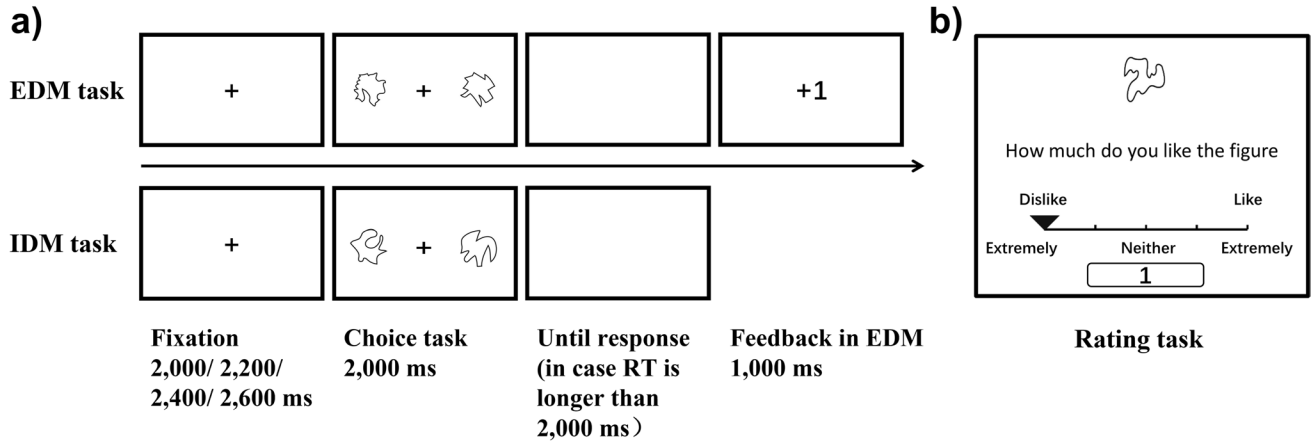

Fig. 7 Experimental procedure. **a** In the EDM task, participants were asked to choose the one considered correct from the two stimuli. In the IDM task, participants were asked to choose the one preferred from the two stimuli. Feedback was not presented in the IDM task. **b** In the rating task, participants subjectively evaluated all stimuli in the IDM task on a 5-point Likert scale (1 = Extremely Dislike, 5 = Extremely Like). The rating task was conducted after the IDM task. The subjective rating data was not used for computational model analyses

impressions, it is possible that the estimated learning rate was estimated to be larger than the true value.

Conclusion

In this study, we implemented the tasks of EDM and IDM using similar experimental procedures (Fig. 7) and applied the computational model analysis for the behavioral data of both decisions. In which EDM was followed by IDM and presented the same stimuli as EDM, we showed that the learned high values in the EDM reflect on the initial preference of the IDM. Stimuli that had been learned to have high value through EDM were also chosen in IDM, further increasing their value through IDM. Moreover, from the results of the final value in IDM, stimuli that were of high value in EDM were still of high value at the end of IDM, but not as high as the novel stimuli in the most preferred IDM. These results demonstrated that externally given criteria have a strong influence on our later preferences and, at the same time, demonstrated that values formed by choice based on one's own criteria can be higher than externally derived values. We propose that the phenomenon of SIV in the IDM may be observed through the reward response to processing self-relevant stimuli in terms of the fundamental self-hypothesis. This study is the first to disentangle the relationship between EDM and IDM, revealing that EDM values influence IDM and determine SIV. This superiority suggests that the values learned through the EDM and IDM are likely to differ. Our findings serve as a window to the comprehensive understanding of the decision-making process.

Methods

Participants

Thirty-eight healthy Japanese university students (male = 17, female = 21, mean age = 20.789, age range = 18–29) participated in the experiment. All participants were native Japanese speakers, right-handed, with regular or corrected-to-normal vision. All experimental protocols were accepted by the Ethics Committee of the Graduate School of Education at the University of Hiroshima. According to the guidelines of the Research Ethics Committee of the University of Hiroshima, all participants provided informed written consent prior to participation. They were compensated for participating in the experiment.

Stimuli

In this study, novel contour shapes were used to avoid the influence of preferences acquired prior to the experiment. Fifteen novel contour shapes were selected from a previous study (Endo et al., 2003). These shapes have been used in EDM and IDM studies and applied to computational model analyses in which their initial values are assumed to be equal (Kunisato et al., 2012; Ohira et al., 2009, 2010; Zhu et al., 2021). Furthermore, to minimize the differences in initial preferences caused by shape differences, these shape differences were made as small as possible. Specifically, we selected shapes with mild complexity (mean = 5.1, $SD = 1.4$), width (mean = 5.5, $SD = 1.8$), smoothness (mean = 4.6, $SD = 1.7$), symmetry (mean = 3.7, $SD = 1.7$), and orientation (mean = 5.0, $SD = 2.1$). The ratings of these characteristics were collected using a 9-point rating scale (1–9) (Endo et al.,

2003). Additionally, these shapes had lower association values (mean = 65.71, *SD* = 6.84). The association value was the percentage (%) of respondents who gave the name of a specific object when asked to name the object recalled by the shape and those who could not write the name but said it resembled something (Endo et al., 2003). These shapes are numbered 29, 31, 35, 36, 37, 39, 42, 44, 45, 56, 63, 65, 81, 87, and 92 in the original study (Endo et al., 2003). The image used in the experiment was 800 × 600 pixels in size. Within 30° of the angle of view, participants could see a picture on the screen.

In both tasks, PsychoPy (Peirce et al., 2019) was used to present each pair on a white background, with one member on the left and the other on the right side of the screen. The order of the trials, as well as the presentation slides of the shapes, were randomized through the participants. The experiment was carried out on a Windows 10 PC with a 1920 × 1080 monitor.

Task

All participants conducted EDM tasks followed by IDM (Fig. 7a). Subjective preference ratings of each stimulus were conducted after the IDM.

EDM Task (Reward Learning Task)

Four shapes were randomly selected from 15 stimuli materials, and pairwise combinations were presented. Participants carried out six blocks of 34 trials. Each trial started with a fixing cross shown. To avoid the influence of pre-stimulus brain activity at specific frequencies on decision-making performance (Bai et al., 2016; Fellinger et al., 2011), the duration of the fixation cross was randomly varied time of 2000 ms, 2200 ms, 2400 ms, or 2600 ms. Subsequently, two shapes of the fixed combinations were presented for 2000 ms on the left and right sides of the fixation cross. Participants were asked to choose one of the two shapes considered to be correct as quickly and correctly as possible by pressing the "f" key (left) or the "j" key (right) on a standard computer keyboard. To limit the exposure period for each stimulus, the stimuli disappeared (i.e., turned to a white screen) after 2000 ms. Participants could make their response even after the two shapes had disappeared, and the white screen disappeared once participants made a response. If a key was pressed within 2000 ms, a white screen was not shown. After the response, they received correct (+ 1) or incorrect (0) feedback for 1000 ms. Participants were informed that + 1 indicated an increase in the number of points earned, meaning that the more points earned, the higher the reward paid to the participant after the experiment. The participants were instructed to earn as many points as possible. In each trial, each stimulus had a certain probability of receiving points after selection, and 90% vs. 10% stimuli pair or 80% vs. 20% stimuli pair appeared randomly. In addition, the left and right positions of the stimuli in each stimuli pair were also random. The participants were informed that each stimulus had a certain probability of obtaining points, but they were not informed of the specific probability. Furthermore, it was worth noting that the reward for each stimulus was generated independently in each trial. After a participant chose one stimulus of a pair and received a reward, it was difficult to infer whether another rejected stimulus could receive one. Although participants may have speculated on the possibility of the outcome of the other stimuli, the behavioral data do not support this possibility (see the supplementary materials for more details).

IDM Task (Preference Judgment Task)

This task was the same as that in the previous study (Zhu et al., 2021). Including the four shapes in the EDM task, all 15 shapes randomly created 105 pairs (i.e., 14 presentations per stimulus) in the IDM task, and each pair of stimuli was presented only once. There were three types of stimuli in the IDM task: novel stimuli and the stimuli presented in EDM consisted of high-probability reward stimuli (90%, 80%; HP) and low probability reward stimuli (20%, 10%; LP). Participants carried out five blocks of 21 preference decision trials and were asked to choose the preferred shape among two shape stimuli presented according to their own preferential criteria in each trial. We also informed participants that there was no objectively correct answer in this task. Stimuli were presented in the same manner as the EDM task, except there was no feedback after the choice.

Rating Task

We conducted a subjective rating task for each shape stimuli (Fig. 7b) to examine the relationship between subjective preferences and final stimulus values estimated by computational model analysis for IDM behavioral data. Following the IDM task, participants carried out a subjective preference rating task. In the rating task, participants were asked to determine their subjective preference, graded on a 5-point Likert scale (1 = Extremely Dislike, 5 = Extremely Like) for each shape. It is worth noting that for the IDM task, we did not use the experimental paradigm with subjective ratings, such as the free choice paradigm (Brehm, 1956) or rate-rate-choice (Chen & Risen, 2010), to avoid the influence of noise-contaminated subjective ratings on CIPC measurement (Izuma & Murayama, 2013). As a result, preference rating data in the Likert scale was not included in the computational model analyses.

Classical Analysis of the Behavioral Data

We first confirmed the correct response rate for each stimulus pair in the EDM task to gauge whether participants learned through the EDM task.

To investigate whether the value learned in EDM had an effect on IDM, prior to computational model analyses, we compared the chosen frequency of the three types of stimuli (HP, LP, and novel) in the IDM task using the paired *t*-test with the Holm multiple-comparison correction. The chosen frequency of each type of stimuli was calculated by dividing the number of times it was chosen across all trials by the number of times it was presented.

Computational Models

To investigate how the values learned in EDM reflected IDM, we prepared four CBL models with different initial stimulus values (Table 1), which were established by the previous study (Zhu et al., 2021). The CBL model's learning process involves increasing the value of chosen items while decreasing the value of rejected items. The CBL model is written as follows:

$$V_i^{IDM}(t+1) = \begin{cases} V_i^{IDM}(t) + \alpha_c\left(1 - V_i^{IDM}(t)\right) & \text{if } i \text{ was chosen} \\ V_i^{IDM}(t) + \alpha_r\left(0 - V_i^{IDM}(t)\right) & \text{if } i \text{ was rejected} \end{cases} \quad (1)$$

The values (V^{IDM}) in CBL models were updated based on whether a participant chose or rejected it. The updated V^{IDM} was kept constant until the trial in which the stimulus was presented. The degree of value change followed by choice was determined by the learning rate (α_c or α_r). When item i was chosen, the learning rate (α_c) was multiplied by $1 - V_i^{IDM}(t)$ and added to the value at trial t, as if it were the prediction error for a correct response (feedback is + 1) in the RL model (see the supplementary materials Eq. s1). In case item i was rejected, the learning rate (α_r) was multiplied by $0 - V_i^{IDM}(t)$ and added to the value at trial t as if it were the prediction error in the RL model for an incorrect response (feedback is 0) in the RL model.

The typical RL model in Equation s1 (see the supplementary materials) does not update the values of the rejected items. In contrast, in CBL, participants updated the value of items based on their own choices; hence, rejected items were considered incorrect answers, and their values were updated to decrease. Notably, although the EDM and IDM are similar in updating values through differences in existing values and feedback, there are differences in updating the values of both or chosen options. These differences arise from task design. In a typical EDM task, including our EDM task, the feedback for the two items is independently determined by probability and participants are only informed of the feedback of the chosen option. Therefore, knowing whether the rejected item was the correct answer was difficult, and only the value of the chosen option would increase or decrease based on feedback from external circumstances (Palminteri et al., 2017). In contrast, in the IDM task, it is clear that what they choose is preferred, and what they do not choose is not preferred. Thus, the value of the chosen option increases, and the value of the rejected option decreases (Brehm, 1956). Zhu et al. (2021) compared CBL models that update the value of both chosen and rejected models that update only one of them and reported that models that update both values have a better fit to the behavior. We used the same IDM task as Zhu et al. (2021). Although not directly related to the aim of this study, we confirmed that the behavioral data from this study are better suited for an RL model that only updates chosen options rather than one that updates both chosen and rejected options (see the supplementary materials).

The initial values of the stimulus types (novel, HP, or LP), wherein the models differed, were estimated as free parameters. Model 1 represented no influence of the values learned in EDM, and the initial value η ($0 \leqq \eta \leqq 1$) was a free parameter, which was the same for all stimulus types in the IDM task. Model 2 represented that only the initial value of the HP stimuli can reflect the EDM value and differed from the other stimuli in the IDM task. The initial value of HP stimuli (η_{HP}) was a free parameter, whereas the initial value of the LP stimuli was fixed at 0.5, the same as with novel stimuli. In contrast to Model 2, in Model 3, only LP was a free parameter, and the others were fixed at 0.5. In Model 4, both HP and LP were free parameters, and the novel stimuli were fixed at 0.5. Since the previous study (Zhu et al., 2021) reported that the model that used different learning rates for chosen and rejected items was unsuitable for model comparison, the above four models were created based on the model that used the same learning rates (i.e., $\alpha_c = \alpha_r$) as the main model-based analysis. Model 1 is a null model that assumes no effect of the EDM on the IDM. If the fit of the other models is superior to that of Model 1, this indicates an effect of the EDM on the IDM. Because behavioral data are more likely to fit a model with a higher number of parameters (Watanabe, 2013) and to avoid the issue of parameter setting for the initial values of other models affecting the fit of Model 1, we set the initial values of Model 1 as free parameters under the constraint of no influence from the EDM. Thus, if other models provided a better fit, we could clearly determine the effect of the EDM on the IDM.

To calculate the probability of choice in the CBL models, the softmax function was applied to the value difference between the two options.

$$P_{chosen} = \frac{1}{1 + \exp(-\beta(V_{chosen}^{IDM}(t) - V_{rejected}^{IDM}(t))} \quad (2)$$

In trial t, $V^{IDM}_{chosen} - V^{IDM}_{rejected}$ with parameter β was used to determine the value of P_{chosen}, which was used to represent the probability that the model chooses the option the participant chose. β determined the softmax function's slope. The higher the value, the more the decision was based on the values, while the lower the value, the more random the decision was and the less reliant on the value.

Simulations and Model-Based Behavioral Data Analyses

All CBL models (Table 1) underwent both simulations of parameter and model recoveries. CBL models that passed the parameter recovery performed model recovery, testing whether the model that generated the artificial data best fit the same model. Finally, we fitted the actual behavioral data to the CBL models to determine which model best explained the behavioral data.

All subsequent simulations and actual data analyses were performed on R (R Core Team, 2020). The hierarchical Bayesian method was used to derive model parameters, and the calculation process was completed by rstan package (Stan Development Team, 2020). This method assumes that a common distribution within the group generates each participant's parameters (e.g., α_n, β_n). As shown in the following Eqs. 3 and 4, the parameters α and β of participant n were assumed to be generated from the normal distributions of μ and σ^2. As with the prior distributions of μ and σ^2, we used the uniform distribution. At the same time, we truncated the normal distribution to ensure that the generated parameters were within a certain range.

$$\alpha_n \sim N\left(\mu_\alpha, \sigma_\alpha^2\right) \tag{3}$$

$$\beta_n \sim N\left(\mu_\beta, \sigma_\beta^2\right) \tag{4}$$

The parameters at the population level (μ_α, σ_α^2, μ_β, σ_β^2) were used as free parameters to infer from data. At the population level, the distribution parameters took a prior distribution into account, and the distribution calculated a posterior distribution using the Bayes estimator. The posterior distribution of parameters was obtained by the Markov Chain Monte Carlo method (MCMC).

Simulation 1 (Parameter Recovery)

We conducted parameter recovery simulations for the CBL models to evaluate whether the experimental settings and models met the goal of estimating model parameters from behavioral data (Wilson & Collins, 2019). We tested whether model parameters used to produce artificial behavioral data could be estimated by model fitting to the artificial data. As parameter recovery indices, Pearson's correlation coefficient was calculated between the simulated and fitted parameters.

We used the same settings as the actual experimental design when generating the artificial behavioral dataset. In each model, we generated artificial data with 15 stimuli and 105 trials for 38 people. Initial values for stimuli were set as in Table 1. The α, β, and η were generated from the normal distribution of μ and σ^2. In the stage of generating artificial data, to make the simulation more consistent with actual behavioral data, μ_α, σ_α^2, μ_β, and σ_β^2 were set based on the analysis results of Zhu et al.'s (2021) research on actual behavioral data. Specifically, the α and β of all models were generated from the normal distribution of $\mu_\alpha = 0.160$, $\sigma_\alpha^2 = 0.180$, and the normal distribution of $\mu_\beta = 9.870$, $\sigma_\beta^2 = 6.480$, respectively. Unlike α and β, η, the parameter for estimating the initial value of the HP and/or LP stimuli (Table 1), was not included in Zhu et al.'s (2021) research and was a new parameter added in this study. Therefore, the η was generated to fall within the range of the stimulus values of 0–1. That is,η of all models were generated from the normal distribution of $\mu_\eta = 0.500$, $\sigma_\eta^2 = 0.300$.

At the stage of model fitting to the data, μ_α, σ_α^2, μ_η, and σ_η^2 were generated from the uniform distribution ranging from 0 to 1, while μ_β and σ_β^2 were generated from the uniform distribution ranging from 0 to 20 and 0 to 10, respectively.

Simulation 2 (Model Recovery)

Model recovery was conducted to test whether the true model showed the best fit for the data generated by that model under the experimental design. The widely applicable Bayesian information criterion (WBIC) (Watanabe, 2013) was used to assess the relative goodness of fit of the models. As shown in Eq. 5, the − WBIC is equal to the approximate value of log marginal likelihood.

$$\mathrm{WBIC} \approx -\log p(X|M_i) \tag{5}$$

$p(X|M_i)$ is the marginal likelihood, which is the probability of generating data X given by the model M_i.

More specifically, we first used the MCMC method after performing the transformation in Eq. 6 to estimate the log posterior density $\log p(X|\theta)^{\frac{1}{\log N}} p(\theta)$.

$$\log p(X|\theta)^{\frac{1}{\log N}} p(\theta) = \frac{1}{\log N}\log p(X|\theta) + \log p(\theta) \tag{6}$$

θ was the vector of parameters such as α, β. N represented the sum of all trials across participants.

The WBIC was then calculated from the S samples of MCMC as Eq. 7.

$$\mathrm{WBIC} = -\frac{1}{S}\sum_{s=1}^{S}\log p(X|\theta^s) \quad (7)$$

S for model recovery was 5000 samples, while those for parameter recovery and behavioral data analysis were 10,000.

A smaller WBIC indicated a better fit of the model to the data. The four individual models shown in Table 1 generated artificial behavioral data in the same way with Simulation 1. Subsequently, each data set was fitted to all models and judged which model best fit the data using WBIC.

To compare which model had the higher probability of generating data, we calculated the Bayes factor (BF). The BF was calculated by the ratio of the marginal likelihood of the two models and was calculated using the marginal likelihood of the model used to generate the data as the numerator. A previous study (Kass & Raftery, 1995) referred to evaluated the Bayes factor of 1–3 as not worth mentioning, 3–20 as positive, 20–150 as strong, and more than 150 as very strong.

Models Fit to the Behavioral Data

To conduct computational model analysis of the actual behavioral data in the IDM task, we first applied participants' behavioral data to all CBL models passing parameter recovery. As in Simulation 2 (model recovery), WBIC was calculated as an index of model fit, and BF was used for inter-model comparison. The estimated model parameters and values of each stimulus from the best-fit model were used for further analyses.

Furthermore, to ensure that the model adequately describes actual behavioral data, we conducted a posterior predictive check (Gelman et al., 1996), which allowed the observation of discrepancies between the observed behavioral data and the data predicted by the model. Specifically, utilizing all samples from the posterior distribution, we generated 12,000 independent datasets, each containing data on the choice behavior of 38 individuals. Subsequently, we calculated the chosen frequencies of different stimuli for each individual in each dataset. Finally, we randomly selected 500 distributions and compared them with the distribution of actual behavioral data to determine whether the model can predict the patterns of observed behavioral data.

Additional Model Comparison Between the CBL and the Models Without Value Updates in IDM

The CBL models used in this study were based on those developed by the previous study (Zhu et al., 2021). Although in their study, the CBL models were validated to show a change in the value of chosen and rejected options after preference-based choices, we cannot rule out the possibility that the value of the stimuli in this study remained unchanged after choices. Therefore, an additional computational model analysis was performed to investigate whether the value did not change in the IDM. Additional models without value changes were constructed based on the best-fit CBL model for the behavioral data in Models 1–4. Specifically, we created four additional models (Table 4): Model A, in which only the value of the HP stimuli was not updated; Model B, in which both the HP and LP stimuli were not updated, and only the value of the novel stimuli was updated; Model C, in which only the value of the novel stimuli was not updated; and Model D, in which the values of all the stimuli were not updated. All new models were analyzed through simulations and the fitting of behavioral data.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42113-024-00198-5.

Author Contribution JZ and TN designed the experiment. JZ wrote the first draft. JZ, KK, MH, and TN reviewed and revised the manuscript.

Funding Open Access funding provided by Hiroshima University. This research was supported by JST COI grant no. JPMJCE1311 and JSPS KAKENHI grant no. 18K03177, 18K03173, 22K07328, and 22H01083.

Data Availability The data of each participant (correct response rate, chosen frequency, subjective evaluation, and the value of each stimulus estimated by the computational modeling analysis) are available from https://osf.io/hwe9v/ (DOI: https://doi.org/10.17605/OSF.IO/HWE9V).

Declarations

Ethical Approval Experimental protocols were approved by the Ethics Committee of the Graduate School of Humanities and Social Sciences at the University of Hiroshima. All methods were performed in accordance with the relevant guidelines and regulations (Declaration of Helsinki).

Consent to Participate All participants provided informed written consent prior to participation.

Consent for Publication Not applicable.

Conflict of Interest The authors declare no competing interests.

References

Akaishi, R., Umeda, K., Nagase, A., & Sakai, K. (2014). Autonomous mechanism of internal choice estimate underlies decision inertia. *Neuron, 81*, 195–206. https://doi.org/10.1016/j.neuron.2013.10.018

Aridan, N., Pelletier, G., Fellows, L. K., & Schonberg, T. (2019). Is ventromedial prefrontal cortex critical for behavior change without external reinforcement? *Neuropsychologia, 124*, 208–215. https://doi.org/10.1016/j.neuropsychologia.2018.12.008

Bai, Y., Nakao, T., Xu, J., Qin, P., Chaves, P., Heinzel, A., Duncan, N., Lane, T., Yen, N. S., Tsai, S. Y., & Northoff, G. (2016). Resting state glutamate predicts elevated pre-stimulus alpha during self-relatedness: A combined EEG-MRS study on "rest-self overlap." *Social Neuroscience*, *11*(3). https://doi.org/10.1080/17470919.2015.1072582

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*, 1293–1295. https://doi.org/10.1126/science.275.5304.1293

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Sciences, 9*, 159–162. https://doi.org/10.1016/j.tics.2005.02.002

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Biele, G., Rieskamp, J., Krugel, L. K., & Heekeren, H. R. (2011). The neural basis of following advice. *PLoS Biology, 9*(6), e1001089. https://doi.org/10.1371/journal.pbio.1001089

Bouton, M. E., & Moody, E. W. (2004). Memory processes in classical conditioning. *Neuroscience and Biobehavioral Reviews, 28*(7), 663–674. https://doi.org/10.1016/j.neubiorev.2004.09.001

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *Journal of Abnormal and Social Psychology, 52*(3), 384–389. https://doi.org/10.1037/h0041006

Camille, N., Griffiths, C. A., Vo, K., Fellows, L. K., & Kable, J. W. (2011). Ventromedial frontal lobe damage disrupts value maximization in humans. *Journal of Neuroscience, 31*(20), 7527–7532. https://doi.org/10.1523/JNEUROSCI.6527-10.2011

Chen, M. K., & Risen, J. L. (2010). How choice affects and reflects preferences: Revisiting the free-choice paradigm. *Journal of Personality and Social Psychology, 99*(4), 573–594. https://doi.org/10.1037/a0020217

Colosio, M., Shestakova, A., Nikulin, V. V., Blagovechtchenski, E., & Klucharev, V. (2017). Neural mechanisms of cognitive dissonance (Revised): An EEG study. *Journal of Neuroscience, 37*(20), 5074–5083. https://doi.org/10.1523/JNEUROSCI.3209-16.2017

Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology, 16*(2), 199–204. https://doi.org/10.1016/j.conb.2006.03.006

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press.

Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron, 36*(2), 285–298. https://doi.org/10.1016/S0896-6273(02)00963-7

de Greck, M., Rotte, M., Paus, R., Moritz, D., Thiemann, R., Proesch, U., Bruer, U., Moerth, S., Tempelmann, C., Bogerts, B., & Northoff, G. (2008). Is our self based on reward? Self-relatedness recruits neural activity in the reward system. *NeuroImage, 39*(4), 2066–2075. https://doi.org/10.1016/j.neuroimage.2007.11.006

Dickinson, A., & Balleine, B. (1995). Motivational control of instrumental action. *Current Directions in Psychological Science, 4*(5), 162–167. https://doi.org/10.1111/1467-8721.ep11512272

Endo, N., Saiki, J., Nakao, Y., & Saito, H. (2003). Perceptual judgments of novel contour shapes and hierarchical descriptions of geometrical properties. *Jpn J Psychol, 74*, 346–353. https://doi.org/10.4992/jjpsy.74.346

Enzi, B., de Greck, M., Prösch, U., Tempelmann, C., & Northoff, G. (2009). Is our self nothing but reward? Neuronal overlap and distinction between reward and personal relevance and its relation to human personality. *PLoS ONE, 4*(12), e8429. https://doi.org/10.1371/journal.pone.0008429

Fellinger, R., Klimesch, W., Gruber, W., Freunberger, R., & Doppelmayr, M. (2011). Pre-stimulus alpha phase-alignment predicts P1-amplitude. *Brain Research Bulletin*, *85*(6). https://doi.org/10.1016/j.brainresbull.2011.03.025

Fellows, L. K., & Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: Judgment under uncertainty or judgment per se? *Cerebral Cortex, 17*(11), 2669–2674. https://doi.org/10.1093/cercor/bhl176

Festinger, L. (1957). *A theory of social cognitive dissonance*. Stanford University Press.

Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*(4), 733–760. http://www.jstor.org/stable/24306036

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), 585–594. https://doi.org/10.1016/j.neuron.2010.04.016

Gluth, S., Rieskamp, J., & Büchel, C. (2014). Neural evidence for adaptive strategy selection in value-based decision-making. *Cerebral Cortex, 24*(8), 2009–2021. https://doi.org/10.1093/cercor/bht049

Guitart-Masip, M., Duzel, E., Dolan, R., & Dayan, P. (2014). Action versus valence in decision making. *Trends in Cognitive Sciences*, *18*(4). https://doi.org/10.1016/j.tics.2014.01.003

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America, 105*(18), 6741–6746. https://doi.org/10.1073/pnas.0711099105

Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D., & Brem, S. (2015). Cognitive flexibility in adolescence: Neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *NeuroImage, 104*, 347–354. https://doi.org/10.1016/j.neuroimage.2014.09.018

Humphreys, G. W., & Sui, J. (2016). Attentional control and the self: The self-attention network (SAN). *Cognitive Neuroscience, 7*(1–4), 5–17. https://doi.org/10.1080/17588928.2015.1044427

Ito, M., & Doya, K. (2009). Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *Journal of Neuroscience, 29*(31), 9861–9874. https://doi.org/10.1523/JNEUROSCI.6157-08.2009

Izuma, K., & Murayama, K. (2013). Choice-induced preference change in the free-choice paradigm: A critical methodological review. *Frontiers in Psychology, 4*, 41. https://doi.org/10.3389/fpsyg.2013.00041

Izuma, K., Matsumoto, M., Murayama, K., Samejima, K., Sadato, N., & Matsumoto, K. (2010). Neural correlates of cognitive dissonance and choice-induced preference change. *Proceedings of the National Academy of Sciences of the United States of America, 107*(51), 22014–22019. https://doi.org/10.1073/pnas.1011879108

Johansson, P., Hall, L., Tärning, B., Sikström, S., & Chater, N. (2014). Choice blindness and preference change: You will like this paper better if you (believe you) chose to read it! *Journal of Behavioral Decision Making*, *27*(3). https://doi.org/10.1002/bdm.1807

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Katahira, K., Fujimura, T., Okanoya, K., & Okada, M. (2011). Decision-making based on emotional images. *Frontiers in Psychology, 2*, 311. https://doi.org/10.3389/fpsyg.2011.00311

Katahira, K., Yuki, S., & Okanoya, K. (2017). Model-based estimation of subjective values using choice tasks with probabilistic feedback. *Journal of Mathematical Psychology*, *79*. https://doi.org/10.1016/j.jmp.2017.05.005

Koster, R., Duzel, E., & Dolan, R. J. (2015). Action and valence modulate choice and choice-induced preference change. *PLoS ONE, 10*(3), e0119682. https://doi.org/10.1371/journal.pone.0119682

Kunisato, Y., Okamoto, Y., Ueda, K., Onoda, K., Okada, G., Yoshimura, S., Suzuki, S. I., Samejima, K., & Yamawaki, S. (2012). Effects of depression on reward-based decision making and variability of action in probabilistic learning. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(4). https://doi.org/10.1016/j.jbtep.2012.05.007

Lee, D., & Daunizeau, J. (2020). Choosing what we like vs liking what we choose: How choice-induced preference change might actually be instrumental to decision-making. *PLoS ONE, 15*(5), e0231081. https://doi.org/10.1371/journal.pone.0231081

Lindström, B., Selbing, I., Molapour, T., & Olsson, A. (2014). Racial bias shapes social reinforcement learning. *Psychological Science, 25*(3), 711–719. https://doi.org/10.1177/0956797613514093

Marco-Pallarés, J., Cucurell, D., Cunillera, T., García, R., Andrés-Pueyo, A., Münte, T. F., & Rodríguez-Fornells, A. (2008). Human oscillatory activity associated to reward processing in a gambling task. *Neuropsychologia, 46*(1), 241–248. https://doi.org/10.1016/j.neuropsychologia.2007.07.016

Marco-Pallarés, J., Münte, T. F., & Rodríguez-Fornells, A. (2015). The role of high-frequency oscillatory activity in reward processing and learning. *Neuroscience and Biobehavioral Reviews, 49*, 1–7. https://doi.org/10.1016/j.neubiorev.2014.11.014

Miyagi, M., Miyatani, M., & Nakao, T. (2017). Relation between choice-induced preference change and depression. *PLoS ONE, 12*(6), e0180041. https://doi.org/10.1371/journal.pone.0180041

Nakamura, K., & Kawabata, H. (2013). I choose, therefore i like: Preference for faces induced by arbitrary choice. *PLoS ONE, 8*(8), e72071. https://doi.org/10.1371/journal.pone.0072071

Nakao, T., Ohira, H., & Northoff, G. (2012). Distinction between externally vs. Internally guided decision-making: Operational differences, meta-analytical comparisons and their theoretical implications. *Frontiers in Neuroscience, 6*, 1–26. https://doi.org/10.3389/fnins.2012.00031

Nakao, T., Bai, Y., Nashiwa, H., & Northoff, G. (2013). Resting-state EEG power predicts conflict-related brain activity in internally guided but not in externally guided decision-making. *NeuroImage, 66*, 9–21. https://doi.org/10.1016/j.neuroimage.2012.10.034

Nakao, T., Kanayama, N., Katahira, K., Odani, M., Ito, Y., Hirata, Y., Nasuno, R., Ozaki, H., Hiramoto, R., Miyatani, M., & Northoff, G. (2016). Post-response βγ power predicts the degree of choice-based learning in internally guided decision-making. *Scientific Reports, 6*, 32477. https://doi.org/10.1038/srep32477

Nakao, T., Miyagi, M., Hiramoto, R., Wolff, A., Gomez-Pilar, J., Miyatani, M., & Northoff, G. (2019). From neuronal to psychological noise – Long-range temporal correlations in EEG intrinsic activity reduce noise in internally-guided decision making. *NeuroImage, 201*, 116015. https://doi.org/10.1016/j.neuroimage.2019.116015

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience, 32*(2), 551–562. https://doi.org/10.1523/JNEUROSCI.5498-10.2012

Northoff, G. (2016). Is the self a higher-order or fundamental function of the brain? The "basis model of self-specificity" and its encoding by the brain's spontaneous activity. *Cognitive Neuroscience, 7*(1–4), 203–222. https://doi.org/10.1080/17588928.2015.1111868

Northoff, G., Vatansever, D., Scalabrini, A., & Stamatakis, E. A. (2022). Ongoing brain activity and its role in cognition: Dual versus baseline models. *Neuroscientist, 29*(4). https://doi.org/10.1177/10738584221081752

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences, 1104*, 35–53. https://doi.org/10.1196/annals.1390.022

Ohira, H., Fukuyama, S., Kimura, K., Nomura, M., Isowa, T., Ichikawa, N., Matsunaga, M., Shinoda, J., & Yamada, J. (2009). Regulation of natural killer cell redistribution by prefrontal cortex during stochastic learning. *NeuroImage*, *47*(3). https://doi.org/10.1016/j.neuroimage.2009.04.088

Ohira, H., Ichikawa, N., Nomura, M., Isowa, T., Kimura, K., Kanayama, N., Fukuyama, S., Shinoda, J., & Yamada, J. (2010). Brain and autonomic association accompanying stochastic decision-making. *NeuroImage*, *49*(1). https://doi.org/10.1016/j.neuroimage.2009.07.060

Palminteri, S., Lefebvre, G., Kilford, E. J., & Blakemore, S. J. (2017). Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLoS Computational Biology, 13*(8), e1005684. https://doi.org/10.1371/journal.pcbi.1005684

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Qin, P., Wang, M., & Northoff, G. (2020). Linking bodily, environmental and mental states in the self—A three-level model based on a meta-analysis. *Neuroscience and Biobehavioral Reviews, 115*, 77–95. https://doi.org/10.1016/j.neubiorev.2020.05.004

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *Journal of Neuroscience, 27*(47), 12860–12867. https://doi.org/10.1523/JNEUROSCI.2496-07.2007

Stan Development Team. (2020). *RStan: The R interface to Stan. R package version 2.21.2*. http://mc-stan.org/

Stevenson, J. G., & Clayton, F. L. (1970). A response duration schedule: Effects of training, extinction, and deprivation. *Journal of the Experimental Analysis of Behavior, 13*(3), 359–367. https://doi.org/10.1901/jeab.1970.13-359

Sugawara, M., & Katahira, K. (2021). Dissociation between asymmetric value updating and perseverance in human reinforcement learning. *Scientific Reports, 11*(1), 3574. https://doi.org/10.1038/s41598-020-80593-7

Sui, J., & Gu, X. (2017). Self as object: Emerging trends in self research. *Trends in Neurosciences, 40*(11), 643–653. https://doi.org/10.1016/j.tins.2017.09.002

Sui, J., & Humphreys, G. W. (2015). The integrative self: How self-reference integrates perception and memory. *Trends in Cognitive Sciences, 19*(12), 719–728. https://doi.org/10.1016/j.tics.2015.08.015

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*. MIT Press.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological*

Review: Monograph Supplements, 2(4), i–109. https://doi.org/10.1037/h0092987

Ugazio, G., Grueschow, M., Polania, R., Lamm, C., Tobler, P., & Ruff, C. (2021). Neuro-computational foundations of moral preferences. *Social Cognitive and Affective Neuroscience, 17*, 253–265. https://doi.org/10.1093/scan/nsab100

Vinckier, F., Rigoux, L., Kurniawan, I. T., Hu, C., Bourgeois-Gironde, S., Daunizeau, J., & Pessiglione, M. (2019). Sour grapes and sweet victories: How actions shape preferences. *PLoS Computational Biology, 15*(1), e1006499. https://doi.org/10.1371/journal.pcbi.1006499

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research, 14*(1), 867–897.

Wilson, R. C., & Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife, 8*, e49547. https://doi.org/10.7554/eLife.49547

Wolff, A., Gomez-Pilar, J., Nakao, T., & Northoff, G. (2019). Interindividual neural differences in moral decision-making are mediated by alpha power and delta/theta phase coherence. *Scientific Reports, 9*(1), 4432. https://doi.org/10.1038/s41598-019-40743-y

Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Braus, D. F., & Büchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *Journal of Neuroscience, 26*(37), 9530–9537. https://doi.org/10.1523/JNEUROSCI.2915-06.2006

Zhang, Y., Wang, F., & Sui, J. (2022). Decoding individual differences in self-prioritization from the resting-state functional connectome. *Research Square*. https://doi.org/10.21203/rs.3.rs-2204324/v1

Zhu, J., Hashimoto, J., Katahira, K., Hirakawa, M., & Nakao, T. (2021). Computational modeling of choice-induced preference change: A reinforcement-learning-based approach. *PLoS ONE, 16*(1), e0244434. https://doi.org/10.1371/journal.pone.0244434