Behavioral/Cognitive

Nutrient-Sensitive Reinforcement Learning in Monkeys

Fei-Yang Huang (黃飛揚)[1,2] and ⓘ Fabian Grabenhorst[1,2]

[1]Department of Experimental Psychology, University of Oxford, Oxford OX1 3TA, United Kingdom and [2]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3DY, United Kingdom

In reinforcement learning (RL), animals choose by assigning values to options and learn by updating these values from reward outcomes. This framework has been instrumental in identifying fundamental learning variables and their neuronal implementations. However, canonical RL models do not explain how reward values are constructed from biologically critical intrinsic reward components, such as nutrients. From an ecological perspective, animals should adapt their foraging choices in dynamic environments to acquire nutrients that are essential for survival. Here, to advance the biological and ecological validity of RL models, we investigated how (male) monkeys adapt their choices to obtain preferred nutrient rewards under varying reward probabilities. We found that the nutrient composition of rewards strongly influenced learning and choices. Preferences of the animals for specific nutrients (sugar, fat) affected how they adapted to changing reward probabilities; the history of recent rewards influenced choices of the monkeys more strongly if these rewards contained the their preferred nutrients (nutrient-specific reward history). The monkeys also chose preferred nutrients even when they were associated with lower reward probability. A nutrient-sensitive RL model captured these processes; it updated the values of individual sugar and fat components of expected rewards based on experience and integrated them into subjective values that explained the choices of the monkeys. Nutrient-specific reward prediction errors guided this value-updating process. Our results identify nutrients as important reward components that guide learning and choice by influencing the subjective value of choice options. Extending RL models with nutrient-value functions may enhance their biological validity and uncover nutrient-specific learning and decision variables.

Key words: food; learning; nutrients; preference; reward; reward prediction error

---

Significance Statement

RL is an influential framework that formalizes how animals learn from experienced rewards. Although reward is a foundational concept in RL theory, canonical RL models cannot explain how learning depends on specific reward properties, such as nutrients. Intuitively, learning should be sensitive to the nutrient components of the reward to benefit health and survival. Here, we show that the nutrient (fat, sugar) composition of rewards affects how the monkeys choose and learn in an RL paradigm and that key learning variables including reward history and reward prediction error should be modified with nutrient-specific components to account for the choice behavior observed in the monkeys. By incorporating biologically critical nutrient rewards into the RL framework, our findings help advance the ecological validity of RL models.

Introduction

According to reinforcement learning (RL) theory, animals choose by assigning values to options and learn by updating these values from experienced rewards (Sutton and Barto, 1998). This framework has been instrumental in identifying learning

and decision variables that explain animal behavior, including object values and action values, which guide choices, and reward prediction errors, which update values from experienced outcomes. Physical implementations of these concepts have been discovered in neurons of the primate dopamine system, striatum, amygdala, and frontal cortex (Schultz et al., 1997; Samejima et al., 2005; Lau and Glimcher, 2008; So and Stuphorn, 2010; Lee et al., 2012; Seo et al., 2012; Tsutsui et al., 2016; Costa et al., 2019; Grabenhorst et al., 2019a). One important factor that limits the biological validity of canonical RL models is that they do not explain how learning and choice depend on the composition of experienced rewards. Nutrients, for example, are biologically critical intrinsic components of food rewards that engage dedicated sensory and physiological mechanisms and are essential for survival (Carreiro et al., 2016; Rolls, 2020; Simpson and Raubenheimer, 2020). Thus, investigating how

nutrients influence learning and choice could enhance the biological validity of the RL framework.

From an ecological perspective, the ability of an animal to adapt its food choices to changing nutrient availabilities determines its survival and long-term health (Simpson and Raubenheimer, 2012). Indeed, foraging monkeys adapt their feeding patterns to uncertainty imposed by regional and seasonal food variations (Cui et al., 2018, 2020), for example, adjusting their diet based on the availability of nutritious foods (seeds, nuts). Primates, including humans, also exhibit subjective preferences for specific nutrients and related sensory food qualities (van der Klaauw et al., 2016; Ma et al., 2017; Takahashi et al., 2019; Huang et al., 2021). Thus, foraging animals consider both food nutrient composition and the probability of obtaining the food. However, the mechanisms underlying such adaptive, nutrient-sensitive food choices remain poorly understood.

Recent proposals in RL theory extended the reward concept to outcomes under homeostatic regulation (Keramati and Gutkin, 2014) and distinguished hedonic reward components from their postingestive consequences (Dayan, 2022). Generally, the importance of linking biological reward components to learning and choice theories is increasingly recognized (Rangel, 2013; Suzuki et al., 2017; Averbeck and Murray, 2020; de Araujo et al., 2020; Stuphorn, 2021). Yet, the field lacks experimental data on how nutrient-reward components influence choices of animals in formal RL paradigms.

We recently showed that monkeys in economic choice tasks exhibit individual preferences for specific nutrients and related sensory food qualities (e.g., oral texture; Huang et al., 2021). Preferences of monkeys were well described by subjective nutrient-value functions that linked reward nutrient composition to choices. In simulations, we showed that incorporating nutrient-value functions into canonical RL models optimized model performance when preferences prioritized specific nutrients (Huang et al., 2021). Here, we extended this approach to examine the behavior of rhesus monkeys (*Macaca mulatta*) in a dynamic foraging task involving choices between rewards with different nutrient (fat, sugar) components under varying reward probabilities. Our task captured features of food choice in ecological foraging contexts (Cui et al., 2018, 2020); the animals could adapt their choices to changing reward availabilities to obtain preferred nutrients or choose less-preferred but more available alternatives.

Previous studies showed that macaques track changing reward probabilities based on the history of recent choices and rewards (Corrado et al., 2005; Lau and Glimcher, 2005; Kennerley et al., 2006; Lee et al., 2012; Grabenhorst et al., 2019b). The influence of reward history on choice is typically modeled by linking subjectively weighted recent rewards to current choices using logistic regression (Lau and Glimcher, 2005) and by dynamically updating expected values from experienced rewards using RL models (Sutton and Barto, 1998). We followed these approaches and examined whether, in a dynamic learning task, monkeys assigned higher value to recent rewards based on the nutrient components of the rewards and whether nutrient-sensitive RL models could account for this process.

Materials and Methods

Animals
Two adult male rhesus macaques (*Macaca mulatta*) were trained in the study, monkey Ya (weight during the experiments, 17–19 kg; age, 6 years) and monkey Ym (12–13 kg; age, 6 years). The animals were trained and tested ∼1–2 h per day, 5 d per week for 6 months, interrupted by regular week-long testing breaks. Both monkeys participated in a related nutrient-choice study using the same dairy-based nutrient

rewards as in this study (Huang et al., 2021). Thus, the different sensitivities of learning from the received nutrients in the probabilistic learning task can be linked to the nutrient preferences of the same monkeys in economic choice tasks without learning. The animals were on a standard diet for laboratory macaques, composed of high-protein dry pellets (percentage of calories provided by protein, 30.36%; fat, 13.29%; carbohydrates, 56.34%), dried fruits, seeds, nuts, and fresh fruits and vegetables. We monitored the health condition and body weight of the monkeys to ensure their welfare after introducing high-calorie rewards. No effects of these rewards on the health of the animals were observed. Each testing day, the animals had *ad libitum* access to the standard diet before and after the experiments and received their main liquid intake in the laboratory. The body weight of the animals increased as expected for growing animals.

All animal procedures conformed to the National Institutes of Health *Guide for the Care and Use of Laboratory Animals*. The experiments have been regulated, ethically reviewed, and supervised by the following United Kingdom (UK) and University of Cambridge (UCam) institutions and individuals: UK Home Office, implementing the Animals (Scientific Procedures) Act 1986, Amendment Regulations 2012, and represented by the local UK Home Office Inspector; UK Animals in Science Committee; UCam Animal Welfare and Ethical Review Body; UK National Center for Replacement, Refinement, and Reduction of Animal Experiments; UCam Biomedical Service Certificate Holder; UCam Welfare Officer; UCam Governance and Strategy Committee; UCam Named Veterinary Surgeon; and UCam Named Animal Care and Welfare Officer.

Experimental Design
Nutrient rewards
We prepared nutrient-controlled liquids with 2×2 fat and sugar levels to examine whether fat and sugar biased learning from reward outcomes [Fig. 1A,B; low-fat low-sugar (LFLS); high-fat low-sugar (HFLS); low-fat high-sugar (LFHS); high-fat high-sugar (HFHS)]. The liquids were matched in flavor (peach or blackcurrant, varied in different sessions), temperature, protein, salt, and other ingredients (Table 1). We used commercial skimmed milk and whole milk (British skimmed milk and whole milk, Sainsbury's Supermarkets) as baseline low-fat and high-fat liquids and flavored the liquids with fruit juice to increase palatability. The energy content of the liquid rewards (Table 1) was estimated based on metabolizable energy using Atwater general energy conversion factors (protein, 4.0 kcal/g; fat, 9.0 kcal/g; carbohydrate, 4.0 kcal/g), following the definition of foraging efficiency in optimal foraging theory (McNamara and Houston, 1997).

Nutrient foraging task
The four nutrient reward types were associated with four untrained visual cues in each session. When a trial started, the monkeys were first presented with two of the four visual cues shown on a horizontally mounted touch monitor. They then made a choice between the two cues by touching one of two blue rectangle target stimuli shown below the cues. Following the choice, they received either a large amount (rewarded trials, 0.5 ml) or a small amount (nonrewarded trials, 0.3 ml) of the cue-associated liquid depending on its prespecified reward probability (P; see Fig. 1A). When the session started, two of the rewards (LFLS/HFHS or LFHS/HFLS) were offered in high reward probabilities ($P = 0.8$), and the other two rewards in low reward probabilities ($P = 0.2$; see Fig. 1C, block A or block B). The reward probabilities were reversed every 100 trials ($P = 0.2 \rightarrow 0.8$; $P = 0.8 \rightarrow 0.2$; see Fig. 1C,D).

Data Analysis
All data were analyzed using MATLAB 2017 software (MathWorks).

Learning curves
Learning curves were plotted by aligning reward-specific choices to the probability reversal trials. In particular, based on the probability before and after reversals, we grouped these curves into incremental ($P = 0.2 \rightarrow P = 0.8$) and decremental ($P = 0.8 \rightarrow P = 0.2$, not shown) learning curves and plotted the incremental curves in Figure 2A. Two-sample

Table 1. Nutrient content of the liquid food rewards

		Peach flavor				Blackcurrant flavor				Water
	Nutrient rewards	LFLS	HFLS	LFHS	HFHS	LFLS	HFLS	LFHS	HFHS	
Recipe	Peach juice (ml)	30	30	30	30	0	0	0	0	NA
	Blackcurrant juice (ml)	0	0	0	0	30	30	30	30	
	Skimmed milk (ml)	270	0	270	0	270	0	270	0	
	Whole milk (ml)	0	270	0	270	0	270	0	270	
	Caster sugar (g)	0	0.81[a]	20.4	21.21	0	0.81	20.4	21.21	
	Total (ml)	300	300	300	300	300	300	300	300	
Nutrient content	Calorie (kcal)	33.5	60.7	60.7	87.9	43.9	71.1	71.1	98.3	0
(per 100 ml)	Fat (g)	0.45	3.33	0.45	3.33	0.45	3.33	0.45	3.33	0
	Sugar (g)	4.50	4.50	11.30	11.30	6.80	6.80	13.60	13.60	0
	Protein (g)	3.24	3.24	3.24	3.24	3.24	3.24	3.24	3.24	0
	Salt (g)	0.11	0.10	0.11	0.10	0.14	0.13	0.14	0.13	0
	Sugar/fat (kcal/kcal)	4.444	0.601	11.161	1.508	6.716	0.908	13.432	1.815	NA

[a] We added 0.81 g of caster sugar for the HFLS reward to compensate for the slightly lower sugar content in commercial whole milk compared with the skimmed milk. NA, not applicable.

t tests were performed to compare choice probabilities averaged across the last 10 plotted trials after learning for each reward (trial 70 to trial 80 after reversals).

Learning latency
The learning latency was defined as the number of trials between the first behavioral change point after probability reversals (see Fig. 2B). The behavioral change points were identified as the significant change points of cumulative choice slopes (Gallistel et al., 2004), based on a two-sample t test with the criterion $p < 0.05$.

Averaged choice probability
We compared the averaged choice probability for each reward to indicate the reward preference (see Fig. 2C). To cancel out the influence of reward probability on choices, we truncated the final trials in unbalanced sessions and computed the averaged choice probability across the same total number of high-probability and low-probability blocks for each reward.

Logistic regression analysis
History model
We used multiple logistic regression (*fitglm* function, MATLAB) to model choices based on recent choices and reward outcomes as follows (see Fig. 3A,B):

$$logit(P_L) = \beta_0 + \beta_1 \times LeftFirst + \sum_{k=1}^{n}(\beta_{k+1} \times Cx_k) + \sum_{k=1}^{n}(\beta_{k+n+1} \times Rx_k),$$

where the probability of choosing the left option (P_L) was modeled by differential choice history (Cx_n) and reward history (Rx_n) up to recent n trials while controlling the presentation sequence ($LeftFirst = 1$, if the left option was shown first; 0, if the right option was shown first). Similarly, the choice history regressors Cx_n and reward history regressors Rx_n were defined as the differences between the history variables of the left and right options as follows:

$$Cx_n = c_n^L - c_n^R, c_n^i = \begin{cases} 1, & \text{if option } i \text{ was chosen } n \text{ trials earlier} \\ 0, & \text{if option } i \text{ was not chosen } n \text{ trials earlier} \end{cases}$$

$$Rx_n = r_n^L - r_n^R, r_n^i = \begin{cases} 1, & \text{if option } i \text{ was chosen and rewarded } n \text{ trials earlier} \\ 0, & \text{otherwise} \end{cases},$$

$$i \in \{L, R\}.$$

Notably, the history regressors for each option were indexed only when the same option was offered by assuming the unoffered options in a past trial did not influence current choices (Wittmann et al., 2020). Therefore, because of the random sampling of offered options, the n-back trials for the left option may not be the same trials as those for the right option.

Nutrient model
Based on the History model, we further included nutrient history interaction terms in the regression model to characterize the individual contribution of fat and sugar content on the effects of recent choices and reward outcomes (see Fig. 3C–G) as follows:

$$logit(P_L) = \beta_0 + \beta_1 \times LeftFirst + \sum_{k=1}^{n}(\beta_{k+1} \times Cx_k) + \sum_{k=1}^{n}(\beta_{k+n+1} \times Rx_k) + \sum_{k=1}^{n}(\beta_{k+2n+1} \times FC_k) + \sum_{k=1}^{n}(\beta_{k+3n+1} \times FR_k) + \sum_{k=1}^{n}(\beta_{k+4n+1} \times SC_k) + \sum_{k=1}^{n}(\beta_{k+5n+1} \times SR_k),$$

where FC_n denoted recent high-fat choices, and FR_n denoted high-fat rewarded trials; SC_n denoted recent high-sugar choices, and SR_n denoted high-sugar rewarded trials. The nutrient history interaction terms were defined as follows:

$$FC_n = c_{t-n}^L \times FatLv_{t-n}^L - c_{t-n}^R \times FatLv_{t-n}^R$$

$$SC_n = c_{t-n}^L \times SugarLv_{t-n}^L - c_{t-n}^R \times SugarLv_{t-n}^R$$

$$FR_n = r_{t-n}^L \times FatLv_{t-n}^L - r_{t-n}^R \times FatLv_{t-n}^R$$

$$SR_n = r_{t-n}^L \times SugarLv_{t-n}^L - r_{t-n}^R \times SugarLv_{t-n}^R,$$

where c_{t-n}^L and c_{t-n}^R indicated whether the left or right option was chosen n trials earlier (1, chosen; 0, unchosen); r_{t-n}^L and r_{t-n}^R denoted whether the left or right option was chosen and was rewarded (1, chosen and rewarded; 0, otherwise); the fat level ($FatLv$) and sugar level ($SugarLv$) of the options were defined as 1 (left > right), 0 (left = right), or −1 (left < right). Model comparison was performed based on the Akaike Information Criterion (AIC) between the History model and the Nutrient model; these models were matched in history lengths up to 10 trials in the past (see Fig. 3E).

RL models
Standard RL model (Q-learning with binary reward outcomes)
We adopted a standard Q-learning algorithm that followed the Rescorla–Wagner learning rule with binary reward outcomes (Rescorla and Wagner, 1972; Sutton and Barto, 1998). The initial reward values (Q_1^i) were set to be 0 for all options ($Q_1^i = 0, \forall i \in \{LFLS, HFLS, LFHS, HFHS\}$) and were

updated by the reward prediction errors (RPE_t) scaled by the learning rate $\alpha \in [0, 1]$ as follows:

$$RPE_t = [R_t^i - Q_{t-1}^i], R_t^i = \begin{cases} 1, if \, rewarded \\ 0, if \, otherwise \end{cases}, i \in \{LFLS, HFLS, LFHS, HFHS\}$$

$$Q_t^i = Q_{t-1}^i + \alpha \cdot RPE_t.$$

Choices were predicted by first transforming the left-right value difference δ_t via the softmax function into choice probability π_t^L as follows:

$$\pi^L(\delta)_t = \frac{1}{1 + \exp(-\beta \cdot \delta_t - \beta_0)} \in [0, 1], \delta_t = Q_t^L - Q_t^R,$$

where Q_t^L and Q_t^R were the expected values for the left and right option on trial t, β was the inverse temperature indicating the sensitivity of choice to value differences, and β_0 was the side bias independent of left-right action values. The three free parameters, learning rate (α), inverse temperature (β), and side-bias intercept (β_0) were fitted based on maximum likelihood estimation (*fminsearch* function, MATLAB) using the likelihood function $\mathcal{L}(\theta)$ below ($I_t^L = 1, I_t^R = 0$ for left choices; $I_t^L = 0, I_t^R = 1$ for right choices) as follows:

$$\mathcal{L}(\theta) = \prod_{t \in T} l_t(\theta) = \prod_{t \in T} \left[\frac{\exp(\beta \cdot Q_t^L) \times I_t^L + \exp(\beta \cdot Q_t^R - \beta_0) \times I_t^R}{\exp(\beta \cdot Q_t^L) + \exp(\beta \cdot Q_t^R - \beta_0)} \right],$$

$$\theta = \{\alpha, \beta, \beta_0\}$$

$$\hat{\theta} = arg\max_{\theta} \mathcal{L}(\theta).$$

Alternative RL models
We systematically included differential learning rates and nutrient-specific learning parameters in the RL models. Specifically, we examined nine combinatorial RL models with three model complexities (Basic, Asym, and Forget) and three nutrient-specific learning parameters (Binary, NutVal, NutValAlpha) as described below.

RL model complexity (Basic, Asym, Forget models). We included differential learning rates for rewarded (α^+), unrewarded (α^-), and unoffered (α^0) options to update the reward values as follows:

$$Q_i(t+1) = Q_i(t) + \alpha \cdot [R_i(t) - Q_i(t)],$$

$$\alpha = \begin{cases} \alpha^+, & if \, rewarded \\ \alpha^-, & if \, unrewarded \\ \alpha^0, & if \, unoffered. \end{cases} \in [0, 1]$$

In the Basic models, the agent equally updated both the rewarded and unrewarded options and kept perfect memory for the unoffered option ($\alpha^+ = \alpha^-, \alpha^0 = 0$). In the Asym model, the agent updated the rewarded and unrewarded with different learning rates, while keeping perfect memory for the unoffered rewards ($\alpha^+ \neq \alpha^-, \alpha^0 = 0$). In the Forget model, the value of the unoffered rewards decayed because of value forgetting, but the rewarded and unrewarded options were updated equally ($\alpha^+ = \alpha^-, \alpha^0 > 0$).

Nutrient-specific learning parameters (Binary, NutVal, NutValAlpha). We examined nutrient preferences by including either nutrient-specific values only (NutVal models) or additional nutrient-specific learning rates (NutValAlpha models). In the NutVal models (see Fig. 4B), the reward values depend on the reward types as follows:

$$Q_i(t+1) = Q_i(t) + \alpha \cdot [V_i(t) \times R(t) - Q_i(t)]$$

$$V_i(t) = \frac{V_F^{I_F(t)} \cdot V_S^{I_S(t)} \cdot V_{FS}^{I_F(t) \times I_S(t)}}{V_F \cdot V_S \cdot V_{FS}},$$

$$R(t) = \begin{cases} 1, & if \, rewarded \\ d, & if \, not \, rewarded \end{cases} \quad d \in [0, 1]$$

$$I_F(t) = \begin{cases} 1, i(t) = HFLS, HFHS \\ 0, i(t) = LFLS, LFHS \end{cases}, I_S(t) = \begin{cases} 1, i(t) = LFHS, HFHS \\ 0, i(t) = LFLS, HFLS \end{cases},$$

where V_F, V_S, and V_{FS} are fixed animal-specific values for fat, sugar, and their combinations, respectively, in addition to the low-nutrient baseline ingredients. Therefore, the experienced reward values $V_i(t)$ were computed based on subjective preferences for the fat level (I_F) and sugar levels (I_S) of the received reward. For computational simplicity, we constrained all reward values between 0 and 1 by normalizing them to the value of HFHS ($V_F \cdot V_S \cdot V_{FS}$). When the animals received a small reward (nonrewarded trials), the reward values were discounted by a constant $d \in [0, 1]$, which scales the experienced reward values according to the large or small reward amounts. Thus, the model included free parameters for subjective values of fat, sugar, fat–sugar interaction, and for discounting of low reward amounts.

In the NutValAlpha models (see Fig. 4C), higher learning rates are used to update the values for high-nutrient rewards as follows:

$$log\left[\frac{\alpha(t)}{1 - \alpha(t)}\right] = \alpha_0 + \alpha_F \cdot I_F(t) + \alpha_S \cdot I_S(t) + \alpha_{FS} \times [I_F(t) \cdot I_S(t)] \in \mathbb{R},$$

$$\alpha(t) \in [0, 1], \forall t \in \mathbb{N},$$

where $\alpha(t)$ denotes the learning rate to update the value of the rewarded option on trial t, which was first transformed from [0, 1] to the real domain and modified by the high-fat level (α_F), the high-sugar level (α_S), or their combination (α_{FS}), depending on the fat levels (I_F) and sugar levels (I_S). The logistic transformation ensured that the learning rates were always between 0 and 1.

Under these specifications, the Standard RL model is, therefore, equivalent to the Basic Binary model, and the best-fitting model in the main text refers to the NutVal-Forget model (Fig. 4).

Energy RL model
In the Energy model (see Fig. 4G), the reward values were determined solely by their energy content as follows:

$$Q_i(t + 1) = Q_i(t) + \alpha \cdot [V_i(t) \times R(t) - Q_i(t)]$$

$$V_i(t) = \begin{cases} \frac{1}{V_E} & , i = LFLS \\ 0.5 & , i = HFLS, LFHS \\ 1 & , i = HFHS \end{cases},$$

$$R(t) = \begin{cases} 1 & , if \, rewarded \\ d & , if \, not \, rewarded \end{cases} \quad d \in [0, 1],$$

where V_E denotes the subjective values for the high energy content of HFHS reward. Values for rewards with the middle energy levels, LFLS and HFLS, were both $\frac{1}{2}V_E$, and all values were normalized to V_E between 0 and 1.

Object value RL model
In the Object value (ObjVal) model (see Fig. 4G), the animals learn from stimulus-specific values that are free parameters fixed in each session. The value function, with all values normalized to V_{HFHS} to be constrained between 0 and 1, is as follows:

$$V_i(t) = \begin{cases} V_{LFLS}/V_{HFHS}, & i(t) = LFLS \\ V_{HFLS}/V_{HFHS}, & i(t) = HFLS \\ V_{LFHS}/V_{HFHS}, & i(t) = LFHS \\ 1, & i(t) = HFHS. \end{cases}$$

Other specifications are similar to the NutVal models, including the learning rate, value-forgetting rate, and the discount factor.

Nutrient prediction error-RL model
In the nutrient prediction error-RL (NutRPE) model (see Fig. 7), we decomposed the nutrient-specific values $Q_i(t)$ into components of fat value $Q_i^F(t)$ and sugar value $Q_i^S(t)$ as follows:

$$Q_i(t) = Q_i^F(t) \cdot Q_i^S(t) \in [0, 1], \forall t \in \mathbb{N}.$$

Importantly, these value components were updated by nutrient prediction errors (NPEs) that were computed as the discrepancies between the experienced nutrient values and the trial-by-trial expected nutrient values as follows:

$$NPE_i^F(t) = V_i^F(t) \times R(t) - Q_i^F(t), V_i^F(t) = \begin{cases} \frac{1}{v_F} & , i(t) = LFLS, LFHS \\ 1 & , i(t) = HFLS, HFHS \end{cases}$$

$$NPE_i^S(t) = V_i^S(t) \times R(t) - Q_i^S(t), V_i^S(t) = \begin{cases} \frac{1}{v_S} & , i(t) = LFLS, HFLS \\ 1 & , i(t) = LFHS, HFHS \end{cases}$$

$$R(t) = \begin{cases} 1 & , if\ rewarded \\ d & , if\ not\ rewarded \end{cases} \quad d \in [0, 1],$$

where NPE_i^F and NPE_i^S denoted the fat and sugar prediction errors for the chosen reward on trial t, $i(t)$; V_i^F and V_i^S were the fixed, animal-specific values for fat and sugar, and Q_i^F and Q_i^S were the current expected values of fat and sugar components for reward i, respectively. The nutrient values were separately updated by corresponding nutrient prediction errors as follows:

$$Q_i^F(t+1) = Q_i^F(t) + \alpha \cdot NPE_i^F(t)$$

$$Q_i^S(t+1) = Q_i^S(t) + \alpha \cdot NPE_i^S(t),$$

where $Q_i^F(t+1)$ and $Q_i^S(t+1)$ are the values for fat and sugar components updated from the previous fat and sugar value estimates, $Q_i^F(t)$ and $Q_i^S(t)$, and the NPEs for fat and sugar multiplied by the learning rate $\alpha \in [0, 1]$.

Notably, although no the fat–sugar interaction term was included when fat values and sugar values were integrated into aggregated reward values, the multiplication of fat and sugar values inherently introduced supra-additivity when fat and sugar were both present. Specifically, a fixed increase in fat value from higher fat content would give rise to a greater influence on the aggregated reward value when the sugar level is higher, and vice versa. Therefore, we used this specification to test whether the monkeys were sensitive to individual fat and sugar content and possible fat–sugar interactions, similar to our nutrient-sensitive logistic regression models and the RL models described above.

Model comparison
We performed model comparisons based on the AIC, which evaluated model prediction while penalizing excessive free parameters and computed the relative model likelihood of model i, P_i (Burnham et al., 2011) as follows:

$$P_i = exp\left(\frac{AIC_{min} - AIC_i}{2}\right).$$

The session-averaged AIC values were compared with our best-fitting model (NutVal-Forget model), and the AIC differences indicated the results of the model comparison. Figure 4F shows the overview of model comparisons across all RL models, with the gray box indicating the statistical threshold compared with the best-fitting NutVal-Forget model. The NutVal-Forget model was directly compared with the Energy model and ObjVal model (see Fig. 4G inset), and with the NutRPE model (see Fig. 7C, inset) based on AIC differences across sessions.

Additionally, we report the number of best-fitting sessions for each model. In each session, we first selected best-fitting models from the nine RL models and calculated the number of best-fitting sessions for each model. We allowed multiple best-fitting models in each session when they were statistically indistinguishable (relative model likelihood > 0.05).

Model validation
Model-independent results
In Figure 4E, we simulated choices using the best-fitting NutVal-Forget model and computed the conditional choice probabilities depending on the reward and choice history of the previous trial, as we did for actual choice data in Figure 2E.

Model recovery analysis
To evaluate how well our model comparison could distinguish between our candidate models, we performed a model recovery analysis (see Fig. 5) to compare the best-fitting models with the actual models used to generate simulated choices in our task (Wilson and Collins, 2019). Specifically, for nine RL models, we simulated choices in the actual trial conditions ($N = 23$ sessions from two monkeys) and performed our AIC-based model comparison to select the best-fitting model for each simulated session (23 sessions × 9 models = 207 simulated sessions). We used the actual trial conditions and the fitted parameters in the 23 sessions from both monkeys to approximate conditions in our model comparison. Next, we computed the conditional probabilities of identifying the correct models given the simulated model (see Fig. 5A, confusion matrix) or the fit model (see Fig. 5B, inversion matrix). The confusion matrix quantifies how sensitive this approach is to recover the underlying model, p(fit model|simulated model); the inversion matrix indicates how reliable a best-fitting model truly identifies the underlying model, p(simulated model|it model), rather than misclassification from other candidate models.

Parameter recovery analysis
To evaluate the reliability of the parameter estimates from the model fitting, we examined how well the fitted parameters of each model were correlated with the true simulated parameters, using the same simulation data as in the model recovery analysis (see Fig. 6A–H). We also reported the cross-correlation between the eight free parameters in the NutVal-Forget model because correlations between parameters might influence the reliability of parameter estimates in the model fitting (see Fig. 6I).

Data and materials availability
All data are available on reasonable request from the corresponding author.

Results

Experimental design
Two monkeys performed in a dynamic foraging task to obtain probabilistic liquid rewards with defined nutrient compositions (Fig. 1A). In each trial, the monkey viewed two visual cues, randomly drawn from a set of four cues, that were first presented sequentially before reappearing in a randomized left-right arrangement as choice options. The monkey was then required to choose between the two cues by touching a cue-
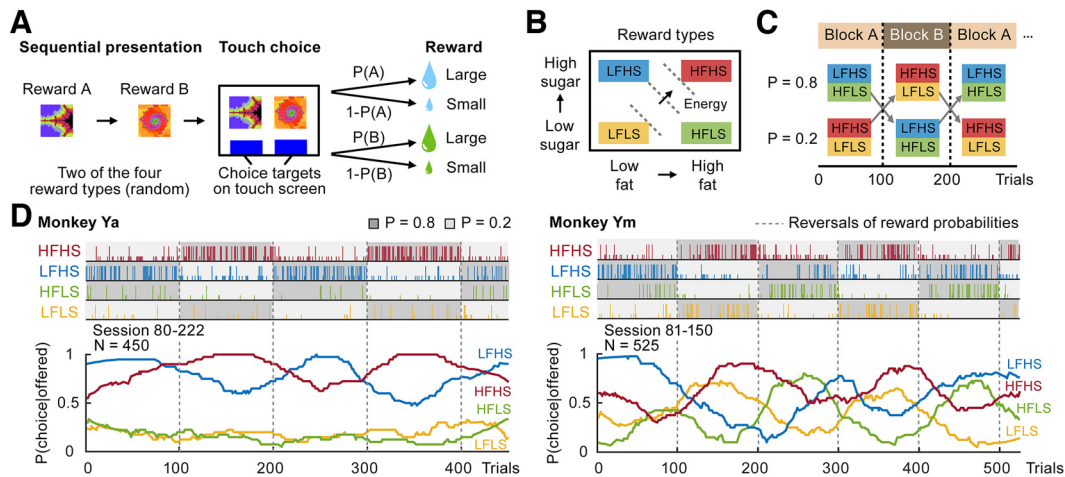
Figure 1. Dynamic foraging task with nutrient-defined rewards. *A*, Task structure. In each trial, the monkeys were first sequentially presented with two visual cues randomly drawn from a set of four cues and then made a left or right touch choice between these two cues when they were simultaneously presented. Following the touch choice, the animals received a large amount (0.5 ml) or a small amount (0.3 ml) of the associated liquid reward depending on a prespecified reward probability (*P*). *B*, Reward design. Four types of liquids with 2 × 2 factorial fat and sugar levels were offered in each session, LFLS (yellow), HFLS (green), LFHS (blue), and HFHS (red). The LFHS and HFLS liquids were isocaloric, and all rewards were matched in flavor (blackcurrant or peach) and other ingredients (e.g., protein, salt, etc.; Table 1). *C*, Reward-probability schedule. The probabilities of receiving large rewards (reward probability) were assigned in two block types. In block A, LFHS and HFLS were associated with a high probability (*P* = 0.8); LFLS and HFHS were associated with a low reward probability (*P* = 0.2). All reward probabilities were reversed in block B. Each session started with either block A or block B, and the reward probabilities changed between the two block types every 100 trials with typically three to five alternations in each session. *D*, Choices and reward outcomes in single sessions for monkey Ya (left) and monkey Ym (right). Tick marks represent choices for specific rewards; long marks indicate large-reward outcomes (rewarded trials), and short marks indicate small-reward outcomes (nonrewarded trials). Reward types in dark-gray blocks were associated with high reward probability (*P* = 0.8) and light-gray blocks were associated with low reward probability (*P* = 0.2). Choice-probability curves show nine trial running averages of choices for each reward (*N* = trials).

associated target on the touch monitor. Following the touch choice, the animal received either a large amount of liquid reward (rewarded trials) or a small amount (nonrewarded trials) depending on a prespecified reward probability (*P*). The visual cues were each associated with one of four different liquid rewards. Cue-reward associations were fixed within each session and we used new, untrained visual cues in each session to avoid influences of prior experience. We designed four liquid rewards that differed only in fat and sugar levels (Fig. 1B; LFLS, HFLS, LFHS, HFHS) while controlling the flavor (blackcurrant or peach) and other ingredients (protein, salt, etc.; Table 1). The LFLS liquid was lowest in energy content; the HFLS and LFHS liquids had matched, intermediate energy content; and the HFHS liquid was highest in energy. This reward set was the same as the one used in our previous study in an economic choice task (Huang et al., 2021).

At the start of each session, two rewards (LFLS/HFHS or LFHS/HFLS) were associated with a high probability of obtaining a large reward (*P* = 0.8), and the other two rewards were associated with a low reward probability (*P* = 0.2; Fig. 1C, block A or block B). We reversed the reward probabilities every 100 trials throughout the session (*P* = 0.2 → 0.8; *P* = 0.8 → 0.2) to encourage learning and value updating from reward outcomes. Notably, this design maintained constant availability of fat and sugar regardless of block type as there were always two high-probability and two low-probability options for both high-fat and high-sugar rewards. The design also held the available energy content constant across blocks. The liquids were matched in flavor (blackcurrant or peach) and other ingredients (protein, salt, etc.; Table 1); thus, differential learning and choice patterns could only be attributed to the nutrient content of the rewards.

In the following sections, we first identify the typical learning and choice patterns of each monkey in representative sessions and aggregated data across sessions. We then model these data using logistic regression based on reward history and compare

the performance of different RL models in accounting for the choices of the monkeys.

Nutrients bias learning and choice for probabilistic rewards: example sessions

The behavior in two example sessions (Fig. 1D) showed that both monkeys exhibited preferences for specific nutrients while tracking the changing reward probabilities. The choices of monkey Ya (Fig. 1D, left) were dominated by a general preference for high-sugar rewards, with a smaller impact of reward probability on choice. Specifically, monkey Ya chose the high-sugar rewards (LFHS, HFHS) frequently, even when they were associated with a lower probability of obtaining a large reward amount compared with the high-fat reward (Fig. 1D, blue and green curves; compare the energy-matched LFHS and HFLS rewards). In addition, choice frequencies tracked the changing reward probabilities, particularly for the high-sugar rewards (compare red and blue curves across trial blocks). In addition, the choices of monkey Ya tracked the changing reward probabilities, particularly for the high-sugar rewards (Fig. 1D, left; compare red and blue curves across trial blocks). By contrast, the choices of monkey Ym (Fig. 1D, right) reflected both a preference for nutrient content and a strong dependence on reward probability. Although monkey Ym also preferred high-nutrient rewards over low-nutrient rewards with matched reward probabilities (Fig. 1D, right; compare red and yellow curves in 200–300 trials, blue and green curves in 100–200 trials), he chose the low-nutrient reward more frequently when it was associated with a high reward probability (yellow curve, e.g., 100–200 trials), indicating a noticeable but weaker preference for fat and sugar than monkey Ya (compare blue and green curves, e.g., 300–400 trials).

Two main results emerged from these single-session data. First, the choices of the monkeys were sensitive to the nutrient content of the rewards; both monkeys preferred sugar over fat
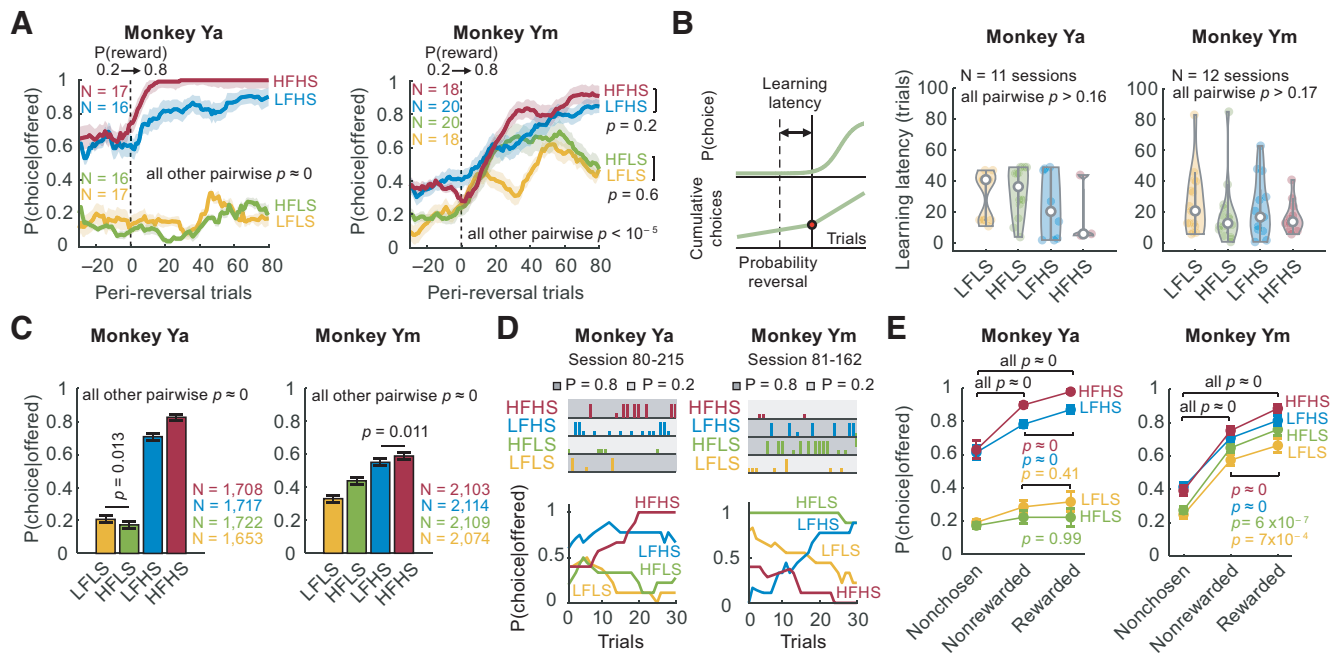
Figure 2. Nutrient-specific learning and choice patterns across sessions. ***A***, Learning curves. Mean choice frequencies (nine trial running averages ± SEM) aligned to reward probability reversals (dashed line, $P = 0.2 \rightarrow 0.8$) indicate how choices responded to changes in reward probabilities depending on reward nutrient content. N, Number of low to high reward probability reversals. Two-sample t test on the choice probability averaged across the 70th and 80th trials after reversals. ***B***, Learning latency was defined as trial intervals between probability reversals and the first significant change of choice patterns (see Materials and Methods). We included only the first probability reversal across sessions to avoid influences of different prereversal choice probabilities on learning latency. Median ± 95% confidence interval; p values, pairwise two-sided Wilcoxon rank sum test. ***C***, Reward preference. Averaged choice frequencies (mean ± SEM) indicate preferences for the four reward types. The choice frequencies were aggregated across sessions, which were truncated to have balanced block types (see Materials and Methods); p values, pairwise two-sample binomial test. $N =$ trial numbers. ***D***, Initial learning from novel visual cues. Choices and reward outcomes in the initial 30 trials of two example sessions for monkey Ya (left, block type A) and monkey Ym (right, block type B) show how the monkeys differentially associated novel visual cues to the reward types while learning from reward outcomes. Tick marks represent choices for specific rewards; long marks indicate large-reward outcomes (rewarded trials), and short marks indicate small-reward outcomes (nonrewarded trials). Reward types in dark-gray blocks were associated with high reward probability ($P = 0.8$) and light-gray blocks were associated with low reward probability ($P = 0.2$). Choice-probability curves show nine trial running averages of choices for each reward. ***E***, History-dependent choice probabilities. The probability of choosing each reward (mean ± SEM) increased after choosing and receiving the specific reward. Such influence of reward history depended on the fat and sugar content of the reward; p values, pairwise two-sample binomial test. The conditional probability of choosing each reward was computed based on the outcomes when the reward was last offered. Left, The reward was not chosen (Nonchosen). Trial numbers LFLS, HFLS, LFHS, HFHS = 1775, 1909, 622, 342 (Ya); 1677, 1507, 1117, 1042 (Ym). Middle, The reward was chosen, but only a small reward was delivered (Nonrewarded). Trial numbers LFLS, HFLS, LFHS, HFHS = 472, 425, 1707, 1996 (Ya); 967, 1171, 1545, 1616 (Ym). Right, The reward was chosen and a large reward was delivered (Rewarded). Trial numbers LFLS, HFLS, LFHS, HFHS = 231, 206, 971, 1099 (Ya); 563, 704, 859, 953 (Ym).

content, although this preference was more pronounced in monkey Ya. Second, although both monkeys tracked changing reward probabilities, they differed in the way in which they integrated reward probabilities with nutrient preferences to make choices, with monkey Ym showing a more balanced integration of reward probability and nutrient content. These results could not be explained by canonical RL models that would only consider binary reward outcomes without also considering how different reward components that are common to different rewards (i.e., fat and sugar) may affect learning and choice.

Nutrients bias learning and choice for probabilistic rewards: aggregated data

The choice patterns observed in single sessions were also found in aggregated data across sessions. Both monkeys successfully tracked changing reward probabilities by choosing an option more frequently when its reward probability switched from low ($P = 0.2$) to high ($P = 0.8$), as evident by averaged choice probabilities around reward-probability reversal points (Fig. 2A). Notably, higher reward probability increased the choices of monkey Ya only for the high-sugar stimuli but not for the low-sugar stimuli, which suggested that learning depended on the nutrient composition of the reward. By contrast, in monkey Ym, an increase in reward probability led to an increase in choices for all stimuli.

We did not find strong evidence that the monkeys adjusted their choices more quickly to changed reward probabilities for the high-sugar and high-fat rewards, indicated by learning latencies (number of trials after probability reversals before significant changes in choice behavior; Fig. 2B). Overall, the monkeys responded differently to probability changes for rewards that differed in fat and sugar content.

When choices stabilized after probability reversals, monkey Ya showed a strong preference for high-sugar rewards, in particular the combination of high-sugar and high-fat content, whereas monkey Ym showed graded preferences for high-sugar over high-fat rewards, followed by the low-nutrient option (Fig. 2A; 60–80 trials postreversals). These preferences were evident in averaged choice probabilities for the four reward types across truncated sessions with balanced trial types (Fig. 2C; see above, Materials and Methods). The choice patterns reflecting nutrient preferences and adaptations to current reward probabilities typically emerged quickly within the first 30 trials of each session when novel visual cues were introduced (Fig. 2D). Importantly, both monkeys preferred the high-sugar (LFHS) reward over the energy-matched high-fat (HFLS) reward. This result is important because it shows that the monkeys preferred specific nutrients, rather than being indifferent between nutrients of the same energy content,
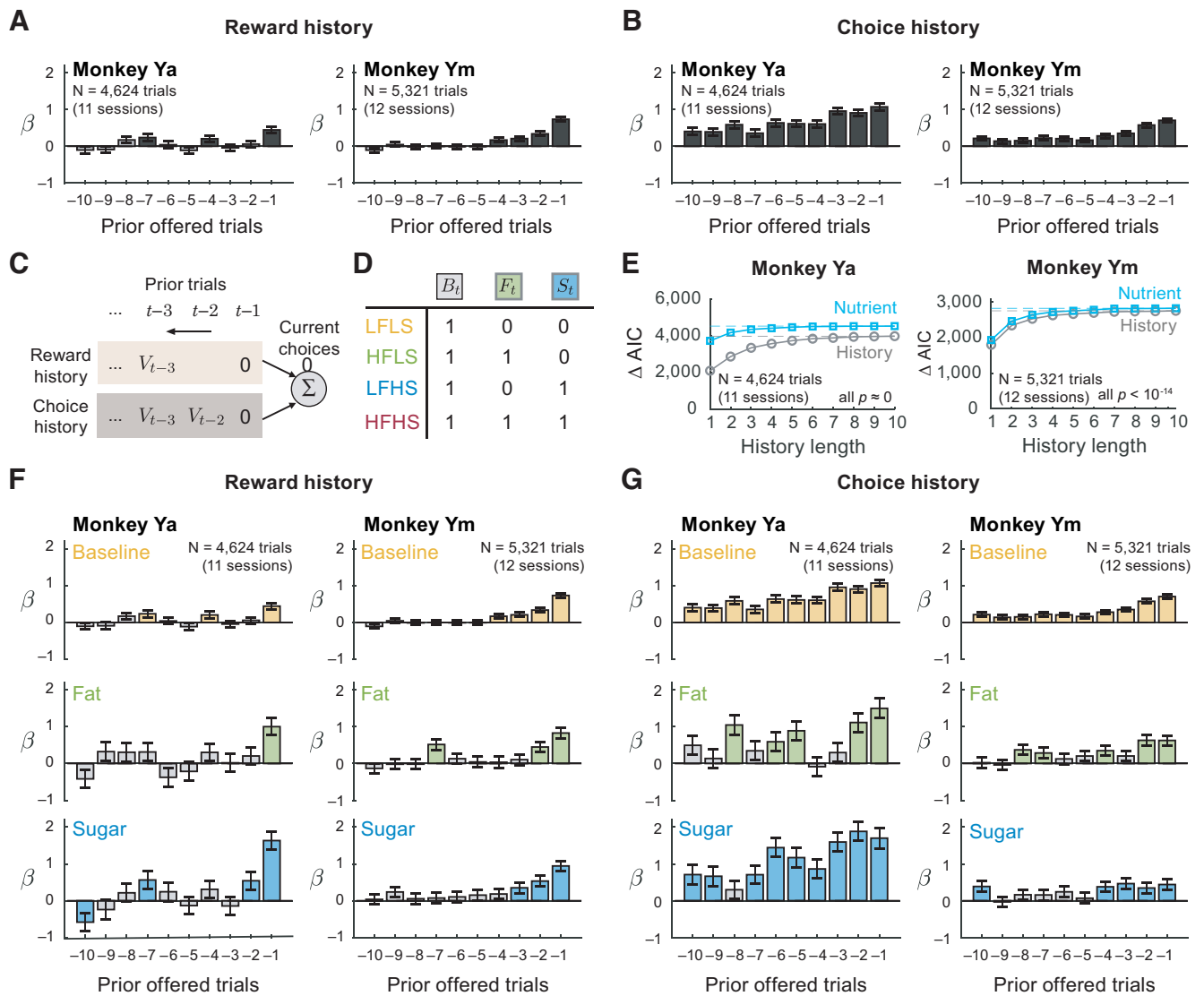
Figure 3. Nutrient-dependent influences of reward and choice history on choice. *A*, *B*, General reward- and choice-history effects in logistic regression. Reward-history effects in logistic regression (*A*). Regression coefficients ($\pm$ SEM) for reward history, modeled across all stimuli, reveal baseline effects of recent large versus small rewards on choice in the last 10 trials preceding current-trial choice. Filled bars, $p < 0.05$; gray bars, not significant. Choice history effects in logistic regression (*B*). Regression coefficients for choice history, obtained from the same logistic regression model as in *A*, reveal the baseline effect of recent choice on current-trial choice regardless of reward outcome. *C–G*, Nutrient-dependent logistic regression model. Schematic of reward and choice history logistic regression model (*C*). Prior trials were indexed only when the same type of reward was offered. Current-trial choices were explained by recently rewarded and chosen rewards. Nutrient-specific history regressors (*D*). In the nutrient-specific logistic regression model (Nutrient model), each reward outcome and choice on trial t was decomposed into effects of baseline low-nutrient ingredients (B_t, gray), additional fat content (F_t, green), and additional sugar content (S_t, blue). Model comparison across history trial lengths (*E*). Performance of the Nutrient model (blue) and History model (gray) matched in history trial lengths up to the past 10 trials. Models were compared based on ΔAIC = AIC (history trial length = 0) − AIC (history trial length = i, $i = 1,2,...,10$) and confirmed by the loglikelihood test (p value). Nutrient-specific reward- and choice-history effects (*F,G*). Effects of nutrient-specific reward history (*F*). Regression coefficients for recent low-nutrient baseline rewards (yellow), fat-containing rewards (green), and sugar-containing rewards (blue) on current-trial choice. Nutrient-specific effects were estimated in the same model as low-nutrient baseline effects; thus, effects of fat and sugar reward history were not accounted for by general effects of reward history. Effects of nutrient-specific choice history (*G*). Regression coefficients for recent low-nutrient baseline choice (yellow), the choice for fat-containing rewards (green), and the choice for sugar-containing rewards (blue) on current-trial choice. Nutrient-specific effects were estimated in the same model as low-nutrient baseline effects; thus, effects of fat and sugar choice history were not accounted for by general effects of choice history.

as would be suggested by ecological energy-maximization models. The nutrient preferences of the monkeys in the present learning task were broadly similar to the nutrient preferences of the same monkeys in a previously investigated choice task (Huang et al., 2021), with the exception that monkey Ya showed less preference for fat reward in the current task.

To assess whether nutrient content influenced learning from recent reward experiences, we compared the conditional choice probabilities of the four reward types based on whether they were chosen or rewarded in the preceding trial (Fig. 2E). Both monkeys demonstrated a baseline preference for high-sugar

rewards; the monkeys were about two times (Ym) or three times (Ya) more likely to choose high-sugar rewards (HFHS and LFHS) than low-sugar rewards (HFLS and LFLS; Fig. 2E, left). Importantly, both monkeys were more likely to repeat choices for recently received large rewards compared with small rewards, except for low-sugar rewards in monkey Ya (Fig. 2E), conditional choice probability after small rewards (Nonrewarded, middle), and large rewards (Rewarded, right)). Notably, because we introduced novel visual cues for each session, these results were not because of preferences for specific visual stimuli. Similarly, these results cannot be simply explained by the preference for calories
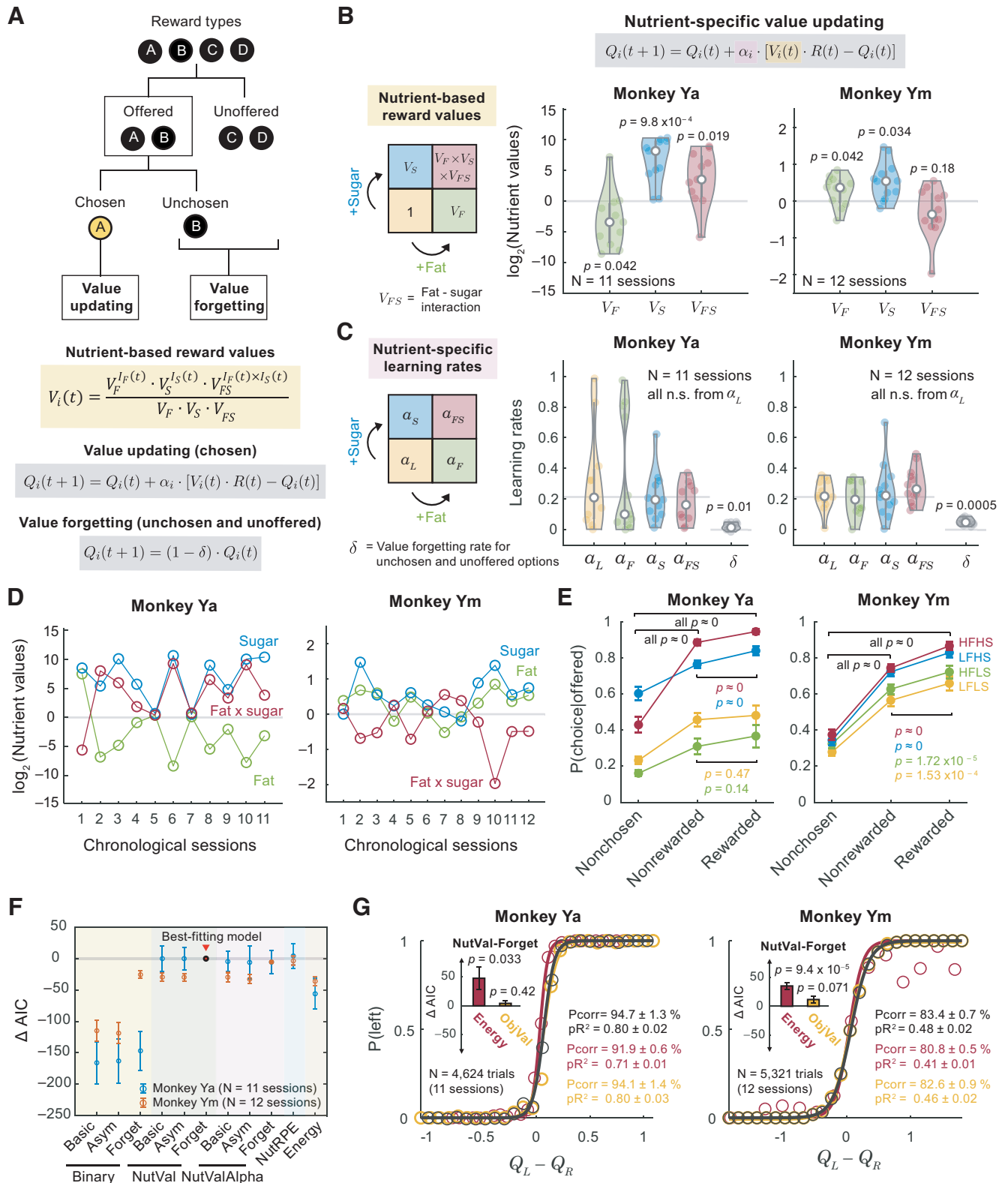
Figure 4. Nutrient-sensitive reinforcement learning. ***A***, Nutrient-sensitive RL model (NutVal-Forget model; see above, Materials and Methods). Expected values for each option (Q_i) were iteratively updated based on subjective reward values, V_i, constructed by animal-specific values for high-fat content (V_F), high-sugar content (V_S), and fat–sugar interaction (V_{FS}), compared with the low-nutrient baseline reward ($V = 1$). $I_F(t)$ and $I_S(t)$ indicate the fat and sugar levels of the reward, respectively. Q_i refers to the value for reward i, which was updated depending both on the reward outcome on trial t, $R(t)$, and the reward type received on trial t, $i(t)$. All reward values were normalized to the highest reward value ($V_F \cdot V_S \cdot V_{FS}$) to be constrained between 0 and 1. Values for unrewarded and unoffered options decayed by a factor of $(1 - \delta)$. δ, forgetting rate $\in [0, 1]$. ***B***, Subjective nutrient values. Fitted parameters indicating subjective values for fat, sugar, and fat–sugar interaction in each session fitted by the NutVal-Forget model. Data were log-transformed (base 2) for visualization; p value, Wilcoxon signed-rank test. Gray lines indicate reference values as null hypotheses for each parameter. ***C***, Nutrient-specific learning rates (compared with low-nutrient baseline α_L) and the value-forgetting rate δ (compared with $\delta = 0$, perfect value memory), fitted in the NutValAlpha-Forget model; p value, Wilcoxon signed-rank test. Gray lines indicate reference values as null hypotheses for each parameter. ***D***, Temporal dynamics of nutrient values. Nutrient values were plotted across chronological sessions for both monkeys. ***E***, Nutrient-based learning behavior of the NutVal-Forget model. Simulated
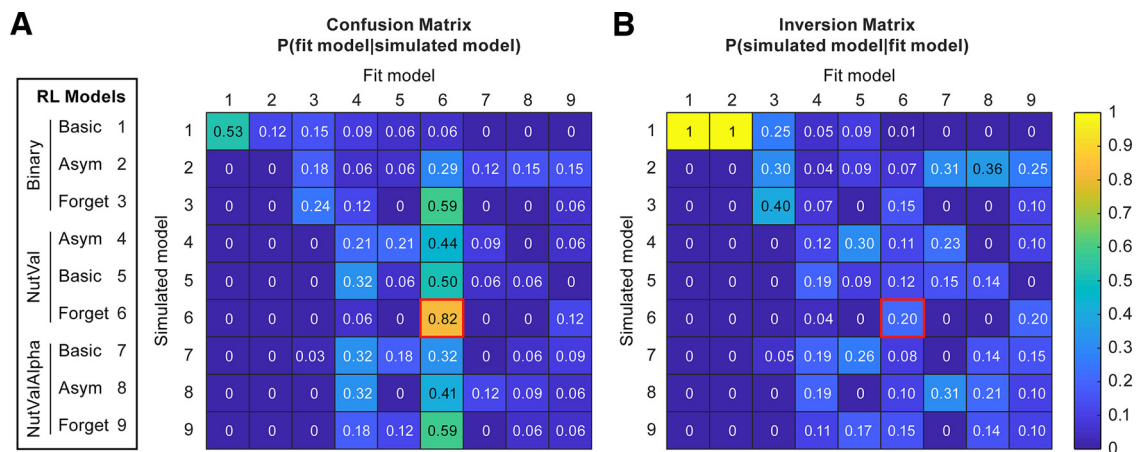
Figure 5. Model recovery analysis. **A, B,** We evaluated the sensitivity and reliability of our model comparison based on (**A**) the proportion of correctly recovered models from a specified simulated model (confusion matrix) and (**B**) the confidence of best-model predictions using our approach (inversion matrix; see above, Materials and Methods). Colors indicate the conditional probabilities stated above each matrix. Left, The nine combinatorial RL models. Red boxes in the matrices indicate data for our best-fitting NutVal-Forget model. The model comparison correctly identified 82% of the simulated sessions from the NutVal-Forget model (compare the conditional probabilities in the sixth row of **A**). Additionally, the NutVal-Forget model was the most probable generative model when predicted as the best-fitting model (compare the posterior probabilities in the sixth column of **B**).

because energy-matched rewards (i.e., LFHS and HFLS) had distinct effects on subsequent choices depending on their fat and sugar content (Fig. 2E, green and blue curves). Thus, the monkeys learned differently from different recent rewards, according to their nutrient preferences.

These aggregated data showed that the monkeys had subjective preferences for specific high-nutrient rewards and tracked changing reward probabilities depending on the nutrient content of the rewards. Because visual cues changed every session, the results were not explained by preferences for specific visual stimuli. Thus, the nutrient composition of the food rewards and the individual preferences of the animals for sugar and fat biased learning and choice. We next modeled these effects in trial-by-trial data using reward-history regression models and a formal RL approach.

Nutrient-specific reward history and choice history influence choices of monkeys
To formally characterize nutrient-dependent learning mechanisms, we followed approaches from previous studies that used logistic regression to link the choices of monkeys to the recent history of rewards and choices (Corrado et al., 2005; Lau and Glimcher, 2005). In this approach, current-trial choices are modeled by the sum of (1) subjectively weighted recent rewards

←

choices based on the NutVal-Forget model reproduced nutrient-specific learning observed in Figure 2E. The probability of choosing each reward (mean ± SEM) increased with previous reward outcomes but to different extents depending on reward fat and sugar content; p values, pairwise two-sample binomial test. **F,** Model comparison across RL models including the nine combinatorial RL models, the NutRPE model (Fig. 5) and the Energy model (Fig. 4G). The model comparison was conducted based on the AIC across testing sessions (mean ± SEM) for monkey Ym (blue) and monkey Ym (orange). The gray line indicates the statistical decision threshold (relative likelihood of a given model < 0.05; see above, Materials and Methods) compared with the best-fitting NutVal-Forget model (red arrowhead). Comparisons between any two of the other models can also be performed by taking their AIC differences. **G,** Psychometric curves relating model-derived reward values to choice probability. Model-fitted reward values from the NutVal-Forget model (black) outperformed the Energy model (Energy RL, red) and performed equally well as the ObjVal model in explaining choices of the monkeys (see above, Materials and Methods). Inset, $\Delta AIC = AIC_{Energy} - AIC_{Nutrient}$. Pcorr, Percentage correctly modeled choices ± SEM pR^2, pseudo-R^2 ± SEM; *p < 0.05, **p < 0.01, ***p < 0.001. n.s.: not significant.

resulting from choices for specific options (reward history) and (2) subjectively weighted recent choices for specific options (choice history); the subjective weights are determined by fitting the logistic-regression model to trial-by-trial data of current choices, past choices, and past rewards of an individual animal. To establish a baseline reference, we first modeled the trial-by-trial choices of the monkeys with a logistic regression that accounted for whether options were chosen in previous offers (choice history) and whether previous choices were rewarded (reward history), regardless of the nutrient composition of the rewards (History model; see above, Materials and Methods). Because options were randomly offered in our experiments, and not every reward appeared on each trial, we indexed the recent rewards and choices only when the specific option was offered (see above, Materials and Methods). In this History model (Fig. 3A,B), the most recent rewards and choices were associated with the highest regression coefficients, indicating a strong effect of recent reward and recent choice on current choice, whereas regression coefficients for more remote rewards and choices declined as a function of past trials, as observed in previous studies (Corrado et al., 2005; Lau and Glimcher, 2005).

We next modeled the contribution of individual nutrient components (Nutrient model, see above, Materials and Methods; Fig. 3C,D) by including nutrient-history interaction regressors that decomposed aggregated influences of reward and choice history (V_t) into contributions from the low-nutrient baseline liquid (B_t, yellow), high-fat content (F_t, green), and high-sugar content (S_t, blue). The Nutrient model outperformed alternative models that encoded only binary outcomes of reward and choice history but ignored reward nutrient compositions (History model; see Materials and Methods), with model comparisons being robust for different history lengths (Fig. 3E). Coefficients from this Nutrient model showed that although more recent rewards and choices generally had stronger effects on current choices, these effects depended on the nutrient composition of recent rewards (Fig. 3F,G). Specifically, in monkey Ya, reward-history coefficients for recent low-nutrient rewards were small, whereas reward-history coefficients for the most recent high-nutrient rewards, especially high-sugar rewards, were large and highly significant (Fig. 3F, left). For monkey Ym, reward-history coefficients for both the baseline and high-nutrient rewards were large and significant and
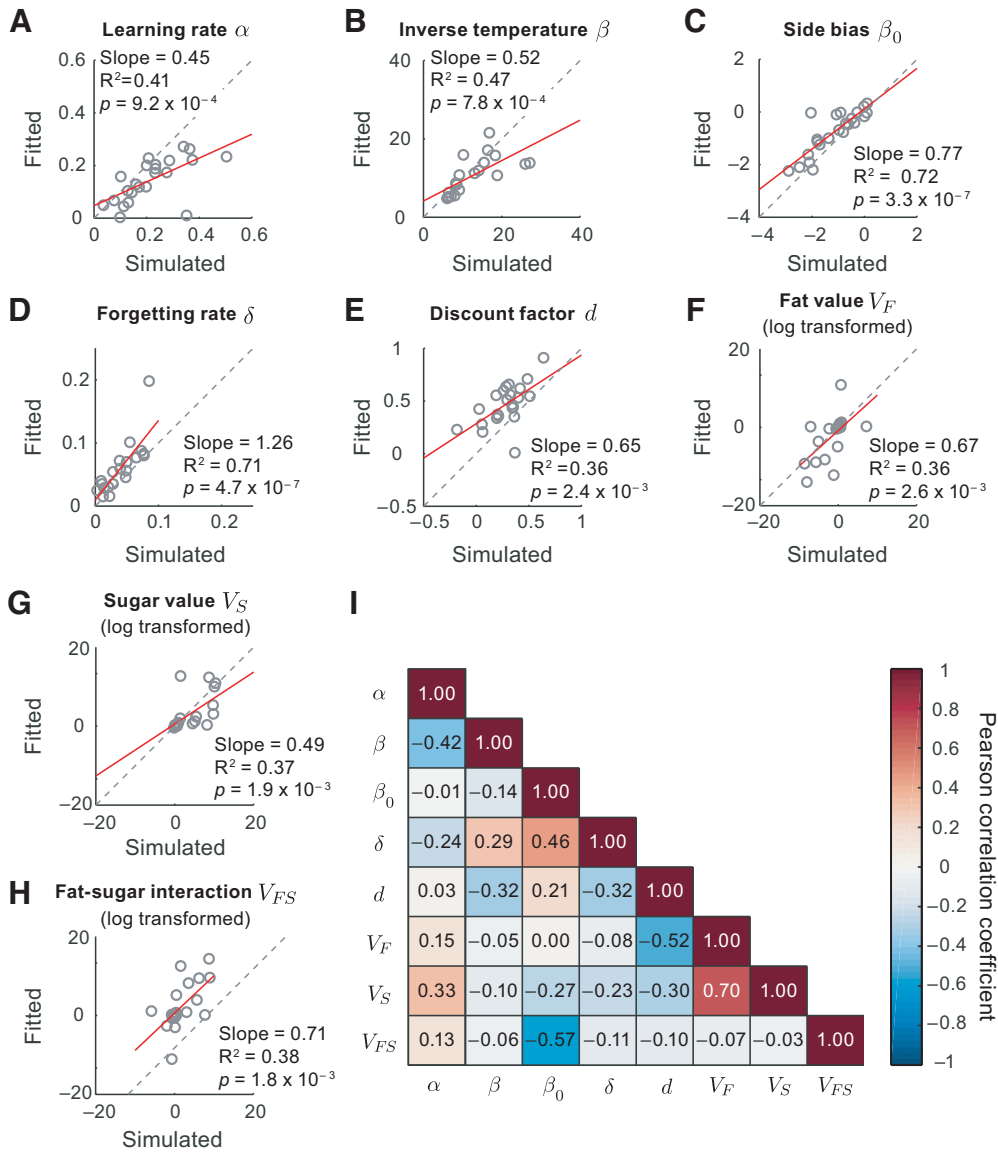
Figure 6. Parameter recovery analysis. **A–H**, Correlations between simulated and fitted parameters for each of the eight free parameters in the NutVal-Forget model, including (**A**) learning rate, α; (**B**) inverse temperature, β; (**C**) side bias, β_0; (**D**) value-forgetting rate, δ; (**E**) discount factor, d; (**F**) fat value, V_F; (**G**) sugar value, V_S; and (**H**) fat–sugar interaction, V_{FS}. The simulation was performed based on session-specific fitted parameters from each monkey to approximate the valid range of parameters (see above, Materials and Methods). The value parameters in **F–H** were log transformed (base 2). Dashed line, Unity line; red line, least-square regression line. Inset, Slope, the fitted slope of the regression line; p value was estimated based on the least-square linear fit of the data points. **I**, Parameter trade-off. Cross-correlation between the eight fitted parameters in **A–H** identified mutual dependence between free parameters in the NutVal-Forget model.

declined for more remote rewards, indicating both nutrient-specific and nutrient-independent effects of reward history (Fig. 3F, right). In both monkeys, nutrient-specific choice history effects for fat and sugar were large and significant for recent choices and declined for more remote choices (Fig. 3G). Critically, because the nutrient-specific history coefficients were modeled alongside the baseline history coefficients, they explained distinct parts of the variation in the choices of the monkeys. In other words, significant coefficients for fat- and sugar-specific reward and choice history could not be explained by general, nutrient-independent history effects, which were captured by the baseline regressors. Notably, nutrient-specific reward history captured effects on choices that derived from the difference between rewarded and unrewarded choices for particular nutrients. By contrast, choice history captured effects related to recent

choices for particular nutrients regardless of reward outcome (Lau and Glimcher, 2005); choice history effects, therefore, reflected the preference of the monkeys for particular rewards independent of reward probability. (Accordingly, in a supplemental analysis without choice-history regressors, nutrient-specific reward-history coefficients increased to reflect nutrient preferences of the monkeys as well as the effect of reward probability.)

In summary, logistic-regression analysis showed that the choices of the monkeys were influenced by the history of recently obtained fat and sugar rewards and by the history of recently made choices for fat and sugar rewards. These nutrient-specific effects of reward history and choice history were not accounted for by the separately modeled general (i.e., nutrient independent) reward- and choice-history effects.

Reinforcement learning based on nutrient-specific values

Having established that the choices of the monkeys depended on the reward history for specific nutrients, we next modeled the effect of nutrients on learning and choice within the RL framework. Learning from rewards is formalized by RL models that update trial-by-trial expected values for each option based on reward outcomes. However, our results show that the monkeys were sensitive not only to whether they received a reward but also to the nutrient composition of the reward. Accordingly, RL models may require nutrient-specific parameters to account for choices in dynamic environments with varying nutrient-reward composition as tested here. We therefore developed a novel, nutrient-sensitive RL model (NutVal-Forget model, see above, Materials and Methods) that incorporated a subjective nutrient-value function to capture how specific nutrients (fat, sugar) may differentially influence the trial-by-trial updating of expected reward values and their influence on choice (Fig. 4A). Instead of updating the value of the chosen option with a binary reward outcome, the NutVal-Forget model updated values based on animal-specific subjective values for fat, sugar, and fat–sugar interaction, as follows (We note that this model assumes that the animals are sensitive to fat, sugar, and their interaction, and combine these nutrient valuations into integrated values for decision-making.):

$$Q_i(t + 1) = Q_i(t) + \alpha \cdot [V_i(t) \times R(t) - Q_i(t)]$$

$$V_i(t) = \frac{V_F^{I_F(t)} \cdot V_S^{I_S(t)} \cdot V_{FS}^{I_F(t) \times I_S(t)}}{V_F \cdot V_S \cdot V_{FS}},$$

$$R(t) = \begin{cases} 1 & \text{, if rewarded} \\ d & \text{, if not rewarded} \end{cases} \quad d \in [0, 1]$$

$$I_F(t) = \begin{cases} 1, i(t) = HFLS, HFHS \\ 0, i(t) = LFLS, LFHS \end{cases},$$

$$I_S(t) = \begin{cases} 1, i(t) = LFHS, HFHS \\ 0, i(t) = LFLS, HFLS \end{cases}.$$

In this model, the expected value for reward i, Q_i, was updated depending on both the reward outcome on trial t, $R(t)$, (rewarded or not rewarded) and the chosen reward type on trial t, $i(t)$. The animal-specific subjective value for reward i, $V_i(t)$, was constructed by the subjective valuation of the fat level of the reward, $I_F(t)$, and sugar level, $I_S(t)$, based on fixed subjective values for fat (V_F), sugar (V_S), and their interaction (V_{FS}). Thus, we use Q_i to refer to the trial-by-trial updated values in the RL model, following the term Q-value in the RL literature. By contrast, we use V_i to refer to the animal-specific subjective values for nutrient components, which were fixed in each session. Additionally, we included a discount factor $d \in [0, 1]$ as a free parameter to capture the reduced but nutrient-dependent effects of the small rewards compared with the large rewards (Fig. 2E). Thus, the free parameters in the nutrient-sensitive RL model included the three nutrient-value parameters—related to fat, sugar, and their interaction—the discount factor for the small reward, as well as other standard RL parameters (see above, Materials and Methods). Because subjective values were measured compared with the low-nutrient baseline reward, nutrient-value parameters (V_F, V_S) larger than 1 suggested a preference for the respective nutrient; a V_{FS} parameter larger than 1 indicated supra-additive effects of fat and sugar on subjective value.

Without loss of generality, we normalized all reward values to the value of the high-fat high-sugar reward ($V_F \cdot V_S \cdot V_{FS}$), to constrain all value parameters between 0 and 1. For options that were not chosen (unchosen) and not offered (unoffered) on a given trial, we allowed their expected values to decay as follows:

$$Q_j(t + 1) = Q_j(t) \cdot (1 - \delta), \forall j \neq i(t),$$

where the values of the unchosen and unoffered rewards, $Q_j(t)$, were discounted according to a forgetting rate (δ), which would be 0 for perfect value memory. We referred to this RL model as nutrient-sensitive because it differentially learned from reward outcomes based on their nutrient content. Specifically, different from canonical RL models, our model constructed reward values for choice options based on subjective values for the nutrient components of each reward. Although canonical RL models can be applied to situations with multiple outcomes, they would not estimate values based on the nutrient (or other intrinsic) components of the outcomes that are common to different outcomes.

We fitted the NutVal-Forget model to the trial-by-trial choice data and reward outcomes in each session, separately for both monkeys (see above, Materials and Methods). The resulting best-fitting parameter values for the nutrient components indicated that across sessions, both monkeys assigned higher values to the high-sugar rewards and that monkey Ym assigned a higher value to fat but monkey Ya did not (Fig. 4B). Fat and sugar showed a small but significant supra-additive effect on choices in monkey Ya and a small, nonsignificant negative interaction in monkey Ym. In an extended model (NutAlphaVal-Forget model; see above, Materials and Methods), we examined whether nutrients had additional influences on the learning rates in RL models, separate from their effect on reward values, but did not find evidence for nutrient-specific learning rates in either monkey (Fig. 4C). (Further, learning latencies resulting from simulated choices for the actual trial sequences that the animals experienced using the NutVal-Forget model showed no significant differences among the four rewards; $p > 0.08$ for all pairwise comparisons.) The nutrient values in the NutVal-Forget model were largely stable across testing sessions, despite some session-specific variation (Fig. 4D). Additionally, both monkeys showed evidence of forgetting the values of unoffered and unchosen options, according to the small but significant value-forgetting rates (Fig. 4C). In both monkeys, our main NutVal-Forget model reproduced the behavioral signatures of nutrient-specific learning from Figure 2E (Fig. 4E). In model comparisons across sessions, the NutVal-Forget model outperformed alternative RL models with variations on value updating (Basic, Asym, Forget) and nutrient-specific parameters (Binary, NutVal, NutValAlpha; Fig. 4F; see above, Materials and Methods). Furthermore, psychometric curves illustrated that subjective values derived from the NutVal-Forget model accounted for the choices of the monkeys, confirmed by model-fit indicators across sessions (pseudo-R^2, monkey Ya = 0.80 ± 0.02; monkey Ym = 0.48 ± 0.02; percentage of correctly modeled choices, monkey Ya = $94.7 \pm 1.3\%$; monkey Ym = $83.4 \pm 0.7\%$; Fig. 4G, black).

To rule out the possibility that the differential learning from rewards was driven by the energy content of nutrients rather than by specific nutrients, as assumed in ecological foraging models, we compared the performance of our main NutVal-Forget model (Fig. 4G, black), which estimated separate effects of sugar and fat components on subjective values, to an RL model based on the energy content of the rewards (Energy model; Fig.

4G, red). The key difference of the Energy model was that it learned equally from the isocaloric HFLS and LFHS rewards, although they differed in fat and sugar levels (Fig. 1B). In both monkeys, the NutVal-Forget model (Fig. 4G, black and inset) provided a significantly better fit to the choices of the monkey than the Energy model (Fig. 4G, red data and inset), consistent with the preference of the monkeys for high-sugar over high-fat stimuli of matched energy content (Fig. 2C).

For completeness, we also constructed an object-specific RL model with separate values for the four different rewards as free parameters that were fixed in each session (ObjVal model). Different from this object-specific model, which specified separate values for all available rewards, our main NutVal-Forget model specified values for the two nutrient components that were common to the different rewards. Our main NutVal-Forget model performed similarly to the ObjVal model, despite having one fewer free parameter (Fig. 4G, yellow). Although both models performed similarly in this dataset, we note that the nutrient-based model is more flexible compared with the stimulus-specific model because it uses a value function for particular reward components (fat, sugar) rather than for each individual reward. Thus, the nutrient-based model could use the same value function for increasing numbers of different rewards (provided they were based on common nutrient components), whereas the object-specific model would require one additional free parameter for each reward type.

We tested the sensitivity and reliability of our model comparison in a model recovery analysis (Fig. 5; see above, Materials and Methods). We found that the model comparison correctly identified 82% of the simulated sessions from the NutVal-Forget model (Fig. 5A) and that the NutVal-Forget model was the most probable generative model given our model comparison result (Fig. 5B). Notably, the winning NutVal-Forget model also generalized well to other candidate models (Fig. 5A), likely because of the negligible influences of nonrelevant, additional parameters, for example, nutrient-specific learning rates (Fig. 4C). The NutVal-Forget model was also associated with relatively lower absolute confidence (Fig. 5B, red square), likely because it was compared with a large number of nested (i.e., overall similar) models. We also tested the identifiability of the parameters in a parameter recovery analysis and examined the trade-offs between parameters (Fig. 6). Importantly, the reliability of the parameter estimates in Figure 4B was supported by the successful recovery of model parameters from simulated sessions (Fig. 6A–H) despite some correlation between V_F and V_S (Fig. 6I). Further, the parameter-correlation matrix (Fig. 6I) indicated a trade-off between specific parameters; (1) we found a positive correlation between V_F and V_S, suggesting that preferences for fat and sugar covaried across sessions; (2) we found that the right-side bias (β_0) was positively correlated with the forgetting rate (δ) and inversely correlated with V_S and V_{FS}, indicating that nutrient-influences on choices were weaker in periods when monkeys were apparently less focused on the task (displaying side-bias and higher forgetting rate); and (3) we found that the discount rate (d) between large and small rewards (which was fixed across reward types) was inversely correlated with nutrient values; this was expected in our models because when any of the nutrient value increases, values for other reward types decrease naturally because of value normalization (see above, Materials and Methods) and the fixed discount rate would decrease to adapt to such inadvertent influences from value normalization.

Thus, the choices of the monkeys for probabilistic rewards with defined nutrient compositions were well explained by a nutrient-sensitive RL model that assigned subjective values of specific fat and sugar components to reward outcomes and choice options.

Value updating based on nutrient-specific reward prediction errors

Our nutrient-sensitive RL model suggested that the monkeys tracked fat and sugar value components of probabilistic reward outcomes and integrated them into a scalar value that guided choices. To better understand the dynamics of nutrient-specific value updating, we examined an extended nutrient prediction error-based RL model (NutRPE). In this model, the reward value on trial t, $Q_i(t)$, was decomposed into fat and sugar value components (Fig. 7A; see above, Materials and Methods). These value components (1) separately adapted to changes in reward probabilities based on fat and sugar reward prediction errors (RPEs) from experienced nutrient outcomes and (2) were flexibly combined into the integrated reward value for each option, based on the nutrient preference of the animal (Fig. 7A). On each trial, the NutRPE model compared the expected fat and sugar value components with the obtained nutrient outcomes and computed nutrient-specific RPEs to update reward expectations. Thus, the model reveals the trial-by-trial dynamics of latent learning variables in dynamic nutrient-reward choices, including nutrient-specific values, integrated reward values, and nutrient-specific RPEs. We examined this model in part because it could serve as a tool in future studies to identify neuronal signals related to nutrient-specific learning and decision variables.

By capturing the dynamic updating and subjective integration of the monkeys of fat-sugar value components, the model closely tracked evolving choice patterns for nutritionally distinct rewards in single sessions (Fig. 7B). The NutRPE model provided a good fit to the choices of the monkeys across sessions according to psychometric curves and model-fit indicators (Fig. 7C; pseudo-R^2, monkey Ya = 0.82 ± 0.01; monkey Ym = 0.46 ± 0.02; percentage of correctly modeled choices, monkey Ya = 95.8 ± 0.8%; monkey Ym = 82.3 ± 1.0%). The trajectories of fat value and sugar value within single sessions characterized the idiosyncratic sensitivity of each monkey to fat and sugar when updating expected reward values (Fig. 7D). Specifically, reward values in monkey Ya were primarily driven by the sugar value component, which was updated after the monkey received high-sugar rewards and tracked their blockwise changes in reward probabilities, whereas the fat-value component varied much less within a session (Fig. 7D, left). In monkey Ym, both fat and sugar value components contributed to value updating and tracked fluctuating reward probabilities throughout the session (Fig. 7D, right).

The NutRPE model also allowed us to examine the dynamics of nutrient-specific RPEs that separately updated the fat and sugar value components. For instance, in the first block of the single-session data in Figure 7E, when the LFHS reward (Fig. 7B,E, blue data, trials 1–100) had a high reward probability, LFHS was frequently chosen and produced mostly positive sugar RPEs. The few negative sugar RPEs were because of the occasional low reward outcomes and fluctuations of values during learning. The magnitude of sugar RPEs decreased during this block while the monkey learned the reward value more accurately. After probability reversal, LFHS was chosen less frequently and mostly produced low-magnitude rewards. However, it produced a few large positive sugar RPEs during that period (Fig. 7E, blue data, trials 100–200), which illustrated that even low rewards can produce large positive sugar RPEs if the reward contained sugar, because reward outcomes were scaled by the nutrient-specific subjective value in the NutRPE model. By contrast, when the high-fat
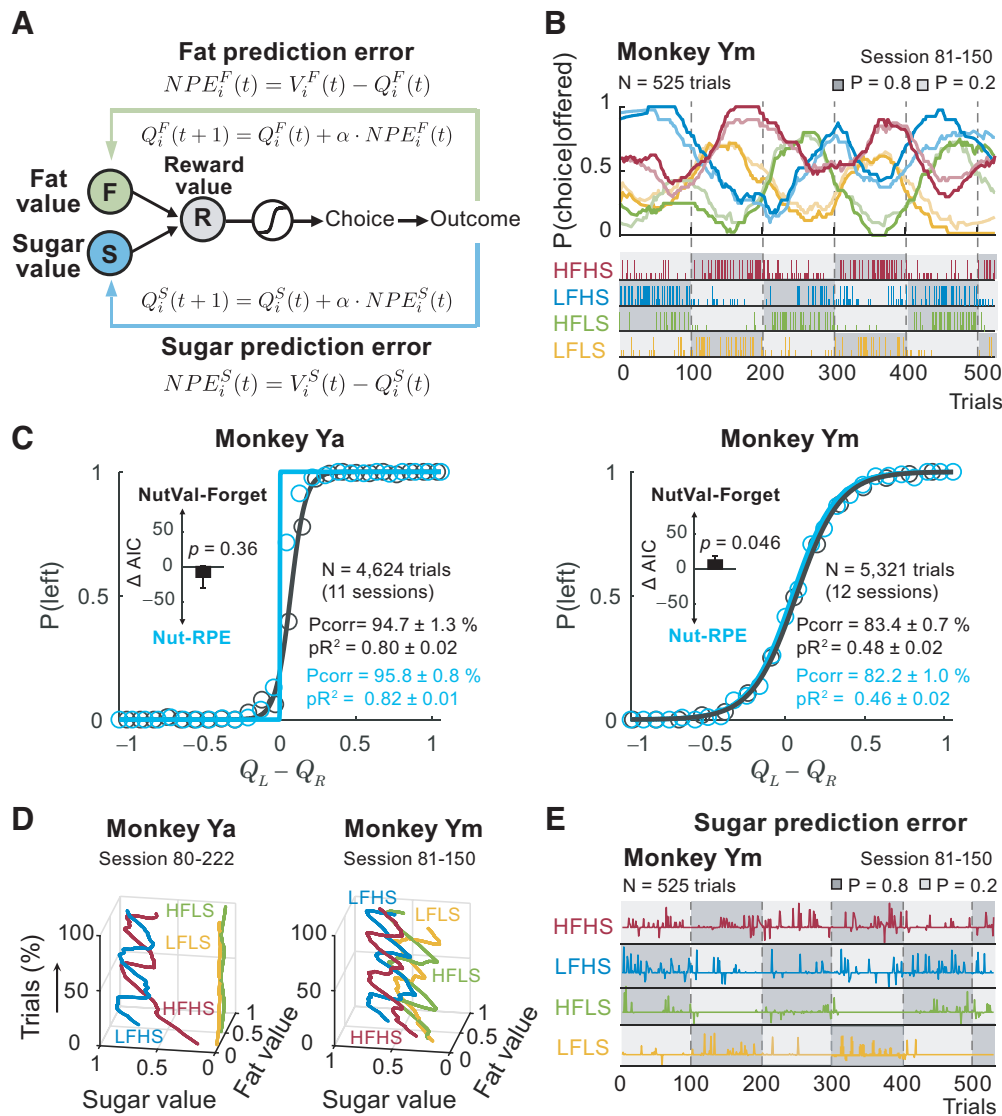
Figure 7. Learning with nutrient-specific reward prediction errors. **A**, NutRPE model. Subjective values for fat (top, green) and sugar (bottom, blue) were updated based on differences between reward outcomes and expected nutrient values (NPEs); fat values and sugar values were multiplied into integrated reward values to guide choices. **B**, Single-session data for choices, rewards, and modeled choice probabilities based on the NutRPE model (monkey Ym). Top, Model-predicted choices (faint lines) tracked the actual choices of monkey Ym (thick lines) for nutrient-defined rewards. Bottom, The trial-by-trial record of choices and reward outcomes. Tick marks represent choices for each reward, long marks indicate large-reward outcomes (rewarded trials); short marks indicate small-reward outcomes (nonrewarded trials). **C**, Psychometric curves based on value estimates from the NutRPE model accounted for the choices of the monkeys with good model-fit indicators. Pcorr, Percentage correctly modeled choices ± SEM pR²: pseudo-R² ± SEM. **D**, Single-session dynamics of nutrient values. Reward values in monkey Ya were dominated by the sugar values. In the NutRPE model, these sugar values were updated after choices of high-sugar liquids and tracked blockwise changes in reward probability. In monkey Ym, both fat values and sugar values contributed to value updating and tracked fluctuating reward probabilities. **E**, Sugar prediction errors in the single session shown in **B**. Prediction errors were sensitive to sugar content and reward size, determined by model-derived nutrient value parameters fitted to the choices of the monkey. Single-session data in **B**, **D**, **E** are the same sessions as in Fig. 1D (right).

low-sugar stimulus was associated with high reward probability in the third block (HFLS; Fig. 7B,E, green data), it consistently produced large rewards but only very small sugar RPEs because of the low sugar content of the rewards.

We note that both the NutRPE model and the NutVal-Forget model have eight free parameters. Although the NutRPE model does not have a free parameter for fat–sugar interaction (because it assumes multiplication of fat values and sugar values), it does have two separate learning rates for fat-value and sugar-value updating. A direct comparison between the NutRPE model and the NutVal-Forget model (Fig. 4F) showed that these two models performed equally well in monkey Ya, and the NutVal-Forget model outperformed the NutRPE model in monkey Ym. Therefore, the NutRPE model revealed trial-by-trial nutrient value dynamics without compromising model performance.

Thus, subjective nutrient-value functions guided the dynamic updating and integration of reward values based on nutrient-specific reward components and related nutrient-specific RPEs. The nutrient RPEs were sensitive to both the presence of a specific nutrient in the experienced reward outcome and the reward size. These results provide a basis for hypotheses about candidate neurons encoding nutrient-specific values and RPEs (Fig. 8), as explained next in Discussion.

Discussion

We investigated choices of monkeys for nutrient-defined rewards under varying reward probabilities and found that the nutrient composition of rewards strongly influenced choices and learning. As in our previous study involving the same animals (Huang et

al., 2021), the choices of the monkeys reflected individual fat and sugar preferences, consistent with the assignment of subjective values to nutrients. Importantly, in the present study we show that these nutrient preferences affected how the animals adapt their choices to changing reward probabilities during learning. Specifically, the monkeys were more likely to repeat choices that led to rewards containing preferred nutrients and frequently chose these options even under low reward probabilities. As in previous studies (Lau and Glimcher, 2005; Tsutsui et al., 2016), more recent rewards had a stronger influence on current choice. Critically, we found that this impact of reward history depended on the nutrient composition of the reward; past rewards that were high in preferred sugar content influenced subsequent choices more strongly than less preferred low-nutrient or high-fat rewards. Choice history regardless of reward outcomes also had a significant nutrient-dependent effect on choice, with stronger effects of past choices for preferred nutrients. We developed a nutrient-sensitive RL model that incorporated these influences of preferred nutrients on learning and choice. The model updated the values of sugar and fat components of expected rewards trial by trial, based on recently experienced rewards, and integrated these nutrient-specific values into scalar reward values that explained the choices of the monkeys. Value updating in response to particular nutrient outcomes was governed by nutrient-specific RPEs. Thus, different from canonical RL models that learn from binary reward outcomes (Sutton and Barto, 1998), our nutrient-sensitive RL model learned based on the nutrient components of the rewards, which influenced the subjective values of choice options and reward outcomes. Nutrient effects on learning and choice were not explained by energy content or other reward properties (flavor, salt, protein), which we controlled, or preferences for visual cues, which we changed every session. Our results suggest that nutrients constitute important reward components that influence subjective valuation, learning, and choice and that RL models can be usefully extended to incorporate nutrient-specific effects on behavior.

Previous studies in macaques revealed important factors influencing learning and choice, including reward and choice history (Corrado et al., 2005; Lau and Glimcher, 2005; Samejima et al., 2005; Kennerley et al., 2006; Lee et al., 2012; Seo et al., 2012; Tsutsui et al., 2016), the variance of recent rewards (Grabenhorst et al., 2019b), novelty and rarity of choice objects (Costa et al., 2019; Rothenhoefer et al., 2021), and social observations (Grabenhorst et al., 2019a). Because these studies did not vary the composition of reward outcomes, they could not test how specific nutrients affect learning and choice. Other studies demonstrated that macaques have sophisticated preferences for different reward types that comply with principles of economic choice theory (Padoa-Schioppa and Assad, 2006; Lak et al., 2014; Raghuraman and Padoa-Schioppa, 2014; Pastor-Bernier et al., 2017) but did not examine how nutrient rewards affect RL. Here, we reasoned that nutrients represent biologically critical, intrinsic reward components that are essential for survival and thus should affect learning and choice in dynamic reward environments to ensure stable nutrient intake. By systematically
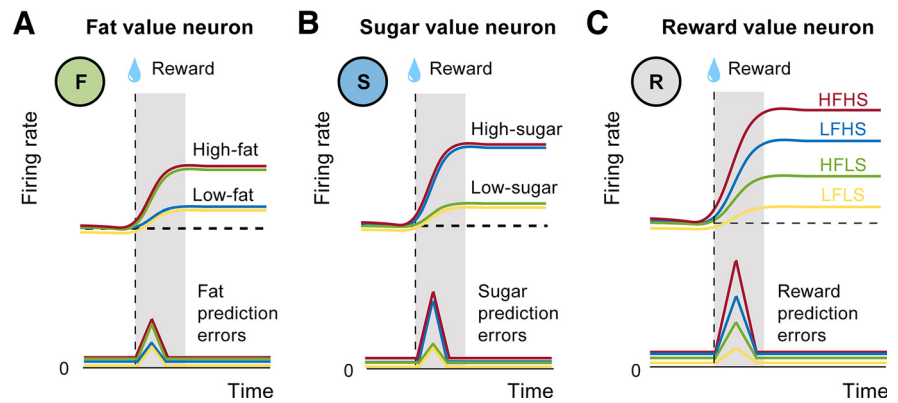

Figure 8. Hypothesized neuron types encoding nutrient-specific learning and decision variables. *A*, Fat value neurons signal the trial-by-trial fat-specific value component and update their activity based on fat-specific reward prediction errors. *B*, A similar process operates for sugar-value neurons. *C*, Inputs from fat and sugar value neurons may converge onto reward value neurons that signal integrated, scalar value, depending on the subjective nutrient preferences and integrated reward prediction errors to guide learning and choice.

manipulating the sugar and fat content of liquid rewards, we showed that monkeys have subjective preferences for specific nutrients and that these preferences influence how monkeys adapt their choices to changing reward probabilities. Similar to one previous study (Raghuraman and Padoa-Schioppa, 2014), our monkeys based their choices on both subjective reward preferences and reward probabilities. However, different from that study, our monkeys were required to learn reward probabilities from experience rather than from explicit pretrained cues; we also systematically varied the nutrient composition of rewards rather than varying reward type regardless of nutrient content. Thus, beyond the general effects of reward type, we determined that specific nutrients differentially affect learning and choice. We do not suggest that nutrients influence learning independently of subjective value. Rather, our data and model suggest that nutrients influence learning by affecting the subjective values that animals assigned to reward outcomes and choice options.

In our study, monkeys adapted their choices to changing reward probabilities to obtain large amounts of preferred nutrient-defined rewards; when probabilities changed, the monkeys could switch their choices to less-preferred but more available alternatives. The animals were not required to learn the nutrient composition of novel foods as the four rewards used were highly familiar to the animals. This situation resembles natural scenarios in which monkeys regularly forage for a range of familiar food rewards and must determine which foods are currently available (Cui et al., 2018, 2019, 2020). Wild macaques solve such foraging problems by adapting their choices to short-term changes in reward availability (e.g., switching between foraging grounds) and to longer-term (e.g., seasonal) changes. For example, preferred foods, such as fat- and protein-rich seeds may only be available in autumn and winter. Accordingly, wild macaques adapt their foraging patterns to pursue different foods in other seasons, such as leaves and herbs, while maintaining a stable nutrient balance (Cui et al., 2019). Our simple repeated-choice paradigm captured these features of the foraging environments of macaques in the laboratory. Our findings suggest that the computations underlying nutrient-adaptive foraging choices involve the assignment of subjective values to specific nutrients, the integration and updating of nutrient values from recently experienced reward outcomes, and the comparisons of values to make choices.

Flavor-nutrient conditioning is an important process in feeding and foraging behavior (de Araujo et al., 2020; Dayan, 2022);

however, it is unlikely to have played a major role in our study. The animals were highly familiar with the four different nutrient-defined rewards through task training over many months. Thus, flavor-nutrient conditioning effects would have stabilized by the time the present experiments were conducted.

We extended a canonical RL model to account for the observed nutrient sensitivity of the choices of the monkeys. Different from RL models that learn from binary reward outcomes, our nutrient-sensitive RL model evaluated rewards using an animal-specific nutrient-value function. This nutrient-value function detected the fat and sugar content of the reward in addition to reward size and weighted the reward according to the preferences of the individual monkey for fat, sugar, and their interaction. As a consequence, the model learned expected values of choice options based on the nutrient composition of the reward and nutrient preferences of the monkeys, rather than based on reward frequency regardless of nutrient content as in standard RL models. An extension of this model decomposed subjective values into separate fat and sugar value components and updated these value components with nutrient-specific prediction errors that considered the difference between experienced and expected values for fat and sugar components. Although not investigated here, our experimental paradigm and model could be extended to incorporate internal, physiological set points as proposed in recent homeostatic RL (Keramati and Gutkin, 2014). Our model could also separately model short-term influences of sensory, hedonic food components on learning and choice, including recently demonstrated fat-related oral-texture effects (Huang et al., 2021) and longer-term effects related to postingestive processing (Dayan, 2022).

Our findings have implications for understanding the neural mechanisms underlying RL. We described a nutrient-specific learning process that updates value estimates for separate fat and sugar reward components and integrates this information into scalar values that guide adaptive choices. At the neural level, this mechanism would require neurons that encode individual nutrient values and dynamically update them via nutrient-specific RPEs (Fig. 8). By decomposing trial-by-trial reward values that guide RL into nutrient-value components, our model specifies computational signals that could help identify such neurons (Fig. 7). Midbrain dopamine neurons, orbitofrontal cortex, and amygdala participate in value-based decision-making, RL, and food evaluation (Padoa-Schioppa and Assad, 2006; Grabenhorst et al., 2010, 2012; Murray and Rudebeck, 2013; Lak et al., 2014; Stauffer et al., 2014; Suzuki et al., 2017; Murray and Rudebeck, 2018; Rolls et al., 2018; Grabenhorst et al., 2019a; Averbeck and Murray, 2020) and thus constitute targets for testing these hypotheses. For example, it will be interesting to determine whether primate dopamine neurons encode RPEs for specific nutrients. Recent studies demonstrated that dopamine neurons encode RPEs in terms of economic utility, thereby integrating influences of reward type, quantity, probability, and risk (Lak et al., 2014; Stauffer et al., 2014). It remains to be tested whether the presently observed nutrient-specific effects on learning and choice are incorporated into the RPE signal of the dopamine neurons or whether nutrient-specific reward components are processed in other reward structures such as the orbitofrontal cortex and amygdala, where single neurons encode nutrients and sensory reward properties (Kadohisa et al., 2005; Huang and Grabenhorst, 2022).

In summary, our results identify nutrients as important reward components that influence subjective valuations, choices, and learning of monkeys. These processes were well described by a nutrient-sensitive RL model that updated the value of expected rewards based on their sugar and fat components using nutrient-specific RPEs. Our data and nutrient-sensitive RL model can serve as tools to guide future studies that aim to uncover nutrient-specific learning and decision computations and their neurophysiological implementations.

References

Averbeck BB, Murray EA (2020) Hypothalamic interactions with large-scale neural circuits underlying reinforcement learning and motivated behavior. Trends Neurosci 43:681–694.

Burnham KP, Anderson DR, Huyvaert KP (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav Ecol Sociobiol 65:23–35.

Carreiro AL, Dhillon J, Gordon S, Higgins KA, Jacobs AG, McArthur BM, Redan BW, Rivera RL, Schmidt LR, Mattes RD (2016) The macronutrients, appetite, and energy intake. Annu Rev Nutr 36:73–103.

Corrado GS, Sugrue LP, Seung HS, Newsome WT (2005) Linear-nonlinear-Poisson models of primate choice dynamics. J Exp Anal Behav 84:581–617.

Costa VD, Mitz AR, Averbeck BB (2019) Subcortical substrates of explore-exploit decisions in primates. Neuron 103:533–545.e5.

Cui Z, Shao Q, Grueter CC, Wang Z, Lu J, Raubenheimer D (2019) Dietary diversity of an ecological and macronutritional generalist primate in a harsh high-latitude habitat, the Taihangshan macaque (*Macaca mulatta tcheliensis*). Am J Primatol 81:e22965.

Cui ZW, Wang ZL, Shao Q, Raubenheimer D, Lu JQ (2018) Macronutrient signature of dietary generalism in an ecologically diverse primate in the wild. Behav Ecol 29:804–813.

Cui ZW, Wang ZL, Zhang SQ, Wang BS, Lu JQ, Raubenheimer D (2020) Living near the limits: effects of interannual variation in food availability on diet and reproduction in a temperate primate, the Taihangshan macaque (*Macaca mulatta tcheliensis*). Am J Primatol 82.

Dayan P (2022) "Liking" as an early and editable draft of long-run affective value. PLoS Biol 20:e3001476.

de Araujo IE, Schatzker M, Small DM (2020) Rethinking food reward. Annu Rev Psychol 71:139–164.

Gallistel CR, Fairhurst S, Balsam P (2004) The learning curve: implications of a quantitative analysis. Proc Natl Acad Sci U S A 101:13124–13131.

Grabenhorst F, Rolls ET, Parris BA, D'Souza A (2010) How the brain represents the reward value of fat in the mouth. Cereb Cortex 20:1082–1091.

Grabenhorst F, Hernádi I, Schultz W (2012) Prediction of economic choice by primate amygdala neurons. Proc Natl Acad Sci U S A 109:18950–18955.

Grabenhorst F, Baez-Mendoza R, Genest W, Deco G, Schultz W (2019a) Primate amygdala neurons simulate decision processes of social partners. Cell 177:986–998.e15.

Grabenhorst F, Tsutsui KI, Kobayashi S, Schultz W (2019b) Primate prefrontal neurons signal economic risk derived from the statistics of recent reward experience. Elife 8: e44838.

Huang F-Y, Grabenhorst F (2022) Nutrient and sensory coding of anticipated food reward in primate amygdala neurons. Society for Neuroscience Abstracts 318. Available at https://www.abstractsonline.com/pp8/#!/10619/presentation/80824.

Huang F-Y, Sutcliffe MPF, Grabenhorst F (2021) Preferences for nutrients and sensory food qualities identify biological sources of economic values in monkeys. Proc National Acad Sci U S A 118:e2101954118.

Kadohisa M, Rolls ET, Verhagen JV (2005) Neuronal representations of stimuli in the mouth: the primate insular taste cortex, orbitofrontal cortex, and amygdala. Chem Senses 30:401–419.

Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9:940–947.

Keramati M, Gutkin B (2014) Homeostatic reinforcement learning for integrating reward collection and physiological stability. Elife 3:e04811.

Lak A, Stauffer WR, Schultz W (2014) Dopamine prediction error responses integrate subjective value from different reward dimensions. Proc Natl Acad Sci U S A 111:2343–2348.

Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav 84:555–579.

Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451–463.

Lee D, Seo H, Jung MW (2012) Neural basis of reinforcement learning and decision making. Annu Rev Neurosci 35:287–308.

Ma CY, Liao JC, Fan PF (2017) Food selection in relation to nutritional chemistry of Cao Vit gibbons in Jingxi, China. Primates 58:63–74.

McNamara JM, Houston AI (1997) Currencies for foraging based on energetic gain. Am Nat 150:603–617.

Murray EA, Rudebeck PH (2013) The drive to strive: goal generation based on current needs. Front Neurosci 7:112.

Murray EA, Rudebeck PH (2018) Specializations for reward-guided decision-making in the primate ventral prefrontal cortex. Nat Rev Neurosci 19:404–417.

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223–226.

Pastor-Bernier A, Plott CR, Schultz W (2017) Monkeys choose as if maximizing utility compatible with basic principles of revealed preference theory. Proc Natl Acad Sci U S A 114:E1766–E1775.

Raghuraman AP, Padoa-Schioppa C (2014) Integration of multiple determinants in the neuronal computation of economic values. J Neurosci 34:11583–11603.

Rangel A (2013) Regulation of dietary choice by the decision-making circuitry. Nat Neurosci 16:1717–1724.

Rescorla RA, Wagner AR (1972) A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical conditioning ii: current research and theory (Black AH. Prokasy WF, eds), pp 64–99. New York: Appleton Century Crofts.

Rolls ET (2020) The texture and taste of food in the brain. J Texture Stud 51:23–44.

Rolls ET, Mills T, Norton AB, Lazidis A, Norton IT (2018) The neuronal encoding of oral fat by the coefficient of sliding friction in the cerebral cortex and amygdala. Cereb Cortex 28:4080–4089.

Rothenhoefer KM, Hong T, Alikaya A, Stauffer WR (2021) Rare rewards amplify dopamine responses. Nat Neurosci 24:465–469.

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599.

Seo M, Lee E, Averbeck BB (2012) Action selection and action value in frontal-striatal circuits. Neuron 74:947–960.

Simpson SJ, Raubenheimer D (2012) The nature of nutrition: a unifying framework from animal adaptations to human obesity. Princeton: Princeton UP.

Simpson SJ, Raubenheimer D (2020) The power of protein. Am J Clin Nutr 112:6–7.

So NY, Stuphorn V (2010) Supplementary eye field encodes option and action value for saccades with variable reward. J Neurophysiol 104:2634–2653.

Stauffer WR, Lak A, Schultz W (2014) Dopamine reward prediction error responses reflect marginal utility. Curr Biol 24:2491–500.

Stuphorn V (2021) Food reward derives from nutrient content and sensory qualities. Proc Natl Acad Sci U S A 118:e2109735118.

Sutton RS, Barto AG (1998) Reinforcement learning. Cambridge, MA: MIT Press.

Suzuki S, Cross L, O'Doherty JP (2017) Elucidating the underlying components of food valuation in the human orbitofrontal cortex. Nat Neurosci 20:1780–1786.

Takahashi MQ, Rothman JM, Raubenheimer D, Cords M (2019) Dietary generalists and nutritional specialists: feeding strategies of adult female blue monkeys (*Cercopithecus mitis*) in the Kakamega Forest, Kenya. Am J Primatol 81:e23016.

Tsutsui K, Grabenhorst F, Kobayashi S, Schultz W (2016) A dynamic code for economic object valuation in prefrontal cortex neurons. Nat Commun 7:12554.

van der Klaauw AA, Keogh JM, Henning E, Stephenson C, Kelway S, Trowse VM, Subramanian N, O'Rahilly S, Fletcher PC, Farooqi IS (2016) Divergent effects of central melanocortin signalling on fat and sucrose preference in humans. Nat Commun 7:13055.

Wilson RC, Collins AG (2019) Ten simple rules for the computational modeling of behavioral data. Elife 8:e49547.

Wittmann MK, Fouragnan E, Folloni D, Klein-Flügge MC, Chau BKH, Khamassi M, Rushworth MFS (2020) Global reward state affects learning and activity in raphe nucleus and anterior insula in monkeys. Nat Commun 11:3771.