Received: 30 July 2019 | Revised: 19 September 2019 | Accepted: 7 October 2019

DOI: 10.1002/hbm.24859

RESEARCH ARTICLE

WILEY

Enhanced striatal and prefrontal activity is associated with individual differences in nonreinforced preference change for faces

Tom Salomon[1] | Rotem Botvinik-Nezer[1,2] | Shiran Oren[1,2] | Tom Schonberg[1,2]

[1]Department of Neurobiology, Tel Aviv University, Tel Aviv, Israel

[2]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

Correspondence
Tom Schonberg, Department of Neurobiology, Faculty of Life Sciences and Sagol School of Neuroscience, Tel Aviv University, Ramat Aviv 6997801, Tel Aviv, Israel.
Email: schonberg@tauex.tau.ac.il

Funding information
H2020 European Research Council, Grant/Award Number: 715016; Israeli Science Foundation

Abstract

Developing effective preference modification paradigms is crucial to improve the quality of life in a wide range of behaviors. The cue-approach training (CAT) paradigm has been introduced as an effective tool to modify preferences lasting months, without external reinforcements, using the mere association of images with a cue and a speeded button response. In the current work for the first time, we used fMRI with faces as stimuli in the CAT paradigm, focusing on face-selective brain regions. We found a behavioral change effect of CAT with faces immediately and 1-month after training, however face-selective regions were not indicative of behavioral change and thus preference change is less likely to rely on face processing brain regions. Nevertheless, we found that during training, fMRI activations in the ventral striatum were correlated with individual preference change. We also found a correlation between preference change and activations in the ventromedial prefrontal cortex during the binary choice phase. Functional connectivity among striatum, prefrontal regions, and high-level visual regions was also related to individual preference change. Our work sheds new light on the involvement of neural mechanisms in the process of valuation. This could lead to development of novel real-world interventions.

KEYWORDS

choice behavior, decision making, learning, ventral striatum, ventromedial prefrontal cortex

1 | INTRODUCTION

The process by which preferences are constructed and modified in the brain is a main theme in the research of value-based decision making (Fellows, 2011; Glimcher & Fehr, 2013; Lichtenstein & Slovic, 2006; Rangel, Camerer, & Montague, 2008; Vlaev, Chater, Stewart, & Brown, 2011). Studies using external reinforcements in humans, revealed significant contribution of the striatum in associative learning (e.g., O'Doherty et al., 2004; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006; Rangel et al., 2008). Following learning, value representation is putatively enhanced within the ventromedial prefrontal cortex (vmPFC; Chib, Rangel, Shimojo, & O'Doherty, 2009; Clithero & Rangel, 2014; Kable & Glimcher, 2009; Rangel et al., 2008).

Alongside these findings, real-world applications for reinforcement-based behavioral change were developed such as weight loss and smoking abstinence programs. These interventions have been proven effective in the short term, however, in the long term, they are often

hard to maintain and have diminished effects over time (Cahill & Perera, 2011; Prochaska, Delucchi, & Hall, 2004).

In contrast to reinforcement-based interventions, in the recently developed cue-approach training (CAT) paradigm, preference modification has been induced without external reinforcements (Schonberg et al., 2014). In this paradigm, the mere association of snack-food stimuli with a neutral cue and a speeded button-press response (Go stimuli), results in enhanced preference for trained stimuli. In a binary-choice probe phase following CAT, Go stimuli are preferred over stimuli that had not been paired with the cue and response (NoGo stimuli) with similar initial-value. Behaviorally, CAT effect has been established in dozens of independent samples, which have also demonstrated the effect is general beyond food-items and induces long-lasting impact for up to 6 months (Bakkour et al., 2016; Botvinik-Nezer, Salomon, & Schonberg, 2019; Salomon et al., 2018; Veling et al., 2017; Zoltak, Veling, Chen, & Holland, 2018).

While preference modification using CAT has been well-established behaviorally, the underlying neural mechanisms have not been fully uncovered by previous imaging studies using snack-food stimuli (Bakkour, Lewis-Peacock, Poldrack, & Schonberg, 2017; Botvinik-Nezer et al., 2019; Schonberg et al., 2014). In these studies, CAT had a stronger effect on preferences for snacks with initial high-value; thus, previous studies focused on the neural modification for high-value stimuli. Schonberg et al. (2014) found that vmPFC fMRI activity during probe choices of Go over NoGo stimuli was modulated by the proportion of trials Go stimuli were chosen. Recently, CAT has also been found to induce a change in the representation of stimuli during passive viewing, resulting in enhanced occipital visual processing of Go stimuli and reduced top–down parietal attention activity (Botvinik-Nezer et al., 2019). However, neuroimaging attempts to examine the change during training itself, had not resulted in conclusive evidence for the mechanisms underlying CAT, neither using standard univariate approach (Schonberg et al., 2014), nor using multivariate pattern analysis (Bakkour et al., 2017).

Thus, in the current work, we aimed to reveal the underlying neural changes during training using face stimuli, motivated by several key features of faces. First, faces are important social stimuli, known to hold innate value features encoded in the brain, most notably in value-processing regions such as the vmPFC, cingulate cortex, and ventral striatum (Aharon et al., 2001; Cloutier, Heatherton, Whalen, & Kelley, 2008; Kranz & Ishai, 2006; Senior, 2003; Smith, Clithero, Boltuck, & Huettel, 2014). The strong preference response naturally evoked by face stimuli, even during preference-irrelevant tasks (Lebreton, Jorge, Michel, Thirion, & Pessiglione, 2009), lends them as prime stimuli to examine processes of preference modification. Second, faces are processed by specialized regions of interest (ROIs) brain network (Fox, Iaria, & Barton, 2009; Kanwisher, McDermott, & Chun, 1997; Yovel & Kanwisher, 2004). Thus, faces uniquely enabled us to focus the fMRI analysis within this network of specialized regions, potentially bridging early visual processing with high-level feature representation, including preferences. Additionally, recent studies with faces have found that CAT enhanced preference for both high- and low-value face stimuli (Salomon et al., 2018). Combining both value-categories, which was not done in previous imaging studies (Bakkour et al., 2017; Botvinik-Nezer et al., 2019; Schonberg et al., 2014), could potentially improve detection power. Finally, a better understanding of the mechanisms by which preferences of faces are modified, could lead to development of novel interventions. For example, preference modification intervention for faces could contest undesired behaviors affected by social biases such as race and gender discrimination (Meissner & Brigham, 2001; Stamarski & Son Hing, 2015). Similarly, in the clinical field, for conditions characterized with biases in attention toward negative affective stimuli, such as depression (Elliott, Zahn, Deakin, & Anderson, 2011; Peckham, McHugh, & Otto, 2010), attentional bias modification therapy has been suggested to affect clinical symptoms by directing attention toward faces of positive valence (Browning, Holmes, Charles, Cowen, & Harmer, 2012). Inducing a preference modification for face stimuli could potentially be used to alleviate these conditions. Therefore, in the current work, we utilized the unique properties of face stimuli and CAT in order to shed new light on the mechanisms underlying nonreinforced preference modification.

2 | METHODS

2.1 | Codes and data accessibility

Our sample size, hypotheses, and a *general* analyses plan were preregistered on the open science framework (OSF), after data collection began, but prior to data analysis (project page: https://osf.io/aqnr4/; preregistration: https://osf.io/k7wn6). Deviations from the preregistration are described at the end of the methods section below. All behavioral and eye tracking data along with the codes used for their analyses and the codes used for fMRI analyses are shared on https://github.com/tomsalomon/CAT_MRI_faces. Magnetic resonance imaging (MRI) data are available in brain imaging data structure (BIDS) format at https://openneuro.org/datasets/ds001818. Uncorrected and cluster-corrected statistical maps of all contrasts described in current work are available at https://neurovault.org/collections/5578/.

2.2 | Participants

A total of $N = 50$ healthy participants were scanned, out of which $N = 42$ valid participants were included in the final analyses. All participants had normal or corrected to normal vision and hearing, and no background of neurological disorders or medication. Participants gave their informed consent to participate in the experiment and received monetary compensation for their time. The study was approved by the ethics committee of Tel Aviv University and institutional review board at the Sheba Tel-Hashomer medical center.

We preregistered our target sample size of $N = 45$ (https://osf.io/35sf3), based on 80% power analysis aimed to find an effect in a vmPFC mask during probe, using data from a previous imaging study with CAT (Schonberg et al., 2014). Our final sample consisted of $N = 42$ valid participants (23 females), ages 19–38 ($M = 25.95$, $SD = 4.32$). Out of the 42 participants, $n = 25$ agreed to return for an

addition follow-up scanning session after a mean period of 34.6 days (*SD* = 15.08; range = 14–98 days; interquartile range = 28–35 days). Due to a scanner upgrade, data collection was terminated early, before we could complete the predetermined sample size.

Eight participants took part in the experiment, but were excluded from the final analysis: Three participants were excluded due to incidental clinical findings; two participants did not complete the tasks inside the scanner; one participant was excluded due to predefined training exclusion criteria (ladder drop below 200-ms, see below) which was used in previous CAT experiments (Salomon et al., 2018); one participant chose the Go stimuli in 100% of probe trials and thus could not be analyzed, and another participant had technical issues with their scans.

2.3 | Materials

2.3.1 | Stimuli

We used a stimulus set of 60 face images, selected from the Siblings Dataset (Vieira, Bottino, Laurentini, & De Simone, 2014), as used in a previous behavioral CAT publication with faces (Salomon et al., 2018). The stimulus set comprised of 30 male and 30 female front-facing individuals, posing a neutral expression with limited facial hair and make-up. The original images were cropped to identical size (400 × 500 pixels) and the original green screen background was replaced with a homogenous gray background. Faces were aligned by positioning each figure's pupils in fixed coordinates symmetrically around the center of the image ([150, 250] and [250, 250] for the left and right figure's pupil, respectively).

2.3.2 | Cue

In the training task, we used a neutral auditory cue of a 180-ms sinusoidal wave tone as a Go signal to which participants were required to respond with a rapid button press, similarly to the cue used in previous cue-approach studies (Bakkour et al., 2016, 2017; Salomon et al., 2018; Schonberg et al., 2014).

2.3.3 | Stimuli presentation

Stimuli were presented using MATLAB R2014b (Mathworks, Inc. Natick, MA), Psychtoolbox-3 (Kleiner et al., 2007) package, using a 21.5'' iMac for the tasks outside of the scanner (initial preference evaluation and demos before the scan, as well as postscan memory task). During scanning, a MacBook Pro was used to present images inside the MRI, which were projected to a NordicNeuroLab 32'' LCD display (1,920 x 1,080 pixels resolution, 120 Hz image refresh rate) that participants viewed through a mirror. The sound during the training part was played in a controlled volume, using Sensimetrics S14 in-ear MRI compatible headphones. Participants' gaze was constantly monitored online and recorded at 500 Hz using SR-Research EyeLink 1000 Plus eye-tracker.

2.4 | Procedure

Overall procedure (Figure 1) followed a similar course to that of previous CAT studies with faces (Salomon et al., 2018)—begining with an initial preference evaluation task, followed by CAT and a binary choice probe phase. Additionally, we introduced immediately before and after CAT, a passive viewing task (following similar procedure to the one used by Botvinik-Nezer et al., 2019), as well as an additional dynamic face localizer task, adapted from previous work (Bernstein et al., 2018), which was the last task inside the MRI, after the probe task. A detailed description of each task appears below.

2.4.1 | Baseline evaluation of subjective preference

Outside the MRI scanner, participants' baseline subjective preference for the 60 individual face stimuli were evaluated using a forced-choice binary ranking procedure. Participants were presented with 300 unique binary choices during which they were asked to select their preferred stimulus out of two randomly paired face stimuli, within an allocated 2,500-ms time window (see Figure 1a). Each stimulus was presented in exactly 10 choice trials to maintain a similar exposure to all stimuli.

Similar to our previous work with face stimuli (Salomon et al., 2018), binary choices acquired in the baseline subjective preference phase, were transformed into ranking scores using the Colley Matrix algorithm (Colley, 2002). In the algorithm, based on the assumption of choice transitivity, outcomes from the binary choices were used to calculate a ranking score indicating the subjective value for each of the 60 face stimuli.

Colley Matrix ranking scores typically range in scores from 0 (least liked) to 1 (most liked), with a fixed mean of 0.5. An intransitive choice pattern is characterized by densely distributed scores around the center of 0.5. We calculated transitivity scores for each participant using the standard deviation (*SD*) of the Colley-ranking scores. From these transitivity scores, we defined an exclusion criterion, similarly to previous studies (Salomon et al., 2018); participants with transitivity scores 3 *SD* below the group mean would be excluded from analysis due to intransitivity of choices. However, in the current study, no participant had been excluded following this criterion.

2.4.2 | Passive viewing: Baseline

Following the initial preference evaluation task, participant entered the MRI scanner. Their first task inside the scanner was a passive

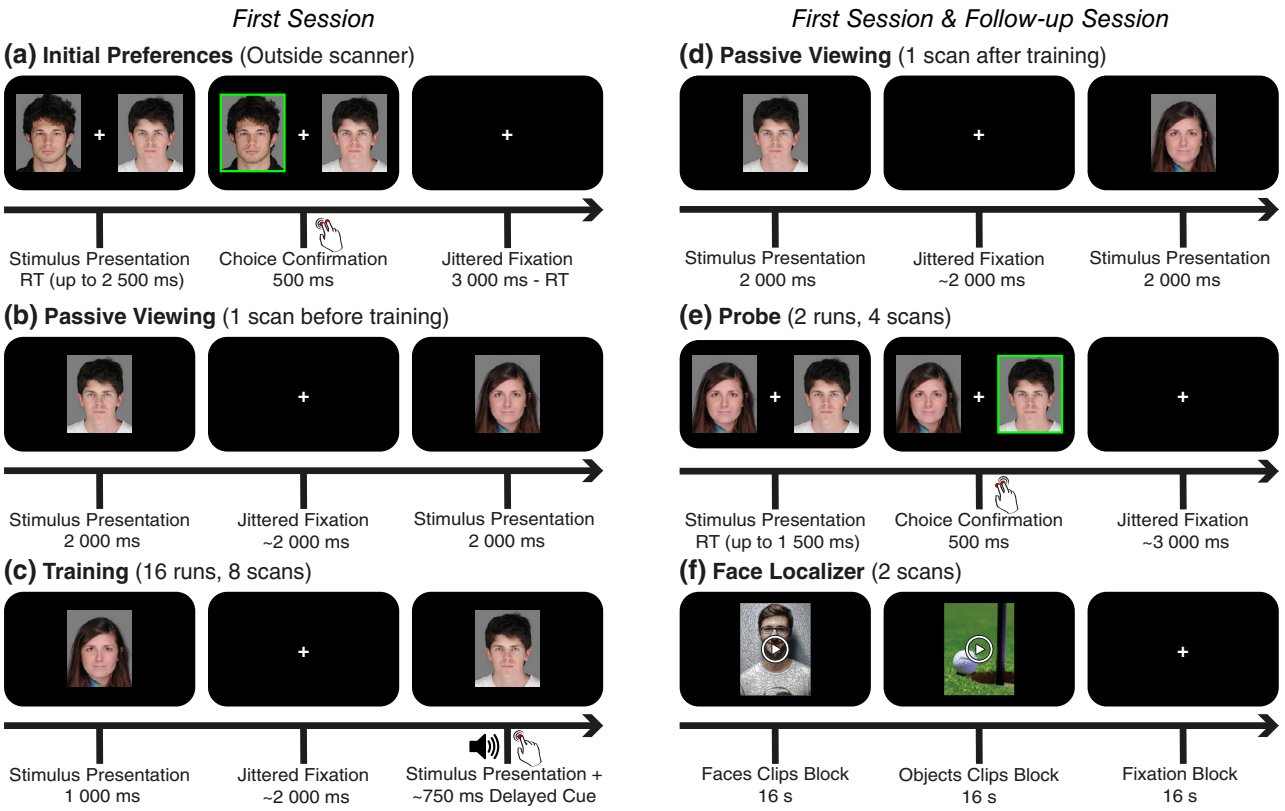

FIGURE 1 Experimental procedure: General outline of the experimental procedure. (a) participants' subjective preference for 60 face images was first evaluated outside the scanner using a binary forced-choice task. A subset of 40 face stimuli were selected for the following tasks inside the MRI (see Figure S1). (b) In a passive viewing task, participants viewed all 40 stimuli while performing a silent counting sham task. (c) During the 40 min cue-approach training (CAT) task, stimuli were presented individually. Twelve stimuli (30%) were designated to be Go stimuli, consistently associated with a delayed neutral auditory cue to which participants were required to respond with a rapid button press response; The remaining 70% of stimuli (NoGo stimuli) appeared without a cue and response. (d) Participants' neural response to the stimuli was tested once again following CAT. (e) During the probe phase, participants were asked to indicate their preferred stimulus out of pairs with similar initial value (either both high-value or both-low value) in which one stimulus was Go and the other NoGo stimulus. (f) Face selective regions were functionally identified using a faces and objects localizer task (Bernstein, Erez, Blank, & Yovel, 2018). Participants were presented with 16-second long blocks of faces, objects, and fixation cross. Each block comprised of 16 short clips, while performing a sham one-back task. In a 30-day follow-up session, participants performed again tasks d–f. Face images (a–e) are included with permission from the copyright holder (Vieira et al., 2014). Illustration of the face localizer task (f) was done using public domain images, similar to the ones actually used in the experiment

viewing task, in which the neural response to 40 face stimuli was measured (see Figure 1b). The stimulus set consisted of 20 high-value (ranked 3–22, above the median rank) and 20 low-value stimuli (ranked 39–58, below the median rank). Each stimulus was presented individually on the screen for 2,000 ms, followed by a jittered interstimulus interval (ISI) presentation of a fixation cross (mean duration 2,000 ms, range 1,000–6,000 ms, 1,000 ms increments), resulting in 170-s long scan. To make sure participants were engaged with the viewing task, participants performed a sham silent counting task of either male or female faces (the gender to be counted was counter-balances across participants, based on the participant code). Participants were asked to report at the end of the scan, how many male or female faces they had counted. One scan was performed at baseline before CAT, and another scan was performed after CAT.

2.4.3 | Cue-approach training

The CAT task's protocol followed that of previous CAT with faces (Salomon et al., 2018). Each stimulus in the training set was presented individually on the screen for 1,000 ms, once during each training run (see Figure 1c). Stimuli were randomly ordered and followed by a jittered fixation cross with an average duration of 2,000 ms (range of 1,000–6,000 ms, 1,000 ms intervals). The task comprised of eight scans (lasting 252 s each). In each scan, two training runs were completed, resulting in a total of 16 presentations for each stimulus throughout the entire training task. The stimulus set included the same 40 stimuli presented in the previous passive viewing task. Out of the 40 stimuli, 30% (12 stimuli: six high-value, M_{rank} = 12.5, and six low-value stimuli M_{rank} = 48.5; see Figure S1) were associated with the Go cue. Participants were instructed to respond to the Go cue by

pressing a response box button using their right-hand index finger, as fast as possible, before stimulus offset. Participants were not informed in advance that the association of stimuli with the cue will be consistent or which stimuli would be Go stimuli. The cue appeared following a delay of 750 ms from stimulus onset, at the beginning of the task. To maintain a balanced difficulty level throughout the training phase, the cue delay was modified according to participants' performance, as conducted in previous cue-approach studies (Bakkour et al., 2016, 2017; Salomon et al., 2018; Schonberg et al., 2014). Each successful response resulted in a 16.66 ms increase of the Go signal delay (GSD), thus making the task more challenging; whereas failing to respond on time resulted in 50 ms decrease (1:3 ratio) of the GSD. GSD was updated independently for high-value (overall, M_{GSD} = 738.37 ms, *SD* = 63.75 ms; last scan M_{GSD} = 756.19 ms, *SD* = 76.69 ms) and low-value trials (M_{GSD} = 730.95 ms, *SD* = 65.94 ms; last scan M_{GSD} = 738.55 ms, *SD* = 79.38 ms). Importantly, participants did not receive any feedback about their performance throughout the task, neither after individual trials, nor at the end of the scan.

2.4.4 | Anatomical scans

After the last training scan, participants were given about 9 min of rest before continuing to the next task, during which FLAIR and T1w anatomical scans were performed.

2.4.5 | Passive viewing: Post training

After the anatomical scans, participants underwent an additional passive viewing task, identical to the one performed before training (Figure 1d). In the second scan, the same stimuli were presented in a new random order. Participants were once again asked to silently count how many male or female faces appeared and report the number at the end of the scan. The target gender to be counted was always the opposite of that used in the baseline passive viewing scan.

2.4.6 | Probe

In the probe task, preference modification following CAT was evaluated in a forced-choice task. On each probe trial, participants had to choose between two stimuli with similar initial value—both either high-value or low-value. In each pair, one of the stimuli was a Go stimulus, previously associated with a cue and a button press during training. Six Go stimuli were compared against six NoGo stimuli of equal mean rank, for a total of 36 (6 x 6) unique comparisons per value category (see Figure S1). In four additional "sanity check" trials, two high-value NoGo stimuli were pitted against two low-value NoGo stimuli. Each trial lasted 2,000 ms—participants had a 1,500 ms time window to make their choice, which was followed by a 500 ms choice confirmation (green frame around the chosen option) and a fixation cross until the end of the trial. Missed trials, in which the participant failed to decide in the allocated time frame, were followed by a 500 ms warning massage prompting to perform faster choices. Trials were separated with a fixation cross presented for a jittered duration (*M* = 3,000 ms, range = 1,000–11,000 ms, 1,000 ms increments; see Figure 1e).

The entire task consisted of 76 unique trials (36 high-value Go vs. NoGo, 36 low-value Go vs. NoGo, and four high-value vs. low-value sanity trials). All choices were repeated twice, once in each of the two probe runs, presented in a random order. In order to ensure each scan will not exceed 6 min, each probe run was equally spilt into two scans (in which half of the trials of each category were presented), for a total of four probe scans, lasting 200 s each.

2.4.7 | Dynamic face localizer

To identify face selective brain regions at the individual level, we included a dynamic face localizer task adapted from previous work (Bernstein et al., 2018). In the task, participants viewed in each 336 s long scan, 16 s blocks of fixation point (five blocks per scan), video clips of faces (eight blocks per scan), and video clips of objects (eight blocks per scan). Each video-clips block consisted of a series of 16 short 1,000 ms video clips (Figure 1f). To maintain task engagement, participants performed a one-back task, in which they were required to press a button when an identical video clip was presented twice in a row. Participants performed two scans of the face localizer task, each with a different stimuli and blocks order.

2.4.8 | Memory

At the end of the experiment, participants performed a recognition task outside the scanner. In this task, a single face stimulus appeared on screen during each trial. The task included 56 trials: 28 face stimuli that appeared during the probe phase, and another 28 new face stimuli that the participants had never encountered before. For each stimulus, participants were first asked to report whether the stimulus had been presented during the experiment, followed by whether the stimulus had been associated with a cue during the training phase. Reporting was done on a 5-points confidence scale (1-certain it was, 2-think it was, 3-not sure, 4-think it was not, 5-certain it was not), without time restriction, though participants were encouraged to respond as quickly as possible.

2.4.9 | Follow-up

All participants had been notified when they signed up for the experiment that there will be an additional follow-up session and were encouraged to return after a predetermined period of 1 month. In the follow-up session, participants completed the same tasks that followed CAT in the first session: passive viewing, probe, and face localizer inside the scanner and a recognition task outside the scanner.

The tasks were identical and included the same stimuli used in each participant's first session, presented in a random order.

2.5 | Imaging data acquisition

Data were acquired using a 3T Siemens Prisma scanner, with a 64-channel head coil. For structural data, T1w high resolution (1 mm^3) whole brain images were acquired with a magnetization prepared rapid gradient echo pulse sequence with repetition time (TR) of 2.53 s, echo time (TE) of 2.88 ms, flip angle (FA) = 7°, field-of-view (FOV) = 224 × 224 × 208 mm, resolution = 1 × 1 × 1. Functional imaging data were acquired with a T2* weighted multiband echo planar imaging protocol, with a repetition time TR = 2,000 ms, TE = 30 ms, FA = 90^0, and multiband acceleration factor of two and parallel imaging factor (iPAT) = 2, scanned in an interleaved fashion. Image resolution was 2 × 2 × 2.5 mm voxels (0.5 mm gap between axial slices), FOV = 194 × 230 × 195 mm (97 × 115 × 78 acquisition matrix). All images were acquired at a 30° angle off the anterior–posterior commissures (AC–PC) line, to reduce signal dropout in the ventral frontal cortex (Deichmann, Gottfried, Hutton, & Turner, 2003).

2.6 | Behavioral data analysis

In order to evaluate the preference-change effect following CAT, we analyzed the proportion of probe trials in which participants chose the Go over the NoGo stimulus. In line with previous findings with CAT (Salomon et al., 2018), we expected participants to choose both high-value as well as low-value Go stimuli over similar value NoGo stimuli above chance level (50% proportion; log-odds = 0; odds-ratio = 1). Thus, CAT preference modification effect was tested using a one-sided repeated measures logistic regression analysis.

We did not expect to find a significantly stronger effect for high-value probe pairs (Salomon et al., 2018); therefore, differences between the two pair-types were tested using a two-sided logistic regression analysis. Each result is reported with the respective odds-ratio (OR) and its 95% confidence interval (CI), as the corresponding effect size estimates. Analyses and visualizations were conducted using lme4 (Bates, Mächler, Bolker, & Walker, 2015) and ggplot2 (Wickham, 2016) R packages, and are available online along with the experimental data at https://osf.io/aqnr4/.

2.7 | Eye-tracking data analysis

In three additional analyses, we inspected the proportion of time participants fixated on the Go versus NoGo stimuli during passive viewing, training, and probe task. In the passive viewing and the training tasks, gaze durations were averaged for each participant across the Go and NoGo trials, and contrasted within participant. In the probe task, we examined in the forced binary choice paradigm the proportion of time participants viewed the Go stimuli versus the competing NoGo stimuli. Relative gaze durations for the competing Go and NoGo stimuli were contrasted within individual trials. Previous work found that during the probe phase, participants fixated on unchosen Go stimuli for a longer duration, compared to unchosen NoGo stimuli (Bakkour et al., 2016; Schonberg et al., 2014).

For some participants, eye-tracking data were not collected or analyzed, mainly due to suboptimal gaze-recording conditions (e.g., participants wearing light-reflecting MRI compatible glasses). Data preprocessing included identification of gaze within the screen regions where stimuli were presented, as well as removal of corrupted data due to poor calibration and blinks. We defined rectangular ROI with a buffer zone around the screen *x* and *y* coordinates of the stimuli presented as well as central fixation cross. The screen ROI were increased by additional 20% of the stimulus width and height around the *x* and *y* axes, respectively (10% additional pixels on each direction). Gaze recorded outside the ROI was treated as invalid data. To exclude data corrupted by blinks, 150 ms time-windows before each blink onset and after each blink offset were also scrubbed before the analysis. The analysis included only the time from stimulus onset to stimulus offset, removing ISI.

We included in the analyses only scans in which the participants fixated on average on the ROI (fixation cross and/or stimuli) for at least 50% of the total time of all trials of the scan. Participants with 50% or more unrecorded or excluded scans (i.e., 1, 4, and 2 scans of the passive viewing, training, and probe task, respectively) were excluded from the task analysis. Results remain consistent when these participants are included in the analyses. See Table 1 for a summary of the sample sizes and excluded data in all tasks.

2.8 | Imaging data analysis

2.8.1 | MRI preprocessing

Raw DICOM format imaging data were converted to NIfTI with dcm2nii tool. The NIfTI files were organized according to the BIDS format v1.0.1 (Gorgolewski et al., 2016). Preprocessing of the functional imaging data was performed using fMRIprep version 1.0.0-rc2 (Esteban et al., 2019), a Nipype-based tool. Each T1 weighted volume was corrected for bias field using N4BiasFieldCorrection v2.1.0 and

TABLE 1 Eye-tracking analyses sample sizes

Session	Task	Valid participants (excluded[a])	Valid scans (excluded[b])
First session	Training (8 scans)	*n* = 32 (1)	234 (22)
	Passive viewing (2 scans)	*n* = 26 (6)	52 (0)
	Probe (4 scans)	*n* = 29 (2)	109 (7)
Follow-up	Passive viewing (1 scan)	*n* = 14 (1)	14 (0)
	Probe (4 scans)	*n* = 15 (1)	56 (4)

[a]Not counting participants for which eye-tracking data were not recorded.
[b]Scans with less than 50% valid data were not included in the analysis.

skull stripped using antsBrainExtraction.sh v2.1.0 (using OASIS template). Cortical surface was estimated using FreeSurfer v6.0.0 (Dale, Fischl, & Sereno, 1999). Skull-stripped T1 weighted volumes were coregistered to skull stripped ICBM 152 Nonlinear template version 2009c using nonlinear transformation implemented in ANTs v2.1.0 (Avants, Epstein, Grossman, & Gee, 2008). Functional data were motion corrected using FSL MCFLIRT v5.0.9 (Smith et al., 2004). This was followed by co-registration to the corresponding T1 weighted volume using boundary-based registration with nine degrees of freedom, implemented in FreeSurfer v6.0.0. Motion correcting transformations, T1 weighted transformation and MNI template warp were applied in a single step using antsApplyTransformations v2.1.0 with Lanczos interpolation. Three tissue classes were extracted from the T1 weighted images using FSL FAST v5.0.9. Voxels from cerebrospinal fluid and white matter were used to create a mask which was used to extract physiological noise regressors using aCompCor. The mask was eroded and limited to subcortical regions to limit overlap with gray matter, and six principal components were estimated. Framewise displacements were calculated for each functional run using a Nipype implementation. For more details of the pipeline using fMRIprep see http://fmriprep.readthedocs.io/en/1.0.0-rc2/workflows.html.

For each scan, we created a motion confound file containing nine regressors: six motion parameters (translational and rotation, each in three directions), a regressor for the *SD* of the root mean squared intensity difference from one volume to the next (DVARS), absolute DVARS values, and voxelwise *SD* of DVARS values. Volumes with excessive head movement (predetermined as framewise-displacement value larger than 0.9 mm) were scrubbed by adding an additional regressor for each volume to be removed. These nine regressors (or more, if included volume scrubbing procedure) were added to each first-level analysis.

2.8.2 | Face region of interest analysis

To examine the involvement of specialized face-processing brain regions, we used an ROI analysis approach to examine three preregistered regions comprising the central part of the face-selective network (Fox et al., 2009; Kanwisher et al., 1997; Yovel & Kanwisher, 2004): the fusiform face area (FFA), occipital face area (OFA), and the posterior division of the superior temporal sulcus (pSTS), separately for each hemisphere. These face-selective regions were identified in MNI space using an independent localizer task contrasting response to video-clips of dynamic faces versus objects (see Section 2.4). We were able to identify face-selective response in the right FFA in 95.24% of participants (n = 40; M_{size} = 126.33 voxels), left FFA in 95.24% (n = 40; M_{size} = 100.00 voxels), right OFA in 59.52% (n = 25; M_{size} = 62.64 voxels), left OFA in 61.90% (n = 26; M_{size} = 53.19 voxels), right pSTS in 95.24% (n = 40; M_{size} = 146.48 voxels), and left pSTS in 78.57% (n = 33; M_{size} = 161.88 voxels) of all 42 valid participants. For each region, we created a binary mask which was used to extract its mean percent signal change (Mumford, 2007). As in the general GLM model, we examined both the group mean as well as the correlation with CAT behavioral effect. Each of the six ROI analyses included data from different subset of participants out of the total N = 42, according to the number of participants in which we were able to localize the ROIs. To account for multiple comparisons, we report Bonferroni-corrected results exceeding p = .0083 as significant (α = .05, corrected for six multiple comparisons).

2.8.3 | Face network support vector machine analysis

Changes at the face processing network were also examined with an additional exploratory support vector machine (SVM) analysis, using the same functionally defined ROIs. The percent signal change of the voxels within each region were averaged, and then used as features in an SVM classification model (each region's mean as a single feature). Participants were median-split according to their CAT behavioral effect (proportion Go stimuli were chosen during probe; below-median group: M = 47.88%, SD = 6.27%, range = 32.64%—56.34%; above-median group: M = 65.48%, SD = 8.12%, range = 56.64%—83.22%), ROI data were z-scored and used as features in the SVM model. Since we were not able to functionally identify the OFA for approximately 40% of participants, we decided to use only FFA and pSTS data as features in the SVM model. For the few participants where either FFA or pSTS were not identified using the functional localizer, the missing feature data were replaced with the group mean value in the SVM model. Using a leave-one-participant-out cross validation (CV) procedure, the model was trained to classify between above-median and below-median participants using data from all participants but one, and the model's accuracy was tested using the last participant, not included in the training of the model. The CV performance of the model was evaluated using a permutation test. The model's accuracy was compared to that of 5,000 randomly permutated models (where participant labels were randomly shuffled) to produce a p value.

2.8.4 | fMRI univariate analysis

FSL's FEAT (FMRIB expert analysis tool; Smith et al., 2004) was used to design a general linear model (GLM) analysis of the fMRI data. In a first level analysis, BOLD response was modeled by convolving each task's regressors (except for motion confound regressors, described above) with a canonical hemodynamic response function (HRF). For each convolved regressor, we included the temporal derivative in the first-level GLM model. Following the first-level analysis, in a second-level fixed effects model, the scans of each individual participant were averaged or contrasted, differently in each task. Finally, in a mixed effects (FLAME 1) group analysis, we analyzed the mean group effect, as well as a correlation analysis between each lower-level contrast and the corresponding behavioral effect measured during the probe phase. Contrasts of high-value stimuli were correlated with the proportion of trials each participant chose the high-value Go stimuli during probe; low-value contrasts were correlated with the proportion of trials each participant chose the low-value Go stimuli during probe; contrast involving both high- and low-value stimuli were correlated

with the mean proportion of trials participants chose any Go stimulus (high or low). A *general* analysis plan with predefined contrasts and ROI was recorded at OSF (https://osf.io/89632/), deviations from this analysis plan are further detailed below at the end of the methods section.

Collecting a relatively large sample of participants with greater power to model behavior variability, enabled us to test the group-level correlation analysis of the CAT behavioral change effect with functional neural response. Thus, we aimed to find activity correlated with stronger CAT effects between participants, that is, which regions showed enhanced activity for participants who demonstrated a stronger preference for Go stimuli. However, considering the correlation analysis requires a large sample size to be properly interpreted (Yarkoni, 2009), we avoided using correlation analysis in the relatively small subsample we had in the follow-up session (n = 25; min. r = .583 detectable with 80% power, α = .01).

Passive viewing

The first-level GLM of the passive viewing task included a total of 13 regressors (excluding the motion confound regressors), similar to the ones used in recent work (Botvinik-Nezer et al., 2019): four regressors for different stimuli of interest (high-value Go, high-value NoGo, low-value Go, and low-value NoGo stimuli), the same four regressors with the same onsets and duration, but with a parametric modulation of the mean-centered proportion of trials the stimulus was chosen during the probe phase, four regressors for stimuli of no interest (high-value "training fillers", high-value "probe sanity", low-value "training fillers", and low-value "probe sanity"), and a final additional regressor with a parametric modulation by the mean-centered initial subjective value (Colley score) of each stimulus, to account for initial-value confound. These 13 regressors were convolved with the canonical HRF and included in the GLM along with their temporal derivative and the motion regressors.

In a second-level analysis, we averaged the scans as well as contrasted the post-training scans with the baseline pretraining scans in order to model "after CAT > before CAT" differences in BOLD response. The follow-up scans were similarly analyzed, contrasting the follow-up scan with the first pre-CAT baseline scan.

Training

The first-level GLM of the training task included a total of 18 regressors (excluding motion), similarly to the training analysis described by Schonberg et al. (2014). Go trials were modeled by 10 regressors: four regressors for high-value Go trials (unmodulated, modulated by proportion of probe choices, modulated by initial value and modulated by GSD; all modulations were mean-centered), similar four regressors were used for low-value Go trials, one additional regressor included onsets and duration of all Go trials modulated by the mean centered reaction time and one regressor modeled missed Go trials (Go trials in which participants failed to respond at all). Eight regressors modeled NoGo trials: three regressors for high-value NoGo trials (unmodulated; modulated by proportion of probe choices; modulated by initial value), similar three regressors were used for the low-value NoGo trials, an additional regressors modeled all NoGo trials of sanity and "fillers" stimuli and the last regressor modeled rare NoGo trials, in which an erroneous button press response was made. In the second-level analysis, we contrasted each scan on its own. Focusing on the first and last scan, the two ends of the training session were contrasted in a "last scan > first scan" contrast as well as in a linear trend weighting the eight scans from earliest to latest, as described in the general analysis preregistration (https://osf.io/89632/).

Probe

In the probe task, 16 regressors were used in the GLM model, as done in previous work (Botvinik-Nezer et al., 2019; Schonberg et al., 2014). In each regressor, trial duration was set to 956 ms, which was the average trial-duration across all trials of all participants. Four trial categories were generated based on the initial value of the two stimuli in the probe pair, and the outcome chosen by the participant (high-value Go, high-value NoGo, low-value Go, and low-value NoGo was chosen). Each of these four trial categories was modeled by three regressors—unmodulated, modulated by the proportion of times the stimulus was selected throughout the entire task; modulated by the difference in initial value between the two stimuli (all modulations were mean-centered); thus, resulting in 12 regressors. Two regressors modeled all high-value and all low-value trials (one regressor for each category) modulated by the reaction time, one regressor modeled sanity trials and one regressor modeled missed trials in which participants failed to respond within the allocated 1,500 ms time frame. Second-level analysis of the probe averaged the four scans of the probe task.

Since regressors were based on participants' responses, in some cases this resulted in a rank deficient design matrix due to an empty regressor of interest (e.g., one participant did not choose the NoGo stimuli in any of the low-value probe trial and was therefore left with an empty low-value NoGo regressor) or a zeroed-out modulation by choice (in case only one choice was made, mean-centered modulation resulted in a modulation column of all-zeros). Scans with empty regressors were excluded from the second-level analysis. Since several modulated NoGo regressors were zeroed out, we decided not to exclude scans in these cases, but rather not focus on these regressors in any further analysis, as was done in a previous work with similar challenges (Botvinik-Nezer et al., 2019; Schonberg et al., 2014). As a result, one participant was excluded from the analysis, and three additional participants had one of their four probe scans excluded. In the second-level analysis, all four probe scans (or three remaining scans in the case of three participants) were averaged.

2.8.5 | Contrasts of interest

We defined several contrasts of interest as our main analyses. In the Section 3 below, we report all statistically significant findings for these contrasts. If one or more of these contrasts are not reported, this is an indication that the contrasts had been examined but yielded no statistically significant results. Our contrasts of interest were based on insights from previous behavioral data (Salomon et al., 2018). Previous imaging studies of CAT with snack-food stimuli focused on high-value stimuli contrasts, as significant behavioral effects were observed

mainly for these high-value stimuli (Bakkour et al., 2017; Botvinik-Nezer et al., 2019; Schonberg et al., 2014). In the preregistration of the current work, based on previous research with CAT and faces (Salomon et al., 2018), we expected that CAT will induce a significant effect both for high-value and low-value stimuli. Therefore, in the current work we focused on three first-level contrasts of interest pooling together both value categories stimuli (termed "*All*" contrasts): (a) *All Go stimuli* contrast—representing the mean response to both high- and low-value Go stimuli; (b) *All Go minus NoGo stimuli* contrast—representing regions with stronger mean response to Go than NoGo stimuli (as in the previous contrast, pooling across both high- and low-value stimuli); (c) *All Go: modulated by choice* contrast—representing regions which activity correlated with the proportion of trials a stimulus was chosen by the individual participant scanned (within scan correlation of BOLD with behavioral effect).

The first contrast, *All Go stimuli*, was examined using a fixed-effects model (second level in FSL) across different time points in the passive viewing and training task. In the passive viewing task, post-training scans were contrasted with the pretraining baseline scans, and 1-month follow-up scans were also contrasted with baseline. In the training task, the last scan of training was contrasted with the first scan, and early training scans were contrasted with late training scans in a linear trend contrast. In a group-level random effect analysis (third level in FSL), the *All Go stimuli* contrast was tested for the mean group-effect. It was also tested in a correlation analysis with the CAT behavioral effect in all three tasks—that is, examining in which brain regions participants with stronger BOLD response to Go stimuli had also demonstrated greater CAT effect (stronger preference for Go stimuli).

The second contrast, *All Go minus NoGo stimuli*, was similarly tested in the passive viewing and training tasks in a fixed-effects analyses comparing different time points (pretraining to post-training and early training to late training in the passive viewing and training task, respectively), as well as in a group-level correlation with the behavioral change analysis in all three tasks. In addition, the *All Go minus NoGo* contrast was also tested without second-level (fixed-effects) contrast in the probe and the post-training passive viewing task, to examine which brain regions responded stronger to Go versus NoGo stimuli within the task.

The third contrast of interest, *All Go: modulated by choice*, similarly to the second contrast of interest, was also tested with fixed-effects contrasts comparing between time-points in the passive viewing (pre-training to post-training) and training tasks (early training to late training), as well as without between-scans contrast in the probe and the post-training passive viewing task. Unlike the previous contrasts of interest, it was not examined in group-level correlation analysis with CAT effect, due to the ambiguity of this contrast's interpretation (i.e., a correlation effect within scan showing correlation pattern with behavioral effect across participants).

2.8.6 | Anatomically defined regions of interest

Our analyses focused on the role of four preregistered anatomically defined ROIs: (a) the vmPFC, which has been implicated in valuation (Bartra, McGuire, & Kable, 2013) and importantly also in previous imaging studies with CAT (Bakkour et al., 2017; Schonberg et al., 2014); (b) the superior parietal lobule (SPL), which is associated with attentional mechanisms (Alho, Salmi, Koistinen, Salonen, & Rinne, 2015; Shomstein & Yantis, 2006) and has been recently found to be related to CAT (Botvinik-Nezer et al., 2019); (c) the striatum, implicated in reward and reinforcement-based learning (O'Doherty, 2004; O'Doherty et al., 2004); and (d) the hippocampus, as we hypothesized that memory processes will be an important factor in the maintenance of CAT (Botvinik-Nezer et al., 2019; Wimmer & Shohamy, 2012).

These four ROIs were preregistered prior to data analysis (https://osf.io/uhk4u; see deviations from preregistration detailed in following section). Following the whole-brain analysis, we performed an additional small-volume corrected (SVC) analysis for these four ROIs, as was done in previous work with CAT (Bakkour et al., 2017; Botvinik-Nezer et al., 2019; Schonberg et al., 2014). In the SVC analysis, anatomical masks for each ROI were defined based on the Harvard-Oxford structural atlas and used to identify clusters within the ROI. We report in the text all contrasts of interest where significant clusters were found with SVC. In contrasts where a whole brain analysis revealed significant results in the ROI, we do not report an additional result using SVC (as these are trivial). All SVC results were corrected for four multiple comparisons using Bonferroni correction. We report clusters exceeding Bonferroni corrected $\alpha = .0125$ as statistically significant, along with corrected p values (original p multiplied by four). Additional relevant regions (Bartra et al., 2013; Smith et al., 2014) were not selected as ROIs. This was done to maintain detection power, as using additional ROIs would require more stringent statistical threshold correction for each region. Nonetheless, we examine and report all significant results in our whole-brain analyses.

2.9 | Functional connectivity: gPPI analysis

To examine the functional dynamics of CAT, we performed an additional generalized psychophysiological interaction (gPPI) analysis (McLaren, Ries, Xu, & Johnson, 2012) for the training task. The analysis examined task-related functional connectivity with key ROI seed regions during the training task—specifically, connectivity with the seeds' response to Go stimuli during training in contrast to the response to NoGo stimuli. Seeds were selected based on significant clusters found in the whole-brain analysis. A 5-mm sphere was defined around the peak activation voxel of each contrast, and was then masked by the original activation cluster, to only include voxels appearing in the original activation cluster. For each seed, an independent model was created. The seed's neural response to Go and NoGo stimuli during the training task was estimated by deconvolving the mean BOLD signal of the voxels in the seed ROI (Gitelman, Penny, Ashburner, & Friston, 2003).

The first-level gPPI model included 21 regressors: one regressor indicating the estimated seed's response to Go stimuli; one regressor indicating the estimated seed's response to NoGo stimuli, one regressor with the mean time series of the seed voxels and all 18 regressors

used in the GLM analysis of the training task, detailed above. Functional connectivity was evaluated by contrasting Go PPI regressor with NoGo PPI regressor. Second-level and group analyses were identical to those described in the GLM analysis.

2.10 | Deviations from preregistration

Prior to the completion of data collection and before any statistical analyses were performed, we preregistered our experimental procedure, sample size, expected results, and a *general* analysis plan for the imaging data. Although we aimed this preregistration to depict as clearly and extensively as possible our analysis plan, some important details were not well-specified. In this section, we highlight important ambiguities and differences between our preregistration and the final methods used in this work.

The first deviation from preregistration regards the use of prehypothesized ROI. In the preregistration, we specified the anatomically defined ROIs (vmPFC, striatum, SPL, and hippocampus) together with the functionally defined ROIs (FFA, OFA, and pSTS). However, we unintentionally neglected to specify how these ROI will be used in analysis. Anatomical ROIs were only intended to be used in SVC analysis, as done in previous work (Bakkour et al., 2017; Botvinik-Nezer et al., 2019; Schonberg et al., 2014). Functionally defined ROI data were transformed to percent signal change and averaged to be used in the reported GLM ROI analysis and an exploratory (unregistered) SVM model. To limit our false discovery rate (Gildersleeve & Loken, 2013), we did not perform any additional exploratory analyses using face ROIs. The data are openly shared to be analyzed by others who might be interested in using different approaches. Another ROI deviation was done with the definition of the parietal ROI—while in the preregistration we intended to use both SPL and intraparietal sulcus (IPS), in the final analysis we did not use the IPS, as we realized this region is not well-defined in the Harvard-Oxford cortical structural atlas.

Second, in our analysis plan (https://osf.io/8yhzr/), we mentioned that group analysis will be done with FSL Randomize as our first option, based on nonparametric permutation testing, and FLAME-1 as a second choice in case Randomize will not yield results (Eklund, Nichols, & Knutsson, 2016). Eventually, we decided to use FLAME-1 which provides valuable modeling of intersubject variability and may be more suitable for our design and sample size. The detailing of fMRI models also lacks in the preregistration. While the general outline of analysis is described for each task, we failed to clearly define the regressors and contrasts in a satisfactory manner. These contrasts of interest described in the current paper were conceptualized prior to data analysis based on previous imaging work with CAT (Bakkour et al., 2017; Botvinik-Nezer et al., 2019; Schonberg et al., 2014), but some of them had not been well-documented in the preregistration. No other unreported analysis models were used.

Finally, in our preregistration, we mentioned we will perform an analysis of the data from the memory behavioral task as well as a multivariate pattern analysis (MVPA) of the imaging data. However, we eventually decided that both memory and MVPA are beyond the scope of the current article and did not perform these analyses.

3 | RESULTS

3.1 | Behavioral

In the first session of the probe task, participants showed enhanced preference for the high-value Go stimuli over high-value NoGo stimuli (prop. = 57.86%, OR = 1.40, 95% CI = [1.18, 1.66], $p = 4.4E^{-5}$, one-sided logistic regression; see Figure 2), as well as for the low-value Go stimuli over low-value NoGo stimuli (prop. = 55.50%, OR = 1.26, 95% CI = [1.07, 1.50], $p = .003$, one-sided logistic regression). The effect was consistent when pooling together trials from both categories (All Go, OR = 1.33, 95% CI = [1.15, 1.54], $p = 8.3E^{-5}$, one-sided logistic regression), with a marginal trend of stronger effect for high-value probe choices (OR = 1.11, 95% CI = [0.99, 1.23], $p = .06$, two-sided logistic regression). In the 1-month follow-up session, enhanced preference for high-value Go stimuli was maintained for the high-value pairs (prop. = 54.78%, OR = 1.23, 95% CI = [0.97, 1.57], $p = .046$, one-sided logistic regression) with a trend for low-value pairs (prop. = 53.13%, OR = 1.13, 95% CI = [0.97, 1.34], $p = .059$, one-sided logistic regression). The effect was consistent when pooling together trials from both categories (All Go, OR = 1.18, 95% CI = [1.01, 1.37], $p = .016$, one-sided logistic regression), with no differential effect found for high- versus low-value pairs (OR = 1.07, 95% CI = [0.94, 1.22], $p = .31$, two-sided logistic regression).

Examining the preference for high-value over low-value stimuli in the sanity-check trials, we found that participants consistently

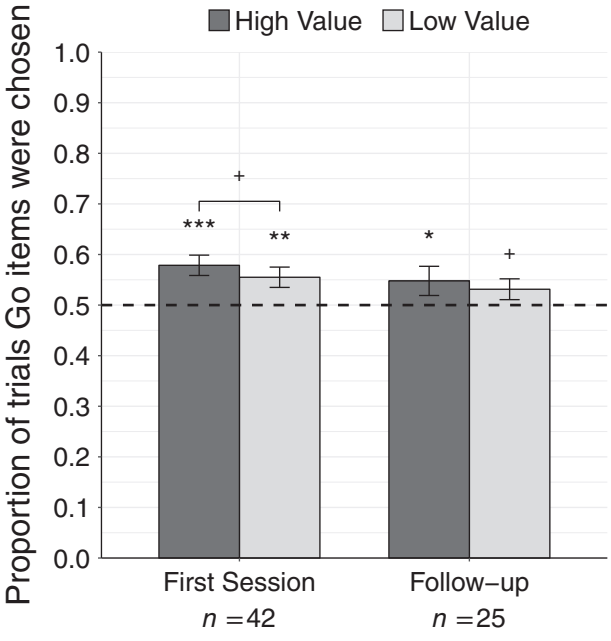

FIGURE 2 Probe results. Mean proportion of trials participants chose Go over NoGo face stimuli, for high-value (dark gray) and low-value (light gray) probe pairs. Dashed line indicates 50% chance level, error bars represent standard error of the mean. Asterisks above each bar reflect statistical significance in a one-sided logistic regression analysis indicating proportions different from chance (log-odds = 0, odds-ratio = 1); Asterisks between bars indicate differential effect for value categories (two-sided logistic regression); ***$p < .001$, **$p < .01$, *$p < .05$, +$p < .1$

preferred stimuli of initial high-value, both in the first session (prop. = 89.25%, OR = 12.83, 95% CI = [6.41, 25.67], $p = 2.7E^{-13}$, one-sided logistic regression), as well as in the 1-month follow-up session (prop. = 86.36%, OR = 15.58, 95% CI = [5.11, 46.81], $p = 6.1E^{-7}$, one-sided logistic regression).

3.2 | Eye-tracking

3.2.1 | Passive viewing task

In the passive viewing task, before CAT, participants gazed the Go and NoGo stimuli for a similar duration of the 2-s presentation (M_{Go} = 84.32%, M_{NoGo} = 84.56%, t[25] = −0.192, p = .849, two-sided linear regression). After CAT, participants maintained similar gaze duration for Go and NoGo stimuli (M_{Go} = 76.70%, M_{NoGo} = 78.78%, t[25] = −1.09, p = .284, two-sided linear regression). Similarly, in the 1-month follow-up session, we did not observe a difference in gaze time of Go versus NoGo stimuli (M_{Go} = 80.70%, M_{NoGo} = 81.71%, t[13] = −0.82, p = .423, two-sided linear regression).

3.2.2 | Training

During the eight training scans (16 CAT runs), participants fixated on the Go stimuli for a mean duration of 88.3% of the 1 s presentation and the NoGo stimuli for a mean of 87.4%. Modeling the proportion of time participants viewed the stimuli, using Go/NoGo and initial subjective value (high/low) as factors, revealed no difference between Go and NoGo stimuli gaze duration (t[31] = 1.45, p = .156, two-sided linear regression) and no differential effect of value was found (t[31] = 0.874, p = .389, two-sided linear regression).

In an additional exploratory analysis, we examined the effect of time across training (modeled as scan number) on gaze pattern. We observed an increasingly growing gaze-bias pattern between Go and NoGo—that is, while in early scans of training no difference was observed in gaze duration for Go (M_{scan1} = 90.29%) versus NoGo (M_{scan1} = 91.44%) stimuli, the mean difference increased over time to a maximal difference in the last scan with longer gaze duration for Go (M_{scan8} = 89.47%) compared to NoGo stimuli (M_{scan8} = 83.97%; linear trend effect: t[31.13] = 4.99, $p = 2.2E^{-5}$, two-sided mixed linear regression). We examined whether this gaze-bias observed in the last training scan (measured as the difference in relative time participants viewed the Go vs. NoGo stimuli) correlated with the effect of CAT as measured in the probe phase. We did not find a significant correlation between gaze-bias and the CAT effect (r = .141, t[57.83] = 1.19, p = .238, two-sided mixed linear regression).

3.2.3 | Probe

In the probe task, we found that participants in the first session overall viewed the selected stimuli more than the unchosen stimuli, as expected in a binary choice task (Krajbich, Armel, & Rangel, 2010). When Go stimuli were chosen, participants fixated on the Go stimuli for a longer duration (M_{Go} = 42.36%, M_{NoGo} = 31.67%, t[28.20] = 9.89, $p = 1.1E^{-10}$, two-sided mixed linear regression) and similarly, when NoGo stimuli were chosen, participants fixated on the NoGo stimuli for a longer duration (M_{Go} = 31.25%, M_{NoGo} = 43.40%, t[28.00] = −10.91, $p = 1.4E^{-11}$, two-sided mixed linear regression).

Participants did not view Go stimuli more compared to NoGo stimuli, neither when comparing the gaze time of the stimuli when they were chosen ($M_{Go/Chosen}$ = 42.36%, $M_{NoGo/Chosen}$ = 43.40%, t[28.90] = −0.83, p = .412, two-sided mixed linear regression), nor when comparing the time when the stimuli were not chosen ($M_{Go/NotChosen}$ = 31.25%, $M_{NoGo/NotChosen}$ = 31.67%, t[29.28] = −0.26, p = .794, two-sided mixed linear regression). Adding the value category (high vs. low) to the regression model had no significant effect (t[27.80] = 0.30, p = .764, two-sided mixed linear regression), indicating the gaze pattern was consistent across all probe trials.

In the 1-month follow-up session, participants demonstrated similar patterns. Participants gazed the chosen stimuli more than the unchosen stimuli, with no differences between Go and NoGo stimuli ($M_{Go/NotChosen}$ = 28.81%, $M_{NoGo/NotChosen}$ = 28.24%, t(57.97) = −0.76, p = .451, two-sided mixed linear regression [close to singular fit]; $M_{Go/Chosen}$ = 38.35%, $M_{NoGo/Chosen}$ = 38.87%, t(13.71) = −0.37, p = .717, two-sided mixed linear regression).

3.3 | Imaging

In the fMRI analyses, we focused on the BOLD response for Go stimuli in order to study the neural mechanisms underlying the preference change induced by CAT. We first present the results focusing on the face-selective network (ROI analysis and exploratory SVM modeling), followed by a whole-brain and SVC univariate GLM analyses for the three tasks scanned (training, passive viewing, and probe), and finally, we present results of the gPPI connectivity analysis for the training task.

3.3.1 | Face regions ROI analysis

An independent face localizer task was used to functionally identify, for each participant, three ROIs of face selective brain regions: FFA (right FFA and left FFA, both identified in 95.24% of participants), OFA (right OFA identified in 59.52% and left OFA in 61.90% of participants) and pSTS (right pSTS identified in 95.24% and left pSTS in 78.57% of all 42 valid participants). For each identified ROI, we extracted the mean response of all voxels within the ROI, and examined the response across the group in our three predefined contrasts: (a) *All Go stimuli*, (b) *All Go > NoGo stimuli*, and (c) *All Go stimuli modulated by choice*. In all three tasks, we were unable to identify stronger responses of face selective regions to Go versus NoGo face stimuli. A single exception was observed in the last scan of the training task, where a stronger response to Go versus NoGo trials was observed in the right pSTS (M_{diff} = 0.012, SE = 0.003, t[39] = 4.22, $p = 1.42E^{-4}$, $p_{corrected} = 8.5E^{-4}$) and the left pSTS ($M_{diff}$ = 0.012,

SE = 0.003, *t*[32] = 3.74, $p = 7.23E^{-4}$, $p_{corrected} = .004$). However, this result was not unique to the late stages of training, and was apparent also in the first scan (pSTS right: $M_{diff} = 0.017$, *SE* = 0.002, *t*(39) = 6.82, $p = 3.83E^{-8}$, $p_{corrected} = 2.30E^{-7}$; pSTS left: $M_{diff} = 0.018$, *SE* = 0.003, *t*(32) = 6.47, $p = 2.78E^{-7}$, $p_{corrected} = 1.67E^{-6}$). In all three tasks (training, passive viewing, and probe), we did not find a significant correlation between the CAT behavioral effect and mean percent signal change in participants' face selective regions. None of the models based on functionally defined face selective ROIs exceeded in statistical significance the threshold of (Bonferroni corrected) $p < .05$.

3.3.2 | Face regions SVM analysis

In an exploratory analysis, we used an SVM to explore whether information stored within face processing network (only FFA and pSTS, see Section 2.8.3) could classify participants according to their consequent behavioral change following CAT, that is, differ participant with above-median behavioral change effect from participants with below-median effect. One SVM model using the FFA and pSTS mean response to all Go stimuli within the last training run (a contrast also presented below in the univariate GLM analysis of the training task and in Figure 3a), was able to accurately classify 66.67% of participants in a leave-one-participant-out CV test (true-positive = 21.43%, true-negative = 45.24%, false-positive = 4.76%, false-negative = 28.57%; see Figure S2). In a permutation test with 5,000 permutations, the SVM model performance was found marginally within the top 5% of models ($p = .049$, one-sided permutation test), resulting in alternating significance conclusion, depending on the number of permutations and randomization seed. Using the ROI response to Go stimuli in the probe and passive viewing task, did not result in classification models exceeding statistical significance of $p < .05$.

3.3.3 | Cue-approach training

In the training task, we analyzed the BOLD response during eight scans, each with two runs (two repetitions of all training stimuli).

Examining the correlation between the BOLD response to Go stimuli during the last scan revealed an association between CAT behavioral effect and BOLD activity of the left ventral striatum, mainly putamen and nucleus accumbens—that is, participants who chose the Go over the NoGo stimuli more during the subsequent probe were characterized with stronger response in the left ventral striatum to Go stimuli during the last training scan (cluster size = 154 voxels, max *Z*-value = 3.76, $p = .042$; Figure 3a). Furthermore, using a striatal SVC analysis, a similar marginal trend was also observed within the contralateral head of caudate nucleus and nucleus accumbens in response to Go minus NoGo stimuli contrast, though after Bonferroni correction these did not exceed statistical significance (cluster size = 67 voxels, max *Z*-value = 3.24, $p = .023$, $p_{corrected} = .092$). These effects were unique to Go stimuli, and not found for NoGo training trials.

When contrasting the last scan with the first scan, no region showed an increase in activation to Go over NoGo stimuli. However, examining the same contrast (Go stimuli > NoGo stimuli [last scan > first scan]) in correlation with the probe effect, showed a positive correlation between the BOLD signal and behavioral change effect, mainly in the premotor regions in the posterior dorsomedial frontal cortex (cluster size = 257 voxels, max *Z*-value = 4.04, $p = .001$; Figure 3b).

Our behavioral eye-tracking analysis revealed that in the last training scan, participants exhibited a gaze-bias effect, fixating more on Go compared to NoGo stimuli (although this did not correlate with behavior change). To further explore this finding, we performed an additional exploratory correlation analysis of fMRI with gaze-bias. The analysis was similar to the correlation with the CAT effect analysis, except in this case, BOLD signal was correlated with the gaze-bias of the last scan (measured as the mean proportion of time a participant viewed the Go stimuli minus that of NoGo stimuli). We found a correlation between gaze-bias

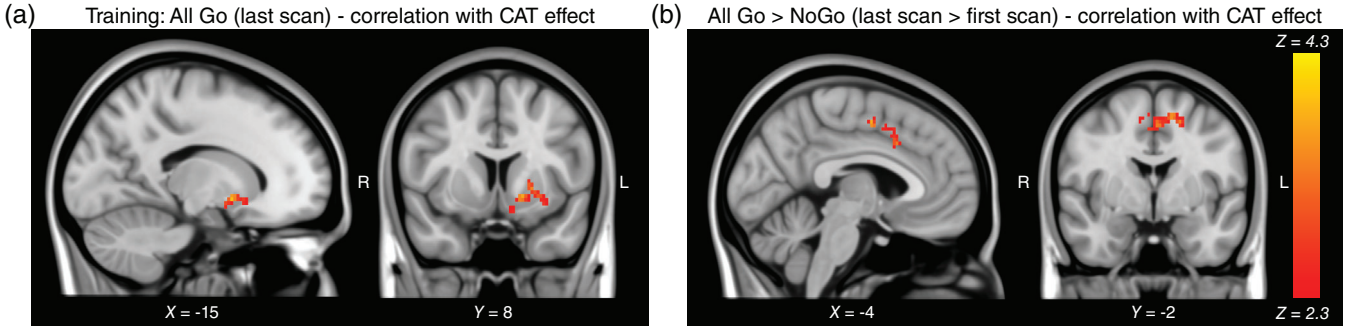

FIGURE 3 Training imaging results—correlation analysis. (a) BOLD signal during the last scan of training task correlated with the cue-approach training (CAT) behavioral change effect, that is, the proportion of trials a participant chose the Go stimulus during subsequent probe phase. Striatal response to Go stimuli in putamen and nucleus accumbens correlated with CAT effect. (b) Increased BOLD response in premotor regions, during the last scan contrasted with first scan of training task, correlated with the CAT behavioral change effect. All results were cluster-corrected at a whole-brain level, $p < .05$. Coordinates reported in standard Montreal Neurological Institute (MNI) space. For a list of anatomical regions within each cluster, see Tables S3-1 and S3-2. Full cluster-corrected and unthresholded maps are available at: https://neurovault.org/collections/NGTUHMTF/images/132467/, https://neurovault.org/collections/NGTUHMTF/images/132468/, https://neurovault.org/collections/NGTUHMTF/images/132469/, and https://neurovault.org/collections/NGTUHMTF/images/132470/

and the BOLD activity of Go minus NoGo stimuli, in the anterior cingulate cortex (ACC; cluster size = 148 voxels, max *Z*-value = 4.06, p = .032) as well as cerebellum and lateral occipital cortex (see: Figure S3a). We also found a correlation between gaze-bias and the BOLD activity mainly within the vmPFC (cluster size = 210 voxels, max *Z*-value = 4.35, p = .003, see Figure S3b) and more dorsal medial PFC, that is,—participants with more positive vmPFC response to Go stimuli, also had larger gaze-bias for Go stimuli in the last training scan. This effect was unique to Go stimuli, and was not observed in NoGo trials.

3.3.4 | Passive viewing

In the passive viewing task, we analyzed participants' response to Go and NoGo stimuli in the absence of response or external cues in three timepoints: once before the training procedure, a second time after training and a third time after a mean period of 1 month. Comparing the response after training to pretraining baseline, revealed no regions with enhanced response to both high and low-value Go stimuli following training. In the 1-month follow-up session, Go stimuli invoked a stronger response in comparison to the pretraining scan within ventral regions of the lateral occipital cortex and fusiform gyrus as well as more dorsal lateral occipital cortex. However, these results were not unique to Go stimuli; similar overlapping regions were also identified when testing the same contrast (follow-up > before training) for NoGo stimuli.

In addition to the main contrasts, we also examined a regressor parametrically modulated by the initial subjective value (Colley score) of all stimuli in the task. Examining this regressor, did not reveal significant clusters which show stronger response to stimuli of higher initial subjective value.

3.3.5 | Probe

Overall, across the entire group, no brain region area showed consistently stronger response to trials where Go stimuli were chosen in comparison to trials where NoGo stimuli were chosen. However, we found that vmPFC activation to Go versus NoGo choices (beyond value category, as preregistered) positively correlated across participants with the proportion of trials each participant chose the Go stimuli in the probe; that is, participants who demonstrated stronger preference for Go stimuli also had stronger response within the vmPFC when choosing Go stimuli versus when choosing the NoGo stimuli (cluster size = 153 voxels, max *Z*-value = 4.07, p = .012; Figure 4). The same contrast revealed additional regions including precuneus, bilateral anterior temporal cortex and bilateral cerebellar cortex (see Table S5 for full statistical details).

In the 1-month follow-up probe, no region showed enhanced response to Go choices over NoGo, nor a correlation with CAT effect, across both value categories.

3.3.6 | Connectivity analysis using gPPI during training

In the gPPI analysis, we used the striatum activation found in the training task (Figure 3) and the vmPFC activation found in the probe task (Figure 4) as seed ROIs. Seed masks of 5-mm sphere of voxels within the activation clusters were defined and were evaluated for response to Go versus NoGo stimuli during training. We examined which brain regions demonstrated significantly larger association with the seeds' activation during Go trials versus NoGo trials. Using response of the vmPFC seeds, we found that several brain regions showed stronger functional connectivity, including a large bilateral region across the SPLs to precentral gyrus, middle, and posterior temporal sulcus (Figure 5a), as well as the right lateral occipital cortex, dorsolateral prefrontal cortex, and cerebellum (see detailed statistical description in Table S5-1). In the case of the striatum seed PPI, we found several regions where stronger connectivity correlated with CAT behavioral effect, including left lateral orbito-frontal cortex (cluster size = 322 voxels, max *Z*-value = 4.19, $p = 7.82E^{-5}$; Figure 5b), right frontal operculum, medial superior frontal gyrus, and cerebellum (see detailed statistical description in Table S5-2).

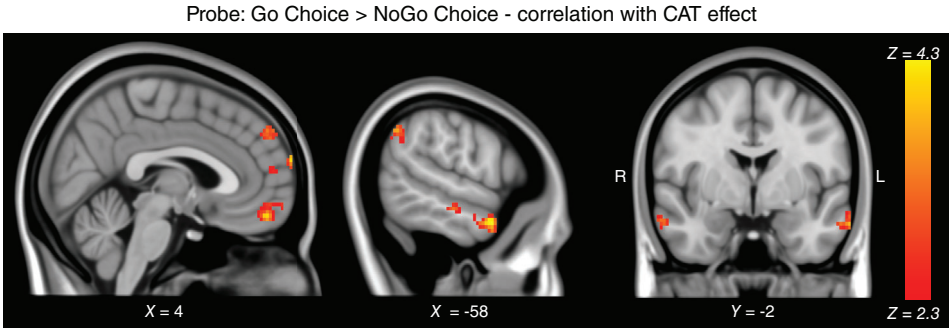

FIGURE 4 Probe task imaging results. Immediately after training, cue-approach training effect correlated with enhanced BOLD response to Go choices over NoGo choices, within several regions including the vmPFC and anterior middle temporal gyri. Results were cluster-corrected at a whole-brain level, p < .05. Coordinates reported in standard Montreal Neurological Institute (MNI) space. For a list of anatomical regions within each cluster, see Table S5. cluster-corrected and unthresholded maps are available at: https://neurovault.org/collections/NGTUHMTF/images/132471/ and https://neurovault.org/collections/NGTUHMTF/images/132472/

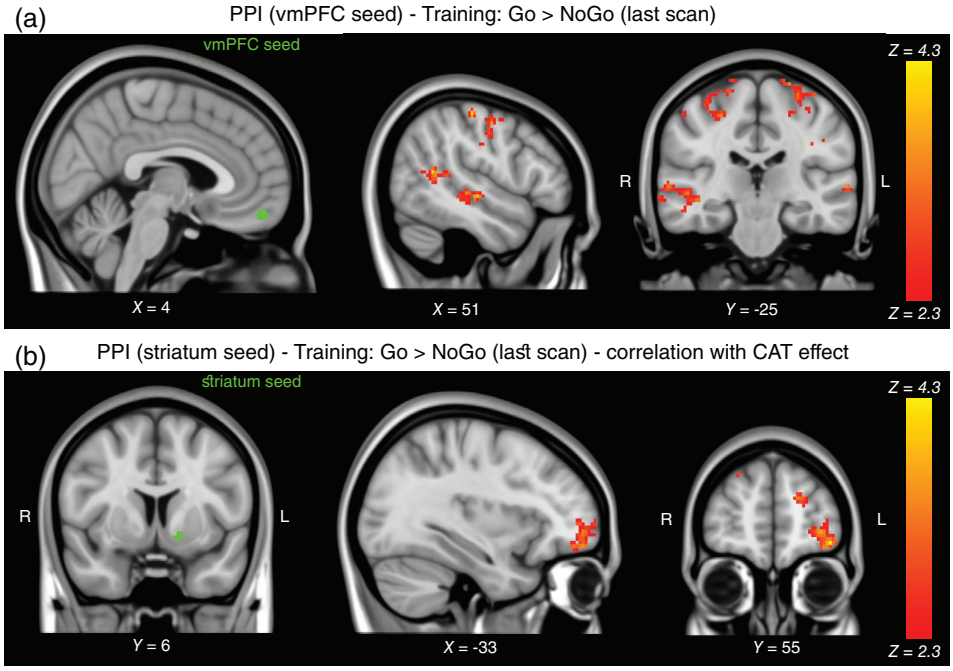

FIGURE 5 Training generalized psychophysiological interaction (gPPI) results. Regions showing greater connectivity with seed's activity to Go over NoGo trials. (a) A vmPFC seed (in green) showed in gPPI analysis stronger connectivity with regions in the superior temporal sulcus (STS) in response to Go versus NoGo stimuli during the last scan of training. (b) Connectivity of striatum seed (green) and orbito-frontal cortex during training, correlated with stronger behavioral effect in subsequent probe phase. Results were whole-brain cluster-corrected, $p < .05$. Coordinates reported in standard Montreal Neurological Institute (MNI) space. For a list of anatomical regions within each cluster, see Tables S5-1 and S5-2. Full maps are available at: https://neurovault.org/collections/NGTUHMTF/images/132473/, https://neurovault.org/collections/NGTUHMTF/images/132474/, https://neurovault.org/collections/NGTUHMTF/images/132475/, and https://neurovault.org/collections/NGTUHMTF/images/132476/

4 | DISCUSSION

The current work examined the neural mechanisms underlying nonreinforced behavioral change using CAT with face stimuli. Behaviorally, we found that CAT resulted in enhanced preference for the associated Go over NoGo stimuli, which was generally maintained 1 month after training. The current study is the first to link individual differences in nonreinforced behavior change following CAT with corresponding neural mechanisms using fMRI. We show a link between individual differences in behavior in the task with enhanced striatal activity during nonreinforced training, and enhanced vmPFC activity during binary-choices. We further show evidence for enhanced PFC connectivity with both the striatum and visual areas associated with nonreinforced preference modification.

The behavioral probe results are in line with the preregistered hypotheses, based on previous work with CAT (Bakkour et al., 2016, 2017; Botvinik-Nezer et al., 2019; Schonberg et al., 2014; Veling et al., 2017; Zoltak et al., 2018) and specifically with faces (Salomon et al., 2018). The behavioral effect of CAT on preferences is a group effect comprised of individual variability across participants, as highlighted in a recent study with prefrontal lesion patients (Aridan, Pelletier, Fellows, & Schonberg, 2019). Thus, our current work aimed to uncover both the common neural modifications across the entire group, as well as the neural signatures associated with individual differences in behavior, using face stimuli.

We selected face stimuli for the current fMRI study due to their unique characteristics, as one of the most prominent stimuli in the life of human beings. Modification of face-related preferences can be potentially developed into applicable means to treat various psychological disorders and undesired social biases, beyond the confinements of the lab (Browning et al., 2012; Meissner & Brigham, 2001; Stamarski & Son Hing, 2015). In addition, from a research perspective, face stimuli are selectively processed by a well-defined network of specialized regions, primarily the FFA, OFA, and pSTS (Fox et al., 2009; Kanwisher et al., 1997; Yovel & Kanwisher, 2004), as well as elicit robust and stable preference response (Aharon et al., 2001; Kranz & Ishai, 2006; Senior, 2003). We used both an ROI analysis approach examining the fMRI signal in these regions as well as an exploratory SVM approach. The ROI analyses did not yield significant results that implicate face-selective regions in nonreinforced behavioral change. During the training phase, an enhanced response of pSTS to Go over NoGo stimuli was observed. However, this effect was not unique to late training scans, thus indicating that enhanced response of pSTS during Go trials was likely due to its adjacency to auditory regions, which were strongly activated by the auditory cue during Go trials.

Using an exploratory SVM model with the response of face ROI in the last training scan, we were able to classify participants with

above-median behavioral-change effect, only slightly better than permuted null-models (CV accuracy = 66.67%, p = .049, one sided permutation test; see Figure S2). Considering the exploratory nature of this analysis and degrees of freedom in design, this finding should be appropriately taken with caution (Carp, 2012; Ioannidis, 2005). Additional replications with stringent statistical significance are needed to decisively conclude that face-selective ROIs play a significant role in nonreinforced behavioral change following CAT.

Interestingly, in a whole-brain analysis examining the neural response during training, we found for the first time a correlation between individual differences of preference modification following CAT and striatal activity, including mainly the left putamen and nucleus accumbens (Figure 3). Thus, participants with greater striatal responses to Go stimuli during the late training phases, also showed greater preference for Go stimuli in the subsequent probe phase. A similar correlation trend was observed in the contralateral striatum (right caudate nucleus and nucleus accumbens) using a striatal SVC analysis, when contrasting Go with NoGo trials (although the results did not exceed statistical significance following Bonferroni correction for the four ROIs tested with SVC).

These results show for the first-time evidence of striatal contribution to preference modification following nonreinforced training. This striatal activity, which is usually observed in reinforcement-learning paradigms (O'Doherty et al., 2004; Pessiglione et al., 2006; Rangel et al., 2008), suggests that even in the absence of external overt feedback or reinforcements, CAT induces preference modification via reinforcement-like mechanisms. Identifying individual differences in learning patterns across the participants' striatal response, that was observed in the current study, resonates with results of previous paradigms using monetary rewards (e.g., Schonberg, Daw, Joel, & O'Doherty, 2007).

It has been suggested that the conjunction of motor control along with reinforcement signals processing within the striatum, are a key feature in the striatum's role in learning (Collins & Frank, 2014; Tricomi, Delgado, & Fiez, 2004). These striatal signals correspond with previous behavioral findings that demonstrated the importance of both a rapid button press and a challenging cue in order to induce behavioral change with CAT (Bakkour et al., 2017). However, the association of a motor response with the cue poses limitations on the ability to disentangle the striatum's role in learning from its functionality in motor control. An alternative explanation could claim that the striatal response observed here, signifies a correlation of choice behavior with enhanced motor planning (Gerfen et al., 1990), also suggested by our result showing correlation with premotor frontal regions. As training progresses, for some participants the Go-cue becomes easier to predict and initiate motor planning processes, which might be in turn associated with both striatal signal and subsequent choices. However, we suggest that it is unlikely that the CAT effect on behavior depends purely on motor planning, as a previous behavioral study showed that training with the hand and choosing with the eyes during probe still yielded an enhanced preference effect (Bakkour et al., 2016). Future imaging studies could aim to replicate a similar effect by scanning training with an independent response module such as eye-movements, in order to differentiate the striatal motor role from its learning one.

Overall, we did not see a clear pattern of longer gaze for Go over NoGo stimuli in the probe and passive viewing task. However, in an exploratory analysis of the training task, we found a developing gaze-bias pattern. This pattern manifested in relatively longer fixations on the Go stimuli compared with NoGo stimuli in the last scans of the training task. However, this difference was not correlated with subsequent choices in the probe task. We further examined the neural correlates of this gaze-bias effect in the last training scan and revealed an interesting correlation with the response of ACC to Go over NoGo stimuli, as well as vmPFC response to Go stimuli; both regions are associated with valuation processes in general (Bartra et al., 2013), and specifically with valuation of face stimuli (Smith et al., 2014). These findings raise interesting hypotheses regarding the interaction of the brain valuation system and gaze pattern observed here. It could be that greater gaze-bias is formed due to a stronger value neural response (Krajbich et al., 2010; Krajbich & Rangel, 2011); although the causal direction might also be inverted, as stimuli attracting more attention and processing could evoke stronger value-related signals (Lim, O'Doherty, & Rangel, 2013). Nonetheless, it is important to note that this gaze-bias was not correlated with greater preference for Go stimuli in the subsequent probe phase. It is possible that a gaze bias develops prior to behavioral change and thus these neural modification in the value-processing brain regions, will later in be related to future preferences modification effect of CAT during probe. Considering the exploratory nature of this analysis, future studies could aim to replicate this effect, as well as examine the causal directionality, for example, by experimentally manipulating gaze and examining the involvement of value-processing brain regions.

We found a correlation between vmPFC activity during the binary choice probe and the individual differences in the CAT behavioral effect. Participants who chose the Go stimuli overall more, also demonstrated stronger vmPFC fMRI response for Go choices over NoGo probe choices. This finding resonates, though does not directly replicates, previous findings with CAT, that showed the enhanced vmPFC activity in the probe task was modulated by a parametric choice effect for specific Go stimuli; that is, vmPFC was more strongly activated for Go stimuli which were chosen more during the probe task (Bakkour et al., 2017; Schonberg et al., 2014). Our finding, in line with these previous results, suggests that the CAT effect on choices engages frontal decision-making neural mechanisms (Bartra et al., 2013; Clithero & Rangel, 2014; Kable & Glimcher, 2009; Rangel et al., 2008). Several differences between the current work and previously published imaging studies with CAT could potentially account for the divergence in imaging results of the current study. Primarily, all previous imaging studies were performed with snack-food items and in the probe phase, choices were made for actual consumption at the end of the experiment (Bakkour et al., 2017; Botvinik-Nezer et al., 2019; Schonberg et al., 2014). In contrast, in the current work, probe choices of preferred face stimuli had no actual consequences. Therefore, it is possible that the different choice context induced less robust within-participant value responses (Seymour & McClure, 2008). This hypothesis can also putatively account for the non-replicated eye-tracking findings, during the probe phase. While in previous CAT studies with snack-food items, unchosen Go stimuli attracted participants' gaze for longer duration compared to unchosen NoGo

stimuli during the probe phase (Bakkour et al., 2016; Schonberg et al., 2014), in the current work we did not find this effect. This hypothesis could be tested in future research by introducing additional features which would enhance the engagement of the participants with the choice procedure, such as introducing a payment method or realization of choices in a prospective part of the task (e.g., Smith et al., 2014).

In the same probe analysis, we found a correlation of the choice effect across participants with activity in middle temporal gyri that had been previously linked to high-level visual processing of faces (Winston, Henson, Fine-Goulden, & Dolan, 2004). These findings might point to the involvement of visual-processing regions, especially high-level processing ones along the STS, in co-operation with prefrontal value-processing regions. This finding is in line with recent findings of enhanced visual processing for Go stimuli during passive viewing, following CAT (Botvinik-Nezer et al., 2019).

In the passive viewing tasks, we did not find regions in which fMRI activity was modulated by the initial subjective value. It is possible that the design of the passive viewing task, with only one short presentation for each stimulus in each time point and relatively short ISI, was not sensitive enough to detect differences in values. In this study, we focused on the training and probe tasks and could not allocate more time for the passive viewing one. In order to increase the detection power, it could have been beneficial to use a different task design, for example, design with longer durations, greater number of repetitions, or a task that requires participants to respond to the value property of each stimulus, similarly to the designs used in previous work examining value processing in the brain (e.g., Aharon et al., 2001; Chib et al., 2009; Kranz & Ishai, 2006; Lebreton et al., 2009). Our follow-up session included a subsample of $n = 25$ participants (out of potential $N = 42$), as data collection was halted due to a scanner upgrade. The smaller follow-up sample size might have resulted in reduced power to detect more subtle effects (Button et al., 2013; Open Science Collaboration, 2015; Yarkoni, 2009). Future studies can address the question of long-term maintenance mechanisms, by conducting a larger-scale longitudinal experimental design.

Finally, in a gPPI analysis of the training task, we found task-related connectivity of the vmPFC seed with pSTS. This finding further resonates the probe finding linking vmPFC and high-level visual processing regions, as well as enhanced visual processing found previously for snacks with CAT (Botvinik-Nezer et al., 2019). These results also correspond with previous work, which had showed that greater connectivity of vmPFC with middle temporal regions was associated with value processing of faces (Smith et al., 2014). Our results could reflect more intense valuation processing (Serences & Yantis, 2006), further suggesting that strengthened visual-frontal associations play a role in integrating visual stimuli information with general value properties encoded in the vmPFC (Lim et al., 2013). The connectivity of the striatum seed with lateral orbitofrontal cortex (OFC) was found to be correlated with the preference effect across the sample. Participants with stronger striatum-OFC connectivity also had stronger preference for Go stimuli. Thus, this connectivity result further implicates that stronger connectivity of striatum with value-related OFC, contributes to successful nonreinforced behavioral change.

In conclusion, the current study sheds new light on the neural mechanisms underlying nonreinforced behavioral change. Our results show that preference modification is unlikely to occur within the face-selective brain network. However, we found for the first time, that individual differences in nonreinforced behavior change following CAT were related to fMRI activations in the striatum, vmPFC, and their connectivity with high-level visual regions. In addition to the theoretical advancement, our findings can also serve as a basis for novel applicative interventions, such as using striatal signals as individualized biomarker for successful learning, and enhancing learning via personalized neurofeedback applications without external reinforcements.

ACKNOWLEDGMENTS

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement n° 715016), and the Israeli Science Foundation granted to Tom Schonberg. Tom Salomon was supported by the Nehemia Levtzion fellowship and the Fields-Rayant Minducate Learning Innovation Research Center. The authors also want to thank Prof. G. Yovel and Ms. L. Kliger for their assistance with face-selective functional localizer.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

DATA AVAILABILITY STATEMENT

Our sample size, hypotheses and a general analyses plan were preregistered on the Open Science Framework (OSF), after data collection began, but prior to data analysis (project page: https://osf.io/aqnr4/; preregistration: https://osf.io/k7wn6). Deviations from the preregistration are described at the end of the methods section below. All behavioral and eye-tracking data along with the codes used for their analyses and the codes used for fMRI analyses are shared on https://github.com/tomsalomon/CAT_MRI_faces. MRI data are available in BIDS format at https://openneuro.org/datasets/ds001818. Uncorrected and cluster-corrected statistical maps of all contrasts described in current work are available at https://neurovault.org/collections/5578/.

ORCID

Tom Salomon https://orcid.org/0000-0002-1417-8163
Rotem Botvinik-Nezer https://orcid.org/0000-0003-2669-1877
Shiran Oren https://orcid.org/0000-0001-7690-3589
Tom Schonberg https://orcid.org/0000-0002-4485-816X

REFERENCES

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron*, *32*(3), 537–551. https://doi.org/10.1016/S0896-6273(01)00491-3

Alho, K., Salmi, J., Koistinen, S., Salonen, O., & Rinne, T. (2015). Top-down controlled and bottom-up triggered orienting of auditory attention to pitch activate overlapping brain networks. *Brain Research*, *1626*, 136–145. https://doi.org/10.1016/j.brainres.2014.12.050

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

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

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

Bakkour, A., Lewis-Peacock, J. A., Poldrack, R. A., & Schonberg, T. (2017). Neural mechanisms of cue-approach training. *NeuroImage*, *151*, 92–104. https://doi.org/10.1016/j.neuroimage.2016.09.059

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of subjective value. *NeuroImage*, *76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063.The

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bernstein, M., Erez, Y., Blank, I., & Yovel, G. (2018). An integrated neural framework for dynamic and static face processing. *Scientific Reports*, *8*. https://doi.org/10.1038/s41598-018-25405-9

Botvinik-Nezer, R., Salomon, T., & Schonberg, T. (2019). Enhanced bottom-up and reduced top-down neural mechanisms drive long-lasting non-reinforced behavioral change. *Cerebral Cortex*, 1–17. https://doi.org/10.1093/cercor/bhz132

Browning, M., Holmes, E. A., Charles, M., Cowen, P. J., & Harmer, C. J. (2012). Using attentional bias modification as a cognitive vaccine against depression. *Biological Psychiatry*, *72*(7), 572–579. https://doi.org/10.1016/j.biopsych.2012.04.014

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376. https://doi.org/10.1038/nrn3475

Cahill, K., & Perera, R. (2011). Competitions and incentives for smoking cessation. Cochrane database of systematic reviews. *Cochrane Database of Systematic Reviews*, (4). https://doi.org/10.1002/14651858.CD004307.pub4

Carp, J. (2012). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fMRI experiments. *Frontiers in Neuroscience*, *6*, 1–13. https://doi.org/10.3389/fnins.2012.00149

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, *29*(39), 12315–12320. https://doi.org/10.1523/JNEUROSCI.2575-09.2009

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, *9*(9), 1289–1302. https://doi.org/10.1093/scan/nst106

Cloutier, J., Heatherton, T. F., Whalen, P. J., & Kelley, W. M. (2008). Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. *Journal of Cognitive Neuroscience*, *20*(6). https://doi.org/10.1162/jocn.2008.20062

Colley, W. (2002). Colley's bias free college football ranking method: The colley matrix explained. Retreived from http://www.colleyrankings.com/matrate.pdf.

Collins, A. G. E., & Frank, M. J. (2014). Modeling interactive learning and incentive choice effects of striatal dopamine. *Psychological Review*, *121*(3), 337–366. https://doi.org/10.1037/a0037015

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Deichmann, R., Gottfried, J. A., Hutton, C., & Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage*, *19*(2), 430–441. https://doi.org/10.1016/S1053-8119(03)00073-9

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Elliott, R., Zahn, R., Deakin, J., & Anderson, I. (2011). Affective cognition and its disruption in mood disorders. *Neuropsychopharmacology*, *36*(1), 153–182. https://doi.org/10.1038/npp.2010.77

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., … Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*, 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fellows, L. K. (2011). The Neurology of Value. In J. A. Gottfried (Ed.), *Neurobiology of Sensation and Reward*. Boca Raton (FL): CRC Press/Taylor & Francis.

Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping*, *30*(5), 1637–1651. https://doi.org/10.1002/hbm.20630

Gerfen, C. R., Engber, T. M., Mahan, L. C., Susel, Z. V. I., Chase, T. N., Monsma, F. J., & Sibley, D. R. (1990). D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science*, *250*(4986), 1429–1432. https://doi.org/10.1126/science.2147780

Gildersleeve, K., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Columbia University. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. *NeuroImage*, *19*(1), 200–207. https://doi.org/10.1016/S1053-8119(03)00058-2

Glimcher, P. W., & Fehr, E. (2013). *Neuroeconomics: Decision making and the brain*. In: P. W. Glimcher & E. Fehr, (Eds.) (2nd ed.). Oxford, UK: Academic Press.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*, 1–9. https://doi.org/10.1038/sdata.2016.44

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), 696–701. https://doi.org/10.1371/journal.pmed.0020124

Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: Consensus and controversy. *Neuron*, *63*(6), 733–745. https://doi.org/10.1016/j.neuron.2009.09.003

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1098/Rstb.2006.1934

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*, 1292–1298. https://doi.org/10.1038/nn.2635

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857. https://doi.org/10.1073/pnas.1101328108

Kranz, F., & Ishai, A. (2006). Face perception is modulated by sexual preference. *Current Biology*, *16*(1), 63–68. https://doi.org/10.1016/j.cub.2005.10.070

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: Evidence from

functional neuroimaging. *Neuron*, *64*(3), 431–439. https://doi.org/10.1016/j.neuron.2009.09.040

Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/CBO978051161803

Lim, S.-L., O'Doherty, J. P., & Rangel, A. (2013). Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. *Journal of Neuroscience*, *33*(20), 8729–8741. https://doi.org/10.1523/jneurosci.4809-12.2013

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. A. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *NeuroImage*, *61*(4), 1277–1286. https://doi.org/10.1016/j.neuroimage.2012.03.068

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces a meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Mumford, J. (2007). A guide to calculating percent change with featquery. Tech Report. Available at: http://mumford.fmripower.org/perchange_guide.pdf

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*, 452–454. https://doi.org/10.1126/science.1094285

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Current Opinion in Neurobiology*, *14*(6), 769–776. https://doi.org/10.1016/j.conb.2004.10.016

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Peckham, A. D., McHugh, R. K., & Otto, M. W. (2010). A meta-analysis of the magnitude of biased attention in depression. *Depression and Anxiety*, *27*(12), 1135–1142. https://doi.org/10.1002/da.20755

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*, 1042–1045. https://doi.org/10.1038/nature05051

Prochaska, J. J., Delucchi, K., & Hall, S. M. (2004). A meta-analysis of smoking cessation interventions with individuals in substance abuse treatment or recovery. *Journal of Consulting and Clinical Psychology*, *72*(6), 1144–1156. https://doi.org/10.1037/0022-006X.72.6.1144

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*, 545–556. https://doi.org/10.1038/nrn2357

Salomon, T., Botvinik-Nezer, R., Gutentag, T., Gera, R., Iwanir, R., Tamir, M., & Schonberg, T. (2018). The cue-approach task as a general mechanism for long term non-reinforced behavioral change. *Scientific Reports*, *8*, 1–13. https://doi.org/10.1038/s41598-018-21774-3

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), 625–630. https://doi.org/10.1038/nn.3673

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

Senior, C. (2003). Beauty in the brain of the beholder. *Neuron*, *38*(4), 525–528. https://doi.org/10.1016/S0896-6273(03)00293-9

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*(1), 38–45. https://doi.org/10.1016/j.tics.2005.11.008

Seymour, B., & McClure, S. M. (2008). Anchors, scales and the relative coding of value in the brain. *Current Opinion in Neurobiology*, *18*(2), 173–178. https://doi.org/10.1016/j.conb.2008.07.010

Shomstein, S., & Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *Journal of Neuroscience*, *26*(2), 435–439. https://doi.org/10.1523/JNEUROSCI.4408-05.2006

Smith, D. V., Clithero, J. A., Boltuck, S. E., & Huettel, S. A. (2014). Functional connectivity with ventromedial prefrontal cortex reflects subjective value for social rewards. *Social Cognitive and Affective Neuroscience*, *9*(12), 2017–2025. https://doi.org/10.1093/scan/nsu005

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., … Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, 208–219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Stamarski, C. S., & Son Hing, L. S. (2015). Gender inequalities in the workplace: The effects of organizational structures, processes, practices, and decision makers' sexism. *Frontiers in Psychology*, *6*, 1–20. https://doi.org/10.3389/fpsyg.2015.01400

Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron*, *41*(2), 281–292. https://doi.org/10.1016/S0896-6273(03)00848-1

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

Vieira, T. F., Bottino, A., Laurentini, A., & De Simone, M. (2014). Detecting siblings in image pairs. *The Visual Computer*, *30*(12), 1333–1345. https://doi.org/10.1007/s00371-013-0884-3

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. A. (2011). Does the brain calculate value? *Trends in Cognitive Sciences*, *15*(11), 546–554. https://doi.org/10.1016/j.tics.2011.09.008

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag.

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270–273. https://doi.org/10.1126/science.1223252

Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, *92*(3), 1830–1839. https://doi.org/10.1152/jn.00155.2004

Yarkoni, T. (2009). Big correlations in little studies. *Perspectives on Psychological Science*, *4*(3), 294–298. https://doi.org/10.1111/j.1745-6924.2009.01127.x

Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron*, *44*(5), 889–898. https://doi.org/10.1016/j.neuron.2004.11.018

Zoltak, M. J., Veling, H., Chen, Z., & Holland, R. W. (2018). Attention! Can choices for low value food over high value food be trained? *Appetite*, *124*, 124–132. https://doi.org/10.1016/j.appet.2017.06.010

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Salomon T, Botvinik-Nezer R, Oren S, Schonberg T. Enhanced striatal and prefrontal activity is associated with individual differences in nonreinforced preference change for faces. *Hum Brain Mapp*. 2020;41:1043–1060. https://doi.org/10.1002/hbm.24859