Appetite 208 (2025) 107950

Contents lists available at ScienceDirect

Appetite

journal homepage: www.elsevier.com/locate/appet

# Shaping food choices with actions and inactions with and without reward and punishment

Huaiyu Liu [a,b,*], Julian Quandt [a], Lei Zhang [c,d,e], Xiongbing Kang [f], Jens Blechert [g], Tjits van Lent [a], Rob W. Holland [a], Harm Veling [a,h]

[a] Behavioral Science Institute, Radboud University Nijmegen, the Netherlands
[b] Department of Imaging Neuroscience, Functional Imaging Laboratory, 12 Queen Square, University College London, United Kingdom
[c] Centre for Human Brain Health, School of Psychology, University of Birmingham, United Kingdom
[d] Institute for Mental Health, School of Psychology, University of Birmingham, United Kingdom
[e] Centre for Developmental Science, School of Psychology, University of Birmingham, United Kingdom
[f] Genome Data Science, Faculty of Technology, Bielefeld University, Germany
[g] Department of Psychology, University of Salzburg, Salzburg, Austria
[h] Consumption and Healthy Lifestyles, Wageningen University and Research, Wageningen, the Netherlands

A R T I C L E I N F O

*Keywords:*
Reinforcement learning
Go/no-go training
Food choice
Bayesian statistics
Preregistration

A B S T R A C T

Enabling people to reduce their consumption of unhealthy appetitive products can improve their health. Over the last decades, progress has been made by uncovering new ways to change behavior toward appetitive products without feedback incentives (e.g., reward or punishment, as in feedback-driven reinforcement learning), but instead by cueing motor responses (e.g., go vs. no go) toward these products in cognitive training tasks. However, it is unclear how this nonreinforced learning compares to reinforcement learning. Moreover, recent work on reinforcement learning has uncovered a basic learning mechanism, the action–valence asymmetry, which points to the possibility that reward and punishment learning may not always outperform learning without any external reinforcement. Here, we report two well-powered preregistered experiments (experiment 1a: N = 72; experiment 1b: N = 81) that examined when reinforcement learning outperforms nonreinforced learning in modifying people's preferences for food. Our findings show that reinforcement learning notably surpasses nonreinforced learning, but only when active responses (go) are rewarded, and inactions (no-go) are reinforced by avoiding punishments. These results shed light on interventions that combine rewards and punishments to facilitate changes in food preferences.

## 1. Introduction

Incremental daily choices can lead to substantial life outcomes. For example, everyday dietary choices are likely to influence long-term physical health (Estacio, 2018). Unfortunately, changing such choices is difficult, as established behavioral patterns are often elicited by deeply learned habitual or hedonic responses to the environment, which may interfere with intentions to change behavior (Estacio, 2018; Marteau et al., 2012). Thus, promoting new behavioral patterns toward appetitive products such as food items requires understanding of how learning mechanisms can be employed to change choices. In the present work, we aim to compare the effects of different learning mechanisms on food choices, and examine how reinforcing go and no-go responses versus not reinforcing go and no-go responses during training influences subsequent food preferences.

One prominent form of learning is reinforcement learning, where different courses of action lead to either reward or punishment, and individuals learn by trial-and-error to find out how to maximize rewards and minimize punishments. Such reinforced responses become preferred and are more likely to be repeated (Sutton & Barto, 2018). There is evidence that food choices can be changed through reinforcement learning. For example, in one study, adult participants performed a binary food choice task in which they chose between appetitive junk food items that had high versus low subjective value as assessed by a pretest calibration procedure. This choice task was preceded by a reinforcement learning training task in which choices for only low value food items

* Corresponding author. Behavioral Science Institute, Radboud University Nijmegen, the Netherlands.
*E-mail address:* huaiyu.liu@ucl.ac.uk (H. Liu).

https://doi.org/10.1016/j.appet.2025.107950
Received 24 October 2024; Received in revised form 29 January 2025; Accepted 27 February 2025
Available online 28 February 2025

were rewarded. In the testing phase, this reinforcement learning increased the probability of choosing low versus high value items in choice pairs where choices for low value food items had been reinforced compared to control choice pairs (Schonberg et al., 2014).

Interestingly, recent studies show that nonreinforced learning tasks in which people respond (or not respond) to food items without reward and punishment can also influence subsequent food choices. These tasks have two key features: responses are cued (e.g., an auditory tone signaling the correct response) instead of reinforced through trial and error, and the absence of external reinforcers after responding (Chen et al., 2019; Schonberg et al., 2014). Over the past two decades, three nonreinforced learning tasks in particular, cue-approach training (CAT) (Bakkour et al., 2016; Schonberg et al., 2014), go/no-go training (GNG) (Chen et al., 2016; Houben & Jansen, 2015; Lawrence et al., 2015; Liu et al., 2022), approach/avoidance training (AAT) (van Alebeek et al., 2023; Veling et al., 2021) have been used to alter food preferences. In these tasks, participants respond to food items labeled as go or no-go based on specific cues (Quandt et al., 2019; Veling et al., 2022), or emit approach or avoidance responses to food items (Becker et al., 2015; van Alebeek et al., 2023). CAT differs from GNG in having fewer go trials and adjusting cue timing based on performance to increase task difficulty (Schonberg et al., 2014), and AAT differs from GNG in training different behavioral responses, i.e., approach/avoidance responses rather than mere go/no-go responses (Veling et al., 2021).

Despite slight differences between the training procedures (for a discussion see Veling et al., 2022), CAT, GNG, and AAT consistently shift preferences toward go (approach) food items, with an approximately 55%–60% choice rate for go (approach) food items over no-go (avoidance) items (Table S1 in supplemental materials). This high consistency in effects across studies and tasks elicited by nonreinforced training caught our attention. It made us curious about the extent to which training effects would be stronger with application of reinforcement learning. While intuitively it seems reasonable to assume reinforced learning is in general more effective than nonreinforced learning to change preferences (see also Van Dessel et al., 2018 for work on adding evaluative consequences to a nonreinforced approach/avoidance training), here we propose reinforcement learning may not always work better than nonreinforced learning to boost food preference change. This counter-intuitive logic might be understood by considering how the so-called Pavlovian bias (Guitart-Masip et al., 2012) during learning can impact subsequent food preferences.

Specifically, previous work showed that learning of go/no-go responses to fractal items is asymmetrical under reward and punishment conditions, respectively. Go responses are easier to learn when they are instrumental to obtain a reward (GoToWin) compared to when they are instrumental to avoid punishment (GoToAvoidPunishment), and no-go responses are easier to learn when they are instrumental to avoid punishment (NoGoToAvoidPunishment) compared to when they are instrumental to obtain a reward (NoGoToWin). This so-called action–valence asymmetry is attributed to a hard-wired Pavlovian bias (Guitart-Masip et al., 2014). It is currently unclear 1) how such valence asymmetries during learning about food items translates into preferences for these food items, and 2) how this compares to nonreinforced learning. We conducted two preregistered experiments to provide preliminary investigations on these questions.

Experiment 1a employed a reinforcement learning GNG task—RL GNG—modified from previous work (Cavanagh et al., 2013; Guitart-Masip et al., 2012). This task comprised four within-subject conditions (GoToWin, GoToAvoidPunishment, NoGoToWin, and NoGoToAvoidPunishment), with participants providing either a go or no-go response to food items in each condition. Participant's task was to learn which response was normatively the best response (e.g., go in the case of a GoToWin food item). In the GoToWin and GoToAvoidPunishment conditions, go responses were reinforced by rewards and by the avoidance of punishments, respectively, while no-go responses resulted in neutral outcomes in both conditions; in the NoGoToWin and NoGoToAvoidPunishment conditions, no-go responses were reinforced by reward and by the avoidance of punishment, respectively, while go responses resulted in neutral outcomes in both conditions.

Immediately after the RL GNG, participants were asked to make binary choices between go and no-go food items from the different conditions, each linked to a specific learning condition (Table 1). We assessed food preferences based on how often participants chose go over no-go items (e.g., if the choice rate of go items is credibly larger than 50%, it indicates that participants favor go items). Given that previous work has shown participants find it much easier to learn the GoToWin and NoGoToAvoidPunishment conditions on the one hand compared to the GoToAvoidPunishment and NoGoToWin conditions on the other hand, we hypothesized that these conditions would differentially influence food preferences for go over no-go food items. Specifically, when after the learning participants make binary choices between go and no-go items, we expected the preference for GoToWin items to be higher in GoToWin_vs_NoGoToAvoidPunishment food choice pairs than in all other food pairs involving a go and no-go item, as the other food pairs involve at least one suboptimal learning condition (i.e., GoToWin versus NoGoToWin; GoToAvoidPunishment versus NoGoToAvoidPunishment; GoToAvoidPunishment versus NoGoToWin). For instance, the probability to choose a go item should be higher in the GoToWin_vs_ NoGoToAvoidPunishment comparison than in the GoToWin_vs_NoGoToWin comparison, because previous research suggests that NoGoToWin is a suboptimal learning condition.

Experiment 1b employed a nonreinforced GNG without feedback. This nonreinforced GNG included two Go and two NoGo conditions as well, but no reward and punishment reinforcers were provided for any go/no-go responses. In the two Go conditions, go responses were cued by two different visual go cues; in the two NoGo conditions, no-go responses were cued by using two other different visual no-go cues. Immediately after the nonreinforced GNG, participants completed a similar binary choice test as explained above. In this test, participants made binary choices between go and no-go food items that had been used as stimuli in the Go and NoGo conditions. For the preference test in Experiment 1b, we hypothesized that participants would prefer go items over no-go items. This hypothesis aligns with findings from previous work described above (Chen et al., 2019).

For both experiments, we preregistered hypotheses, power-analysis, and strategy of data-analysis on Open Science Framework (Experiment 1a: https://osf.io/gnj52; Experiment 1b: https://osf.io/yzjtd). Experimental materials, data, and analyses scripts are also publicly available at (Experiment 1a: https://osf.io/7hrka/; Experiment 1b: https://osf.io/4qpxn/).

We pre-registered the following hypotheses tests: 1) For the RL GNG, a higher *learning* accuracy should be found in the GoToWin (versus GoToAvoidPunishment) and NoGoToAvoidPunishment (versus NoGoToWin) conditions, respectively. 2) For the preference test in Experiment 1a, we expected stronger preference for GoToWin items in the GoToWin_vs_NoGoToAvoidPunishment comparison than in the other go versus no-go comparisons. 3) For the preference test in Experiment 1b,

**Table 1**
Experimental choice pairs in the food choice task of Experiments 1a and 1b.

| Experiment 1a | Experiment 1b |
|---|---|
| GoToWin_vs_NoGoToWin | GoBlue_vs_NoGoOrange |
| GoToWin_vs_NoGoToAvoidPunishment | GoBlue_vs_NoGoPink |
| GoToAvoidPunishment_vs_NoGoToWin | GoTeal_vs_NoGoOrange |
| GoToAvoidPunishment_vs_NoGoToAvoidPunishment | GoTeal_vs_NoGoPink |
| GoToWin_vs_GoToAvoidPunishment | GoTeal_vs_GoBlue |
| NoGoToWin_vs_NoGoToAovidPunishment | NoGoOrange_vs_NoGoPink |

Note: In the column Experiment 1b, the words Blue, Orange, Teal, Pink refer to the colors of frames that surround go or no-go food pictures. It should be noted that such frame colors are arbitrary and have no theoretical meaning, unlike those in Experiment 1a. In other words, the go/no-go choice pairs in Experiment 1b are equivalent to one another.

we expected an overall preference for go over no-go items.

## 2. Method

### 2.1. Transparency and openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and we follow Journal article reporting standards (Appelbaum et al., 2018). All data, analysis code, and research materials are available at (Experiment 1a: https://osf.io/7hrka/; Experiment 1b: https://osf.io/4qpxn/). Data were analyzed using R, version 4.2.1 (R Core Team, 2022) and the R package ggplot2, version 3.3.6 (Wickham, 2016) and the R package brms, version 2.17.0 (Bürkner, 2017).

### 2.2. Sample size and participants

We conducted a Bayesian power analysis to calculate the sample size based on our two hypotheses for the learning phase of reinforcement learning GNG. Simply put, we conducted 100 simulations, each with a hypothesized sample size. We used prior parameter estimates that extracted from previous work to generate data for each simulation. Subsequently, we applied our pre-registered statistical models to each dataset. If at least 80% of the simulations supported our hypotheses—aligning with Frequentist criteria—we considered this sample size justified. See our preregistration for details. Simulations revealed that a sample size of 72 participants ensured a power of 80% to reach credible differences.

We recruited 72 (19 males, 53 females, $M_{age} = 24.17$ years, $SD_{age} = 3.73$) participants in Experiment 1a. Experiment 1b had the same planned sample size as Experiment 1a, but the actual sample size was 81 (55 males, 23 females, 3 non-binary, $M_{age} = 23.07$ years, $SD_{age} = 2.61$) due to multiple online sign-ups for the final timeslot. All participants provided informed consent and the ethical committee at the Faculty of Social Science at the Radboud University approved the study (ECSW-2019-003).

### 2.3. Experimental procedure overview

Both experiments included three identical sequential tasks (food rating task, reinforced/nonreinforced GNG, food choice task), except for the structure of the GNG (Figs. 1 and 2). We programmed all three tasks in JavaScript using JsPsych (de Leeuw, 2015) and ran both experiments online via the platform Pavlovia (Bridges et al., 2020) due to COVID pandemic.

#### 2.3.1. Food rating task

Participants first performed a rating task in which they evaluated sixty food items that were selected from the food-pic database (Blechert et al., 2019). Each trial started with a specific food picture presented at the center of the screen. Participants were asked to rate the food item on a 200-point scale (0 = Not appealing at all, 200 = Very appealing).

After the rating task, the experimental program ranked all sixty food items from the highest value to the lowest value of their perceived attractiveness, and then selected four items with rank 29 to 32 for each participant, so that for each participant the experimental items were comparable in value and there was room to increase and decrease in value. Next, these four items were allocated to four conditions that would be used in the RL/nonreinforced GNG (panel B in Fig. 1 and panel C in Fig. 2) in a counterbalanced manner to match the average item rank among the four conditions across participants. In addition to these four experimental items, the program also selected 12 filler items to create pairs differing in value to validate that participants choose items with the highest subjective value on these filler trials. See Fig. S1 (supplemental materials) for details of this item selection and allocation procedure.

#### 2.3.2. Reinforcement learning GNG

Next, participants in Experiment 1a performed the reinforcement learning GNG where four conditions were included and each condition contained one specific food item. Each trial started with presenting one of four food items (for 1500 ms), and participants had only two action choices—go (press the 'F' key) and no-go (do *not* press any key)—toward the food items. After participants performed the action choice, they received one of three outcomes displayed for 2000 ms: a reward ('WIN 1 point'), a punishment ('LOSE 1 point') or an neutral outcome ('WIN 0 point'/'LOSE 0 point'). Outcomes depended on the four conditions and were probabilistic. In the GoToWin condition performing go responses resulted in a reward ('WIN 1 point') 80% of the time and a neutral outcome ('WIN 0 point') 20% of the time, while performing no-go responses resulted in a neutral outcome ('WIN 0 point') 80% of the time and a reward ('WIN 1 point') 20% of the time. In the GoToAvoidPunishment condition performing go responses resulted in a neutral outcome ('LOSE 0 point') 80% of the time and a punishment ('LOSE 1 point') 20% of the time, while performing no-go responses resulted in a neutral outcome ('LOSE 0 point') 20% of the time and a punishment ('LOSE 1 point') 80% of the time. In the NoGoToWin condition performing no-go responses resulted in a in a reward ('WIN 1 point') 80% of the time and a neutral outcome ('WIN 0 point') 20% of the time, while performing go responses resulted in a neutral outcome ('WIN 0 point') 80% of the time and a reward ('WIN 1 point') 20% of the time. In the NoGoToAvoidPunishment condition, performing no-go responses resulted in a neutral outcome ('LOSE 0 point') 80% of the time and a

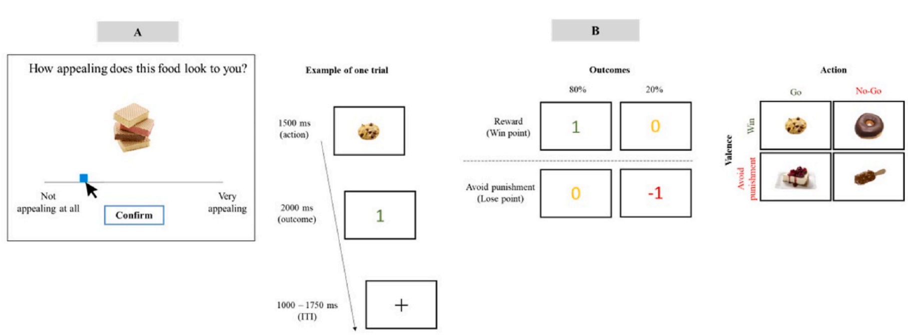


**Fig. 1.** Experimental procedure. (Panel A) Food rating task (Experiments 1a and 1b). Participants rated 60 food items on their perceived attractiveness. (Panel B) Reinforcement learning go/no-go task (Experiment 1a). Participants learned to choose the go or no-go action toward a specific food item via feedback-driven learning. Specifically, participants' go or no-go actions were reinforced through obtaining rewards for responding or not responding to specific foods (GoToWin and NoGoToWin conditions, respectively), or through avoiding punishments by responding or not responding to specific food items (respectively GoToAvoidPunishment and NoGoToAvoidPunishment conditions). See text for details.

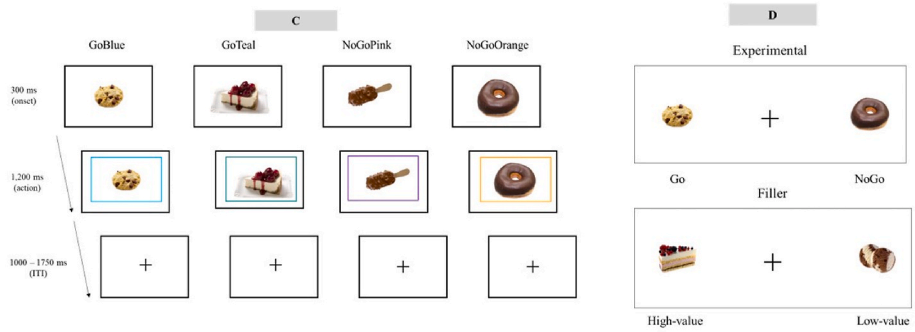


**Fig. 2.** Experimental procedure. (Panel C) The nonreinforced go/no-go task (Experiment 1b). Participants were trained both go and no-go actions toward specific food items without feedback. During go trials two different go cues were consistently paired with two different food items, and participants were instructed to perform go actions when go cues were presented. During no-go trials two no-go cues were consistently paired with two different food items, and participants were instructed to withhold responding when no-go cues were presented. (Panel D) Food choice task (Experiments 1a and 1b). Participants were asked to make binary food choices and choose the food items they would prefer to eat at the moment. Experimental pairs contained food items from the training which were matched on subjective value as assessed with the food rating task; in filler pairs one item had higher value than the other within each pair. See text for details.

punishment ('LOSE 1 point') 20% of the time, while performing go responses resulted in a neutral outcome ('LOSE 0 point') 20% of the time and a punishment ('LOSE 1 point') 80% of the time. In short, the probability of the correspondingly assigned outcome was 80%, and the neutral outcome ("zero points") was received otherwise (20% probability). Finally, each trial ended with an inter-trial-interval (ITI) randomly varied from 1000 ms to 1750 ms, in steps of 150 ms.

This reinforcement learning GNG included four blocks and all four conditions were intermixed in each block. There were 15 trials per condition in each block, resulted in 60 trials per block and 240 trials in total.

Before starting the task, participants were instructed that each condition included two action choices, and that one action choice would lead to better outcomes (i.e., more points which would at the end be converted to money) than the other action choice. Thus, participants were asked to learn from responding and observing the outcomes to find out which was the best action choice for each food item. Participants were instructed about the probabilistic nature of the task as well.

#### 2.3.3. Nonreinforced GNG

In the nonreinforced GNG (Experiment 1b), participants were trained on both go and no-go actions toward specific food items without reinforcement (i.e., outcomes), but the responses were cued. There were two trial types: (1) go trials where a visual go cue was consistently paired with a specific food item and participants were instructed to respond to such cues (i.e., 100% contingency), and (2) no-go trials in which a visual no-go cue was consistently paired with another food item, and participants needed to withhold responses when no-go cues were presented. Note that to mimic the reinforcement learning, this nonreinforced GNG that had also four conditions: We included two Go conditions using two different go cues, and two NoGo conditions with two different no-go cues.

Each trial stared with presenting one specific food item for 300 ms. Next, a frame—the cue—appeared and surrounded this picture for 1200 ms (irrespective of giving a response or not), and participants chose the go or no-go action based on the frame color. In one Go condition the cue was a teal frame, and that for the other Go condition was a blue frame. For one NoGo condition the cue was an orange frame, and for the other NoGo condition was a pink frame. Finally, each trial ended with an inter-trial-interval (ITI) randomly varied from 1000 ms to 1750 ms, in steps of 150 ms. Both block and trial structures were identical to the reinforcement learning GNG. Note that we paired each food item with each cue with a 100% contingency as previous work suggests that effects of nonreinforced training with GNG weaken when this contingency is lower (Jones et al., 2016).

#### 2.3.4. Food choice task

Finally, participants performed a food choice task. Each trial started with presenting two food items simultaneously (for 1500 ms), and participants were requested to choose the food item they would prefer to eat at the moment (within this 1500 ms) by pressing either the left arrow key (choosing the left item) or the right arrow key (choosing the right item). After making a choice, the chosen item was surrounded by a green frame for 2000 ms as confirmation. This food choice task included eight blocks, each block included 12 choice pairs that were presented in random order to participants, resulting in 96 choice pairs in total. We counterbalanced the left-right position within each choice pair across the blocks.

Each block included two types of choice pairs: Six experimental pairs and six filler pairs. These six experimental pairs (Table 1) were constructed of all possible combinations between the four food items from the preceding reinforced/nonreinforced GNG. Note that the important pairs are those between the go and no-go items (four pairs), whereas the remaining two pairs (one between two go items and the other between two no-go items) were not subjected to analysis as they did not directly contribute to our primary research questions.

For the six filler pairs, each pair was composed of one high-value and low-value item, and there was an incremental within-pairs value difference across all filler pairs. These filler pairs served as validation for the rating and choice tasks, because participants should prefer high-value (vs. low-value) items, and the difference in preference should be larger when the ranking difference is larger. See Supplemental Table S2 for details.

### 2.4. Confirmatory analyses

We ran Bayesian multilevel logistic regression models for all confirmatory analyses in Experiments 1a and 1b. We fitted all models using the package brms (Bürkner, 2017) in the statistical software R (R Core Team, 2022). For parameters of interest, we report the posterior mean estimate and its 95% highest density interval (HDI). If the 95% HDI does not include zero, we deem a "significant" effect in the Bayesian fashion (e.g., difference between two conditions) credible (Kruschke & Liddell, 2018).

#### 2.4.1. Reinforcement learning GNG

We compared the probability of correct responses between the GoToWin and GoToAvoidPunishment conditions, and that between the NoGoToWin and NoGoToAvoidPunishment conditions.

#### 2.4.2. Food choice

In Experiment 1a, we compared the probability of choosing go items under the comparison GoToWin_vs_NoGoToAvoidPunishment to the remaining three comparisons (i.e., GoToWin_vs_NoGoToWin, GoToAvoidPunishment_vs_NoGoToWin, and GoToAvoidPunishment_vs_NoGoToAvoidPunishment), respectively.

In Experiment 1b, we first collapsed data across all four go vs. no-go choice pairs (i.e., GoBlue_vs_NoGoOrange, GoTeal_vs_NoGoOrange, GoBlue_vs_NoGoPink, and GoTeal_vs_NoGoPink), and then obtained the overall probability of choosing go items and examined whether this probability was higher than 50%.

### 2.5. Exploratory analyses

#### 2.5.1. Transfer from reinforcement learning GNG to food choice

We explored how learning during RL GNG translated into subsequent food choice. The rationale is that, based on previous work (Swart et al., 2018), participants may prefer food items liked to larger go action weight, or put simply stronger strength of go action. To investigate this, we first estimated the strength of go and no-go action using computational models (see details below). Next, for each go versus no-go comparison, we calculated a score showing the difference in strength of go actions (e.g., between conditions like GoToWin and NoGoToAvoidPunishment) and measured how often participants chose the go item. Finally, we checked if the difference in strength of go action was correlated to how often participants chose the go items for each type of choice comparison.

We ran two Bayesian hierarchical reinforcement learning models (M1 and M2), based on previous work (Ahn et al., 2017; Guitart-Masip et al., 2012), to explore the reinforcement learning GNG data. In brief, reinforcement learning models rely on the assumption that on each trial, individuals have different expected values for different actions—in our case the go and no-go actions—and that these expected action values will be updated by incorporating information from actual reinforcers (i.e., outcomes) that followed the selected action. Accordingly, M1 assigned each action $a_t$ on trial $t$ an expected action value, and updated such expected value using a simple Rescorla-Wagner equation (Sutton & Barto, 2018):

$$Q_{t+1}(a_t, s_t) = Q_t(a_t, s_t) + \varepsilon(r_t - Q_t(a_t, s_t)) \quad (1)$$

Where the subscripts $t$ and $t+1$ referred to the current trial and the next trial, the $Q$ value represented the expected action value, $s$ referred to the item (e.g., GoToWin item), $\varepsilon$ ($0 < \varepsilon < 1$) is a free parameter learning rate that determined how much of the difference between the $Q$ value (on current trial) and the actual reinforcements should be incorporated in updating the $Q$ value (on the next trial), and $r$ represents the reinforcement. Note that $r_t \in \{-1, 0, 1\}$, in which $-1$ represents punishment, 0 represents neutral outcome, and 1 represents reward.

Next, M1 further calculated the probability of choosing the go or no-go action based on a softmax function:

$$p_{Go} = \frac{1}{1 + e^{-\tau^*(Q_{Go} - Q_{NoGo})}} \quad (2)$$

where $\tau$ was another free parameter that captures random noise during decision-making (Zhang et al., 2020). Note that in M1 includes the update of action values only, meaning that M1 does not consider the update of item values.

However, previous work (Guitart-Masip et al., 2012) showed that the updating of action values was influenced by the so-called Pavlovian bias. According to this Pavlovian bias, appetitive items automatically trigger go tendencies, whereas aversive items automatically trigger no-go tendencies. Therefore, the direction of action value updates is opposite to the direction of item value updates under the GoToAvoidPunishment and NoGoToWin conditions, resulting in impaired learning of go and no-go actions in these two conditions.

To explore the influence of Pavlovian bias on action learning for food items, based on M1, M2 added a Pavlovian bias parameter $\pi$ ($\geq 0$). that scales such influence of item values. This Pavlovian parameter increased the go tendencies when the item values were positive but decreased such tendencies when the item values were negative. Note that M2 assums that participants updated the item values following the same preceding Rescorla-Wagner rule as updating the $Q$ values in M1 (equation (4)). In addition, since M2 incorporates influences of both action and item values, we use the action weight (on each trial) $W$ to denote such incorporation and correspondingly, participants calculated the probability of choosing go action by comparing the differences between the go and no-go action weights.

$$W_t(a, s) = \begin{cases} Q_t(a, s) + \pi V_t(s) & \text{if } a = Go \\ Q_t(a, s) & \text{else} \end{cases} \quad (3)$$

$$V_{t+1}(s) = V_t(s) + \varepsilon(r_t - V_t(s)) \quad (4)$$

$$p_{Go} = \frac{1}{1 + e^{-\tau^*(W_{Go} - W_{NoGo})}} \quad (5)$$

#### 2.5.2. Food choice

First, we combined the choice data from Experiments 1a and 1b, and included experiment (1a vs. 1b) as a between-subject factor. We then fitted these choice data to an overall Bayesian multilevel logistic regression model. Next, we compared the probabilities of choosing go versus no-go items between Experiment 1b and the relevant conditions in Experiment 1a.

Second, we also explored other types of choice pairs such as the filler choice pairs in both experiments. We elaborate on the details of these exploratory analyses in S1 Text.

## 3. Results

### 3.1. Confirmatory analyses

#### 3.1.1. Reinforcement learning GNG

As predicted, the probability of correct responses was credibly higher under the GoToWin condition than the GoToAvoidPunishment condition, $\beta = .043$, 95% HDI = [.021, .066]. This indicates that go responses were easier to learn under GoToWin (vs. GoToAvoidPunishment) condition. Although in the predicted direction, unexpectedly there was no credible difference between the NoGoToAvoidPunishment and the NoGoToWin conditions, $\beta = .020$, 95% HDI = [-.005, .047]. This indicates that participants learned no-go responses equally well under these two no-go conditions. See Fig. 3 for data patterns of all conditions and see Table 2 for model-derived parameter estimates.

#### 3.1.2. Food choice

For experiment 1a, strikingly, the probability of choosing go over no-go items was .8 in the GoToWin vs NoGoToAvoidPunishment comparison. Next, the probability of choosing go items under the choice pair GoToWin_vs_NoGoToAvoidPunishment was credibly higher than the choice pair GoToWin_vs_NoGoToWin, $\beta = .195$, 95% HDI = [.068, .324], and compared to the choice pair GoToAvoidPunishment_vs_NoGoToWin, $\beta = .330$, 95% HDI = [.133, .532]. However, against the prediction, there was no credible difference between the choice pairs GoToWin_vs_NoGoToAvoidPunishment and GoToAvoidPunishment_vs_NoGoToAvoidPunishment, $\beta = .112$, 95% HDI = [-.023, .265]. In summary, this indicates that participant's learning of go and no-go actions translated into the subsequent food choices.

For Experiment 1b (orange bar in Fig. 4). What is noteworthy is that the probability of choosing go over no-go items is .63, and thus in a similar range as previous work using similar training (Chen et al., 2019) with many more items (Table 1). As predicted, the overall probability of choosing go items was credibly higher than 50%, $\beta = .180$, 95% HDI = [.093, .264]. This is line with previous work (Chen et al., 2019) and indicates that participants overall preferred go over no-go items. See

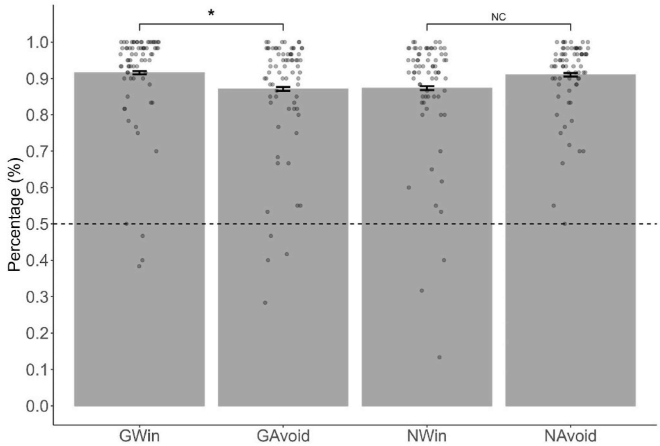


**Fig. 3.** Percentage of correct responses per condition (i.e., GoToWin, GoToAvoidPunishment, NoGoToWin, and NoGoToAvoidPunishment) for the reinforcement learning GNG in Experiment 1a. For illustrative purposes, conditions GoToWin, GoToAvoidPunishment, NoGoToWin, and NoGoToAvoidPunishment are denoted as GWin, GAvoid, NWin, and NAvoid, respectively. Dots refer to the average percentage for each participant. Error bars refer to within-participant standard errors. Credible difference is denoted by the *, and non-credible difference is denoted by NC.

Fig. 4 for data patterns of all conditions and see Table 2 for model-derived parameter estimates.

### 3.2. Exploratory analyses

#### 3.2.1. Transfer from reinforcement learning GNG to food choice

We set out to examine whether there was an influence of the Pavlovian bias during participants' learning of go and no-go responses. First, we fitted a basic Rescorla-Wagner model (M1) (Sutton & Barto, 2018; Zhang et al., 2020) to the reinforcement learning GNG data. This basic model M1 included a learning rate $\varepsilon$ and an inverse temperature parameter $\tau$. M2 added the Pavlovian bias parameter $\pi$ to scale the weight of appetitive and aversive item values. Note that M2 used a same Rescorla-Wagner update rule as the update of item values in M1 according to previous work (Guitart-Masip et al., 2012; Swart et al., 2018). Next, for each model we estimated model evidence using the Widely Applicable Information Criterion (WAIC) (Ahn et al., 2017; Zhang et al., 2020), which provides a metric of model goodness. We then performed model comparison by comparing the WAIC between these two candidate models, and selected the winning model with the lower value of WAIC (Table 3). Results showed that M2 is the winning model. The corresponding posterior parameter estimations are: $\varepsilon = .39$, 95% CI = [.34, .45]; $\tau = .4.81$, 95% CI = [4.22, 5.41]; $\pi = .03$, 95% CI = [.01, .06].

In line with previous work (Guitart-Masip et al., 2012), the winning model included the Pavlovian bias parameter $\pi$, indicating that there was an influence of Pavlovian bias during the learning of go and no-go actions for food items. Additionally, we also explored the strength of go action for all conditions. We found that the ranking of the go action strength across these 4 conditions was as follows, GoToWin > NoGoToWin > GoToAvoidPunishment > NoGoToAvoidPunishment, and the condition difference was credible for each pairwise comparison. See S1 text for details of posterior parameter estimates.

Next, we explored whether the relative differences in the strength of go actions between go and no-go items were correlated with proportion of choosing go (vs. no-go) items in Experiment 1a. Since M2 was the winning model, we obtained these action weights from M2. The correlations were positive under two comparisons GoToWin_vs_NoGoToAvoidPunishment, $r$ (70) = .32, $p < .01$, and GoToAvoidPunishment_vs_NoGoToWin, $r$ (70) = .53, $p < .001$. However, there were no significant correlations under the other four comparisons: GoToWin_vs_NoGoToWin, $r$ (70) = .03, $p$ = .82; GoToAvoidPunishment_vs_NoGoToAvoidPunishment, $r$ (70) = .01, $p$ = .92; GoToWin_vs_GoToAvoidPunishment, $r$ (70) = −.002, $p$ = .99; NoGoToWin_vs_NoGoToAvoidPunishment, $r$ (70) = .18, $p$ = .12. These findings provide some preliminary evidence that reinforcement learning may drive food behavior change by changing item-associated go action strength, but that such relations are not consistently found.

#### 3.2.2. Comparing reinforced and nonreinforced training effects on food choice

Comparing the choice data between Experiments 1a and 1b, it revealed that the overall probability of choosing go items (Experiment 1b) was credibly lower than the probability of choosing go items under the pair GoToWin_vs_NoGoToAvoidPunishment (Experiment 1a), but was comparable to the remaining three choice pairs: GoToWin_vs_NoGoToWin, GoToAvoidPunishment_vs_NoGoToWin, and GoToAvoidPunishment_vs_NoGoToAvoidPunishment (Experiment 1a). This indicates that the reinforcement learning GNG outperformed the nonreinforced GNG only under one specific condition, that is, when go responses were reinforced with obtaining rewards and no-go responses with avoiding punishments. Fig. 3 shows the data patterns and see S1 text for details of behavioral results and model-derived parameter estimates.

#### 3.2.3. Food choice on filler pairs

Results showed that participants choose high-value items with a higher probability than low-value items, and this preference increased as the within-pair value differences increased. See S1 text for details.

## 4. Discussion

### 4.1. Key findings

In the present work, we conducted two experiments to compare effects of reinforced and non-reinforced go/no-go training on subsequent preferences for go and no-go food items. With regard to the learning of go and no-go responses during reinforcement learning, and in line with previous work (Guitart-Masip et al., 2012), we found people learn go

**Table 2**
Behavioral results and model-based parameter estimates for the reinforcement learning GNG and the food choice task in Experiments 1a and 1b.

| Study | Task | Trial type | Response | Percentage | Probability |
|---|---|---|---|---|---|
| 1a | Reinforcement learning GNG | GoToWin | Correct responses | .91 (.005) | .96 [.94, .97] |
| | | GoToAvoidPunishment | | .87 (.005) | .91 [.89, .94] |
| | | NoGoToWin | | .87 (.005) | .91 [.89, .94] |
| | | NoGoToAvoidPunishment | | .91 (.005) | .93 [.92, .95] |
| 1a | Food choice | GoToWin_vs_NoGoToAvoidPunishment | Choosing go items | .80 (.020) | .91 [.85, .97] |
| | | GoToWin_vs_NoGoToWin | | .63 (.020) | .72 [.59, .84] |
| | | GoToAvoidPunishment_vs_NoGoToWin | | .55 (.020) | .58 [.39, .77] |
| | | GoToAvoidPunishment_vs_NoGoToAvoidPunishment | | .67 (.020) | .80 [.65, .94] |
| 1b | Food choice | Go_vs_NoGo (across all conditions) | Choosing go items | .63 (.010) | .68 [.59, .76] |

Note: The numbers under the Percentage column are the mean values with corresponding standard errors inside the parentheses. The numbers under the Probability column are the posterior mean estimates with corresponding 95% HDI inside the brackets.

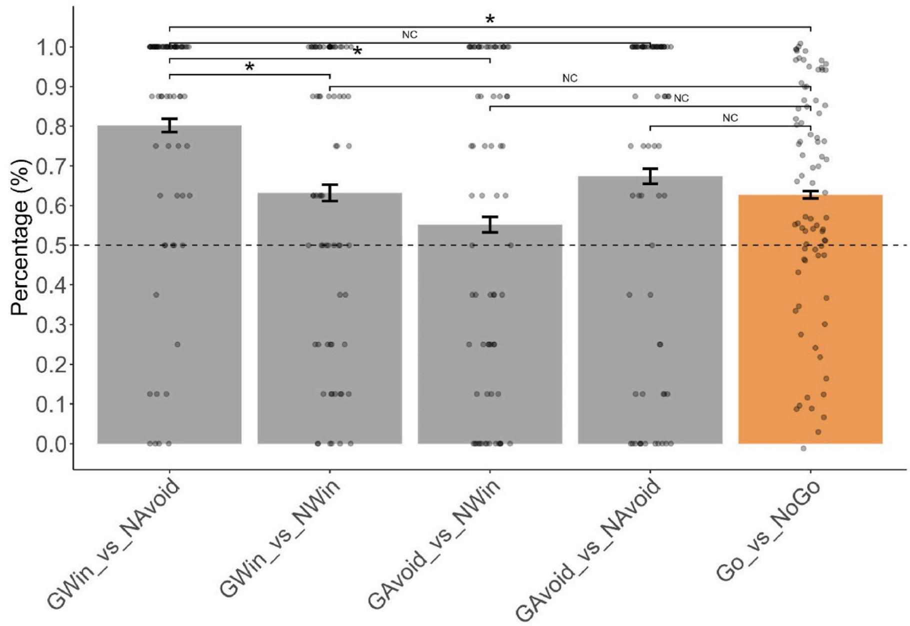


**Fig. 4.** Percentage of choosing go items in Experiment 1a and Experiment 1b. Grey bars represent percentages of choosing go items for four go vs. no-go comparisons (i.e., GoToWin_vs_NoGoToAvoidPunishment, GoToWin_vs_NoGoToWin, GoToAvoidPunishment_vs_NoGoToWin, and GoToAvoidPunishment_vs_NoGoToAvoidPunishment) in Experiment 1a. For illustrative purposes, conditions GoToWin, GoToAvoidPunishment, NoGoToWin, and NoGoToAvoidPunishment are denoted as GWin, GAvoid, NWin, and NAvoid, respectively. The orange bar represents the overall percentages of choosing go items for the go vs. no-go comparison in Experiment 1b. Dots refer to average percentage for each participant. Error bars refer to within-participant standard errors. Credible difference is denoted by the *, and non-credible difference is denoted by NC. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

**Table 3**
WAIC for two candidate models and the winning model (M2) is presented in boldface font.

| Model no. | Model parameters | No. of parameters | WAIC value |
|---|---|---|---|
| 1 | $\varepsilon, \tau$ | 2 | 11882 |
| **2** | $\boldsymbol{\varepsilon, \tau, \pi}$ | **3** | **11708** |

Note: The winning model is represented in boldface type. $\varepsilon$ is the learning rate, $\tau$ is the inverse temperature that captures random noises during decision-making; $\pi$ is the Pavlovian bias. WAIC refers to the Watanabe Akaike Information Criteria that measures the out-of-sample predictive accuracy.

responses better when they are instrumental in obtaining rewards and no-go responses better when they lead to the avoidance of punishment than vice versa. More importantly, we found strongest preferences for go over no-go food items when responses to go foods were instrumental in obtaining rewards and no-go responses toward no-go food items were instrumental in avoiding punishments. The increased preference of around 80% for go over no-go food items was substantially stronger than the often observed 55–60% preference for go over no-go items found for nonreinforced learning (see Table S1) and statistically higher than we found in the current nonreinforced go/no-go training (63%). Another noteworthy finding is that aligning reward and go responses and avoidance of punishment and no-go responses during go/no-go training led to stronger preferences for go over no-go items compared to when both go and no-go responses were instrumental in obtaining rewards. Below we discuss the applied and theoretical implications of these findings.

### 4.2. Applied implications

In the present work we used food items that were similar in (un) healthiness as we were interested in providing first evidence of how reinforcement learning and nonreinforced learning compare in changing food preferences. Nonreinforced go/no-go training has been employed as an intervention tool to promote preferences for healthy over unhealthy food items (Lawrence et al., 2015). Interestingly, a publicly available training app (Aulbach et al., 2021; Keeler et al., 2022) already incorporates this reinforcement schedule (i.e., rewarding successful go responses and punishing commission errors to no-go items), but to date research has not directly examined how this reinforcement influences preferences. The present findings suggest this kind of reinforcement may be beneficial to boost strength in preference change compared to no reinforcement or rewarding both go and no-go responses.

A difference between the implementation of reinforcement in FoodT and the present work is that in FoodT the contingency between reward/punishment and responses is 100%. In light of this, another practical recommendation is to examine whether it may be beneficial to employ the stimulus-outcome contingencies used in the currently employed RL GNG. For example, one may provide the reward feedback 80% of the correct go responses and provide the feedback of avoidance of punishments 80% time of the correct no-go responses. Based on associative learning literature (Chan & Harris, 2019), partial reinforcement is more resistant to extinction and is therefore more likely to support longer-lasting GNG-training effects. Future work is need to examine this possibility.

Practitioners might consider implementing RL GNG training at the category level to examine whether this would further generalize changes in food preferences. For instance, reinforce go actions to vegetables in GoToWin condition only, whereas reinforce no-go actions to chocolates in NoGoToAvoidPunishment condition only. Thus, the present work provides several avenues to further examine how to optimize applied training applications.

### 4.3. Theoretical implications

Two hypotheses were not confirmed in Experiment 1a. First, Participants did not learn no-go actions better under the NoGoToAvoidPunishment condition than the NoGoToWin condition; second, participants did not choose more go items under the GoToWin_vs_NoGoToAvoidPunishment comparison compared to the GoToAvoidPunishment_vs_NoGoToAvoidPunishment comparison. The reason could be that participants had very high accuracies on all four reinforced conditions compared to previous literature (de Boer et al., 2019; Guitart-Masip et al., 2012; Swart et al., 2018), which may make it harder to detect condition differences. Deviations from previous work, and the high levels of accuracy may be due to the use of food items in the present work instead of meaningless fractals used in previous work. Food items have meaning and are more visually distinctive than fractals, which could have boosted learning during the reinforcement learning GNG. Future work could investigate how item type influences reinforcement learning.

Instrumental learning theories suggest both reward and punishment reinforcers should strengthen item-action associations (Maia, 2010; Moutoussis et al., 2021), and hence should increase the probability of choosing go over no-go items more in reinforced tasks compared to nonreinforced go/no-go (GNG) tasks. Our findings show that this is not always the case. We attribute this effect to the influence of Pavlovian bias (Guitart-Masip et al., 2014), which points to the compatibility between rewards and go actions, as well as avoidance of punishments and no-go actions. This finding is noteworthy in light of work in which affective consequences were added to a nonreinforced approach/-avoidance task (Van Dessel et al., 2018). This work showed that adding evaluative consequences to cued actions boosted effects of nonreinforced training on food choice. Here, we show that adding consequences of actions does not always boost nonreinforced learning, but only under conditions where action is coupled to rewards and inaction to avoidance of punishments.

Additionally, we considered how the temporal proximity of items and outcomes during reinforcement learning might affect food item valuation. In the reward conditions (GoToWin and NoGoToWin) foods may be associated with positive value of rewards, while in punishment conditions (GoToAvoidPunishment and NoGoToAvoidPunishment) foods may be associated with negative value of punishments. Our winning model (M2) supported this, showing that item values in winning conditions became more positive, and more negative in losing conditions. This influenced the strength of reinforcement on subsequent choices, as learned responses (No-Go and Go) in the NoGoToWin and GoToAvoidPunishment conditions conflicted with their respective acquired valences (positive and negative).

### 4.4. Limitations and future directions

One limitation of the present study is that participant's food choices were hypothetical due to the online set-up caused by the Covid-19 situation. Yet, we do not think this undermines the validity of the food choice data. First, participant's consistent preference for high-value (vs. low-value) foods on filler trials, indicates sensitive, value-based decisions akin to those in incentive-compatible experiments (Sullivan & Huettel, 2021; Veling et al., 2017). Additionally, the magnitude of the probability of choosing go items in the nonreinforced GNG aligns with results in previous incentive-compatible choice tasks (Chen et al., 2019).

Another limitation is the difference in contingency rates between the reinforced and nonreinforced Go/No-Go (GNG) tasks. In the nonreinforced GNG task, there was a 100% contingency between cues and food stimuli, whereas in the reinforced learning task, the contingency between food stimuli and actions was 80%. This difference is a result from the fact that we used a standard nonreinforced go/no-go task and a standard action-valence asymmetry go/no-go task, and adjusting the one task in the direction of the other task could undermine the effectiveness of the task in changes responses for food. For instance, there are indications in the literature that nonreinforced go/no-go training works less well to change behavior toward food when the contingency between cues and responses is not 100% (Jones et al., 2016). Conversely, reinforcement learning may be impaired when learning is not probabilistic. Future research may systematically examine contingency levels across tasks to isolate the specific effects of reinforcement on learning.

Furthermore, there was a gender imbalance in Experiments 1a (approximately 26% male participants) and 1b (approximately 68% male participants). Since both experiments were conducted online during the COVID-19 pandemic including university students recruited through convenience sampling, it is challenging to determine what explains this imbalance. Nonetheless, we do not think this has influenced the results much in case of the nonreinforced Go/No-Go task (GNG), as previous work demonstrated training effects on food preferences where the majority of participants were women (Chen et al., 2019). This suggests that gender balance in the sample may not be very influential in obtaining nonreinforced GNG effects on food preferences. Future research could systematically explore the role of gender with regard to RL GNG.

Our exploratory analyses showed that RL GNG influenced the strengths assigned to Go actions for specific items. These strengths reflect a mix of action value (linked to instrumental learning) and item value (linked to Pavlovian learning). However, in the winning model (M2), action value and item value are closely connected in mathematical terms, making it unclear how much each type of learning influences food choices. Future research could improve M2 by adding an instrumental learning bias (Swart et al., 2018), using the drift diffusion model (DDM) (Sullivan & Huettel, 2021), or a combined RL-DDM (Fontanesi et al., 2019; Kutlikova et al., 2023) to separate the effects of action and item values on food choices.

Future research directions for current research include examining the longevity of choice preferences from reinforcement learning GNG over varying time frames (e.g., weeks to months) and comparing their persistence against nonreinforced GNG. Second, to investigate generalization within the reinforcement learning GNG task, it would be valuable to train responses to specific food categories (e.g., junk food and healthy food) and assess whether the learned responses generalize to other untrained items that from the same category. Finally, current reinforcement learning models focused on immediate reward and punishment. Future studies may incorporate temporal difference (TD) learning (Sutton & Barto, 2018) to consider long-term expected rewards to understand of reinforcement learning GNG data.

## 5. Conclusion

In conclusion, our study compared reinforcement learning with nonreinforced learning in the go/no-go task. We demonstrated that reinforcement learning was superior only under one condition, namely when rewarding go actions and avoiding punishment for no-go actions. Thus, rewarding go actions toward desired food and making sure people avoid punishment by not responding to undesired food is perhaps the most effective way to change food choices in this type of training.

## CRediT authorship contribution statement

**Huaiyu Liu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Julian Quandt:** Writing – review & editing, Software, Methodology, Investigation. **Lei Zhang:** Writing – review & editing, Methodology. **Xiongbing Kang:** Resources. **Jens Blechert:** Writing – review & editing, Supervision. **Tjits van Lent:** Writing – review & editing. **Rob W. Holland:** Writing – review & editing, Supervision. **Harm Veling:** Writing – review & editing, Supervision, Conceptualization.

## Ethical statement

This research complies with the Declaration of Helsinki (2023), and received approval from a local ethics board (ID: ECSW-2019-003).

### Declaration of generative AI in scientific writing

No artificial intelligence assisted technologies were used in this research or the creation of this article.

### Funding

This research did not receive any funding.

### Declaration of competing interest

No potential conflict of interest was reported by the author(s). All authors declare no conflicts of interest.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.appet.2025.107950.

### Data availability

All data and code for have been made publicly available on the Open Science Framework. Please see text for detail.

### References

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Computational Psychiatry, 1*(0), 24. https://doi.org/10.1162/CPSY_a_00002

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25. https://doi.org/10.1037/amp0000191

Aulbach, M. B., Knittle, K., van Beurden, S. B., Haukkala, A., & Lawrence, N. S. (2021). App-based food Go/No-Go training: User engagement and dietary intake in an opportunistic observational study. *Appetite, 165*, Article 105315. https://doi.org/10.1016/j.appet.2021.105315

Bakkour, A., Leuker, C., Hover, A. M., Giles, N., Poldrack, R. A., & Schonberg, T. (2016). Mechanisms of choice behavior shift using cue-approach training. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00421

Becker, D., Jostmann, N. B., Wiers, R. W., & Holland, R. W. (2015). Approach avoidance training in the eating domain: Testing the effectiveness across three single session studies. *Appetite, 85*, 58–65. https://doi.org/10.1016/j.appet.2014.11.017

Blechert, J., Lender, A., Polk, S., Busch, N. A., & Ohla, K. (2019). Food-Pics_ Extended—an image database for experimental research on eating and appetite: Additional images, normative ratings and an updated review. *Frontiers in Psychology, 10*, 307. https://doi.org/10.3389/fpsyg.2019.00307

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, Article e9414. https://doi.org/10.7717/peerj.9414

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using stan. *Journal of Statistical Software, 80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Cavanagh, J. F., Eisenberg, I., Guitart-Masip, M., Huys, Q., & Frank, M. J. (2013). Frontal theta overrides pavlovian learning biases. *Journal of Neuroscience, 33*(19), 8541–8548. https://doi.org/10.1523/JNEUROSCI.5754-12.2013

Chan, C. K. J., & Harris, J. A. (2019). The partial reinforcement extinction effect: The proportion of trials reinforced during conditioning predicts the number of trials to extinction. *Journal of Experimental Psychology: Animal Learning and Cognition, 45*(1), 43–58. https://doi.org/10.1037/xan0000190

Chen, Z., Holland, R. W., Quandt, J., Dijksterhuis, A., & Veling, H. (2019). When mere action versus inaction leads to robust preference change. *Journal of Personality and Social Psychology, 117*(4), 721–740. https://doi.org/10.1037/pspa0000158

Chen, Z., Veling, H., Dijksterhuis, A., & Holland, R. W. (2016). How does not responding to appetitive stimuli cause devaluation: Evaluative conditioning or response inhibition? *Journal of Experimental Psychology: General, 145*, 1687–1701. https://doi.org/10.1037/xge0000236

de Boer, L., Axelsson, J., Chowdhury, R., Riklund, K., Dolan, R. J., Nyberg, L., Bäckman, L., & Guitart-Masip, M. (2019). Dorsal striatal dopamine D1 receptor availability predicts an instrumental bias in action learning. *Proceedings of the National Academy of Sciences, 116*(1), 261–270. https://doi.org/10.1073/pnas.1816704116

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Estacio, E. V. (2018). Health psychology: Theory, research and practice. *Health Psychology*, 1–832.

Fontanesi, L., Gluth, S., Spektor, M. S., & Rieskamp, J. (2019). A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin & Review, 26*(4), 1099–1121. https://doi.org/10.3758/s13423-018-1554-2

Guitart-Masip, M., Duzel, E., Dolan, R., & Dayan, P. (2014). Action versus valence in decision making. *Trends in Cognitive Sciences, 18*(4), 194–202. https://doi.org/10.1016/j.tics.2014.01.003

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage, 62*(1), 154–166. https://doi.org/10.1016/j.neuroimage.2012.04.024

Houben, K., & Jansen, A. (2015). Chocolate equals stop. Chocolate-specific inhibition training reduces chocolate intake and go associations with chocolate. *Appetite, 87*, 318–323. https://doi.org/10.1016/j.appet.2015.01.005

Jones, A., Di Lemma, L. C. G., Robinson, E., Christiansen, P., Nolan, S., Tudur-Smith, C., & Field, M. (2016). Inhibitory control training for appetitive behaviour change: A meta-analytic investigation of mechanisms of action and moderators of effectiveness. *Appetite, 97*, 16–28. https://doi.org/10.1016/j.appet.2015.11.013

Keeler, J. L., Chami, R., Cardi, V., Hodsoll, J., Bonin, E., MacDonald, P., Treasure, J., & Lawrence, N. (2022). App-based food-specific inhibitory control training as an adjunct to treatment as usual in binge-type eating disorders: A feasibility trial. *Appetite, 168*, Article 105788. https://doi.org/10.1016/j.appet.2021.105788

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Kutlikova, H. H., Zhang, L., Eisenegger, C., van Honk, J., & Lamm, C. (2023). Testosterone eliminates strategic prosocial behavior through impacting choice consistency in healthy males. *Neuropsychopharmacology, 48*(10), 1541–1550. https://doi.org/10.1038/s41386-023-01570-y

Lawrence, N. S., O'Sullivan, J., Parslow, D., Javaid, M., Adams, R. C., Chambers, C. D., Kos, K., & Verbruggen, F. (2015). Training response inhibition to food is associated with weight loss and reduced energy intake. *Appetite, 95*, 17–28. https://doi.org/10.1016/j.appet.2015.06.009

Liu, H., Holland, R. W., Blechert, J., Quandt, J., & Veling, H. (2022). Devaluation of NoGo stimuli is both robust and fragile. *Cognition & Emotion, 36*(5), 876–893. https://doi.org/10.1080/02699931.2022.2067132

Maia, T. V. (2010). Two-factor theory, the actor-critic model, and conditioned avoidance. *Learning & Behavior, 38*(1), 50–67. https://doi.org/10.3758/LB.38.1.50

Marteau, T. M., Hollands, G. J., & Fletcher, P. C. (2012). Changing human behavior to prevent disease: The importance of targeting automatic processes. *Science, 337*(6101), 1492–1495. https://doi.org/10.1126/science.1226918

Moutoussis, M., Garzón, B., Neufeld, S., Bach, D. R., Rigoli, F., Goodyer, I., Bullmore, E., Fonagy, P., Jones, P., Hauser, T., Romero-Garcia, R., Clair, M. S., Vértes, P., Whitaker, K., Inkster, B., Prabhu, G., Ooi, C., Toseeb, U., Widmer, B., … Dolan, R. J. (2021). Decision-making ability, psychopathology, and brain connectivity. *Neuron, 109*(12), 2025–2040.e7. https://doi.org/10.1016/j.neuron.2021.04.019

Quandt, J., Holland, R. W., Chen, Z., & Veling, H. (2019). The role of attention in explaining the no-go devaluation effect: Effects on appetitive food items. *Journal of Experimental Psychology: Human Perception and Performance, 45*(8), 1119–1133. https://doi.org/10.1037/xhp0000659

R Core Team. (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/.

Schonberg, T., Bakkour, A., Hover, A. M., Mumford, J. A., Nagar, L., Perez, J., & Poldrack, R. A. (2014). Changing value through cued approach: An automatic mechanism of behavior change. *Nature Neuroscience, 17*(4). https://doi.org/10.1038/nn.3673. Article 4.

Sullivan, N. J., & Huettel, S. A. (2021). Healthful choices depend on the latency and rate of information accumulation. *Nature Human Behaviour, 5*(12). https://doi.org/10.1038/s41562-021-01154-0. Article 12.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning. *An introduction* (2nd ed.). MIT Press.

Swart, J. C., Frank, M. J., Määttä, J. I., Jensen, O., & Cools, R. (2018). Frontal network dynamics reflect neurocomputational mechanisms for reducing maladaptive biases in motivated action. *PLoS Biology, 16*(10), Article e2005979. https://doi.org/10.1371/journal.pbio.2005979. Ouden, H. E. M. den.

van Alebeek, H., Veling, H., & Blechert, J. (2023). Disentangling go/no-go from motivational orientation to foods: Approaching is more than just responding. *Food Quality and Preference, 106*, Article 104821. https://doi.org/10.1016/j.foodqual.2023.104821

Van Dessel, P., Hughes, S., & De Houwer, J. (2018). Consequence-based approach-avoidance training: A new and improved method for changing behavior. *Psychological Science, 29*(12), 1899–1910. https://doi.org/10.1177/0956797618796478

Veling, H., Becker, D., Liu, H., Quandt, J., & Holland, R. W. (2022). How go/no-go training changes behavior: A value-based decision-making perspective. *Current Opinion in Behavioral Sciences, 47*, Article 101206. https://doi.org/10.1016/j.cobeha.2022.101206

Veling, H., Chen, Z., Tombrock, M. C., Verpaalen, I. A. M., Schmitz, L. I., Dijksterhuis, A., & Holland, R. W. (2017). Training impulsive choices for healthy and sustainable food. *Journal of Experimental Psychology: Applied, 23*(2), 204–215. https://doi.org/10.1037/xap0000112

Veling, H., Verpaalen, I. A. M., Liu, H., Mosannenzadeh, F., Becker, D., & Holland, R. W. (2021). How can food choice best be trained? Approach-Avoidance versus go/no-go training. *Appetite, 163*, Article 105226. https://doi.org/10.1016/j.appet.2021.105226

Wickham, H. (2016). Programming with ggplot2. In H. Wickham (Ed.), *Ggplot2: Elegant graphics for data analysis* (pp. 241–253). Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4_12.

Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience, 15*(6), 695–707. https://doi.org/10.1093/scan/nsaa089