PLOS | BIOLOGY

RESEARCH ARTICLE

Differential Activation Patterns in the Same Brain Region Led to Opposite Emotional States

Kazuhisa Shibata[1,2¤], Takeo Watanabe[1,2], Mitsuo Kawato[1]*, Yuka Sasaki[1,2]

1 Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International, 2-2-2 Hikaridai, Keihanna Science City, Kyoto, Japan, **2** Department of Cognitive, Linguistics, & Psychological Sciences, Brown University, Providence, Rhode Island, United States of America

¤ Current address: Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan
* kawato@atr.jp

OPEN ACCESS

Citation: Shibata K, Watanabe T, Kawato M, Sasaki Y (2016) Differential Activation Patterns in the Same Brain Region Led to Opposite Emotional States. PLoS Biol 14(9): e1002546. doi:10.1371/journal.pbio.1002546

Academic Editor: Leslie Ungerleider, National Institutes of Mental Health, UNITED STATES

Received: January 7, 2016

Accepted: August 5, 2016

Published: September 8, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was conducted under the 'Development of BMI Technologies for Clinical Application' of the Strategic Research Program for Brain Sciences supported by Japan Agency for Medical Research and Development (AMED) (http://www.nips.ac.jp/srpbs/). K. S. was partially supported by JSPS (https://www.jsps.go.jp/english/e-ab/index.html). T.W was partially supported by NIH R01 EY019466 (https://projectreporter.nih.gov/project_info_description.cfm?aid=8885171&icde=27667392&ddparam=&ddvalue=&ddsub=&cr=3&csb=default&cs=ASC), Y.S.

Abstract

In human studies, how averaged activation in a brain region relates to human behavior has been extensively investigated. This approach has led to the finding that positive and negative facial preferences are represented by different brain regions. However, using a functional magnetic resonance imaging (fMRI) decoded neurofeedback (DecNef) method, we found that different patterns of neural activations within the cingulate cortex (CC) play roles in representing opposite directions of facial preference. In the present study, while neutrally preferred faces were presented, multi-voxel activation patterns in the CC that corresponded to higher (or lower) preference were repeatedly induced by fMRI DecNef. As a result, previously neutrally preferred faces became more (or less) preferred. We conclude that a different activation pattern in the CC, rather than averaged activation in a different area, represents and suffices to determine positive or negative facial preference. This new approach may reveal the importance of an activation pattern within a brain region in many cognitive functions.

Author Summary

Although it is well studied how averaged activation of a brain region relates to behavior, it is still unclear if specific patterns of activation within regions also relate to cognitive function. In recent years, several methods have been developed for manipulating brain activity in humans. Real-time functional magnetic resonance imaging decoded neurofeedback (fMRI DecNef) is a method that allows the induction of specific patterns of brain activity by measuring the current pattern, comparing this to the pattern to be induced, and giving the subjects feedback on how close the two patterns of neuronal activity are. Using fMRI DecNef, we manipulated the pattern of activation in the cingulate cortex—a part of the cerebral cortex that plays a role in preference to different categories including faces and

by NIH R01 MH091801 (https://projectreporter.nih.gov/project_info_details.cfm?aid=8661290&icde=27667409&ddparam=&ddvalue=&ddsub=&cr=1&csb=default&cs=ASC) and NSF BCS 1539717 (http://www.nsf.gov/awardsearch/showAward?AWD_ID=1539717&HistoricalAwards=false), and M.K. by Japan MIC "Novel and Innovative Brain R&D" (http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc387330.html). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: I have read the journal's policy and the authors of this manuscript have the following competing interests. There is a potential financial conflict of interest; the authors are the inventors of patents related to the multi-voxel pattern induction method used in this study, while the original assignee of the patents is ATR, with which the authors are affiliated.

Abbreviations: CC, cingulate cortex; DecNef, decoded neurofeedback; fMRI, functional magnetic resonance imaging; ROI, region of interest.

daily items—and tested whether we could change these preferences. In the experiment, a specific activation pattern in the cingulate cortex corresponding to higher (or lower) preference was induced by fMRI DecNef while subjects were seeing a neutrally preferred face. As a result, these neutrally preferred faces became more (or less) preferred. Our finding suggests that different patterns of activation in the cingulate cortex represent, and are sufficient to determine, different emotional states. Our new approach using fMRI DecNef may reveal the importance of activation patterns within a brain region, rather than activation in a whole region, in many cognitive functions.

Introduction

A traditional approach in human studies is to examine how averaged activation in a brain region relates to behavior. Results obtained by this approach led most theories of cognitive functions in the human brain to assume that a different region or a group of regions in the human brain play a role in a different function. Although this approach has greatly advanced the understanding of neural mechanisms of human cognitive functions, it cannot effectively reveal a differential role of a pattern of activity within the same region in a different cognitive function. In animal studies, the importance of a role of activity of a certain group of neurons in a region rather than mean activity of the region has been observed [1,2]. Thus, to better understand neural mechanisms of human cognitive functions, it is necessary to investigate how a different pattern of activation within a region plays a different role.

Facial preferences influence a wide range of social outcomes from face perception to social behavior [3–13] and, therefore, has been a subject of great interest. Theories of facial preferences have also been developed, with the general consensus that positive and negative facial preferences are represented by different brain regions including the amygdala, basal ganglia, insular cortex, occipitotemporal cortex, orbit frontal cortex, lateral prefrontal cortex, and cingulate cortex (CC) [14–20]. The CC has also been reported to play roles in preference to different categories including faces and daily items [21–23].

A recently developed online functional magnetic resonance imaging (fMRI) decoded neurofeedback (DecNef) has allowed us to induce a different multi-voxel pattern of activation within the same brain region [24]. In the present study, using fMRI DecNef, we tested whether a different pattern of activations within a single brain region can sufficiently change facial preferences in a positive or negative direction.

In the experiment, we chose the CC as the target brain region for fMRI DecNef because the CC was found to be the best region whose multi-voxel activation patterns represent both positive and negative facial preferences in the current study (see below) among the regions previously implicated in facial preference [14–20]. We tested whether subjects' preferences to neutrally preferred faces could be changed toward a positive (or negative) direction by fMRI DecNef, which induced multi-voxel activation patterns in the CC that correspond to higher (or lower) preference with presentations of the neutrally rated faces to generate a new association between the faces and manipulated preferences. As a result, the previously neutrally rated faces became significantly more (or less) preferred. Although subjects' facial preferences were successfully changed, subjects remained unaware of the aim to change their facial preferences. On the contrary to the previous belief that a different brain region plays a role in positive or negative facial preference, our results are in accord with the hypothesis that it is a different activation pattern in the CC that represents and suffices to determine positive or negative facial preference.

Results

Determination of the CC as a Target Region

To determine a single region that would be used as a target for fMRI DecNef, we first conducted a pilot experiment (see Fig 1A, S1 Fig, and Pilot Experiment in Materials and Methods). Results of the pilot experiment showed that the CC most accurately reflects subjects' behavioral preference ratings both in the negative and positive directions among the regions previously implicated in facial preference (S1A Fig) [14–20]. Thus, we determined the CC as the target region for fMRI DecNef in the main experiment.

Main Experiment with fMRI DecNef

This experimental design was made with the aim to test whether induction of activation patterns in the CC, which represent higher (or lower) preference with presentations of neutrally preferred faces, can make these faces more (or less) preferred. The main experiment consisted of five stages (see Fig 1A and Main Experiment in Materials and Methods for details): pre-test (1 d), fMRI decoder construction (1 d), induction (fMRI DecNef, 3 d), post-test (20 min after the offset of the induction stage), and interview of subjects (immediately after the post-test stage). In the pre-test stage, we measured a distribution of behavioral preference ratings to faces for each subject. In the fMRI decoder construction stage, for each subject, we constructed a preference decoder to estimate the preference ratings, which were represented by the activation patterns in the target region (the CC). In the induction stage, fMRI DecNef was administered based on the preference decoder. Through fMRI DecNef, activation patterns in the CC were made to be similar to specific patterns, which represent higher (or lower) preference

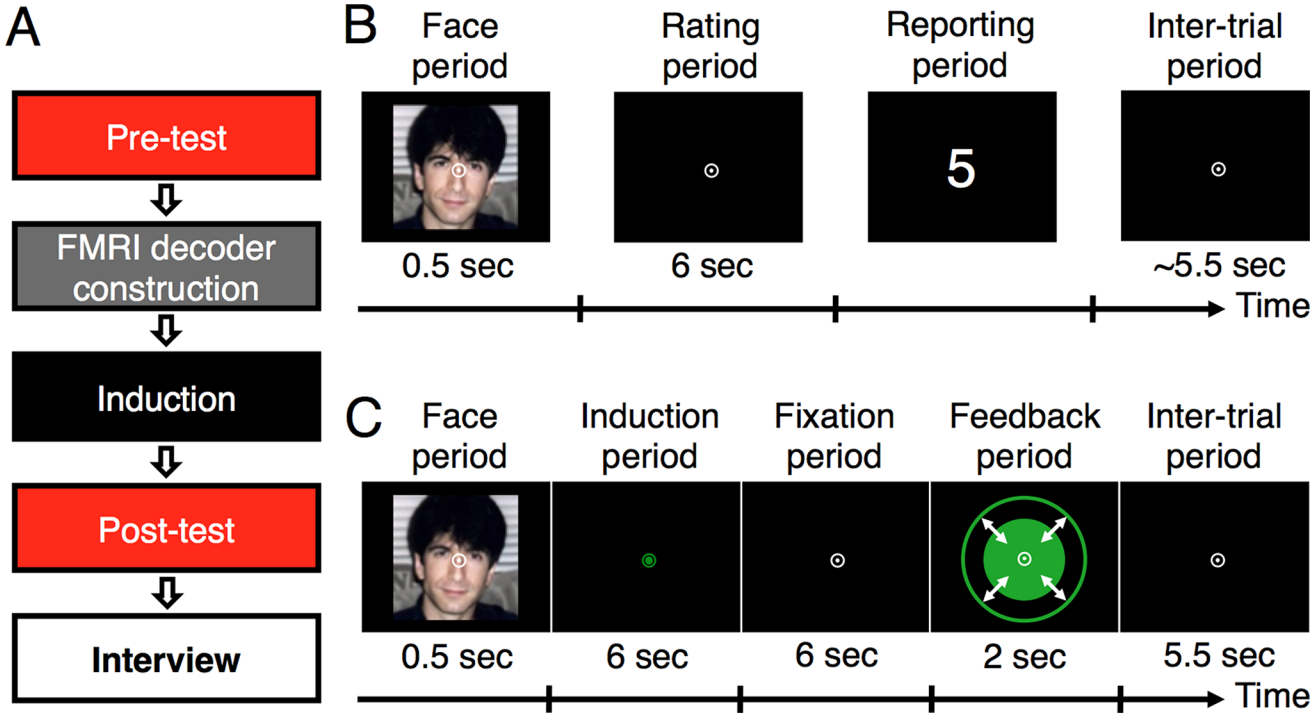

Fig 1. Procedures of the experiments. (A) Five stages of the main experiment. The pilot experiment consisted only of the first two stages. (B) Procedure of a trial in the pre-test, post-test, and fMRI decoder construction stages. (C) Procedure of a trial in the induction stage. The outer green circle in the feedback period indicates the maximum disk size.

doi:10.1371/journal.pbio.1002546.g001

ratings, in association with presentations of neutrally rated faces. In the post-test stage, subjects' behavioral preference ratings to the same faces as in the pre-test stage were measured so that we could test whether behavioral preference ratings to the previously neutrally rated faces were changed due to induction of the specific activation patterns in the CC.

In the pre-test stage, subjects' behavioral preference ratings to 400 face pictures were measured. In each trial (Fig 1B), after the brief presentation of each face, subjects determined their preference to the face on a scale of one to ten (one for the lowest, ten for the highest) during a 6 s rating period. In a subsequent reporting period, they were asked to report the determined preference rating.

Based on the behavioral preference ratings recorded in the pre-test stage, we selected neutrally rated faces, which would be used in the subsequent stages. For each subject, one set of 15 neutrally rated faces was randomly selected for use in the induction stage and called "induction faces." Another set of 15 neutrally rated faces was also randomly selected as "baseline faces" for a control set, which was not shown during the induction stage. Because both induction and baseline faces were neutrally rated in the pre-test stage, the average behavioral ratings were the same between the two sets of faces in the pre-test stage. Thus, comparison of changes in the behavioral preference ratings between the induction and baseline faces in the post-test stage would indicate whether pairings of specific activation patterns in the CC with the induction faces during the induction stage are sufficient to change the behavioral preference ratings to the induction faces.

In the fMRI decoder construction stage, we constructed the preference decoder (sparse linear regression [25]), which would be used during the subsequent induction stage for the target region (the CC). In the fMRI decoder construction stage, subjects again conducted the preference-rating task in the fMRI scanner (see Fig 1B and Main Experiment in Materials and Methods). Based on the fMRI signals measured during the rating period and corresponding behavioral preference ratings for each subject, we constructed the preference decoder to estimate the subject's behavioral preference ratings from activation patterns in the CC.

The purpose of the three-day induction stage was to associate the induction faces with specific activation patterns in the CC that represent higher (or lower) preference ratings through fMRI DecNef using the preference decoder. Subjects were randomly assigned to either a higher-preference (n = 12) or lower-preference (n = 12) group but were not informed of their assigned group. Each trial consisted of face, induction, fixation, feedback, and inter-trial periods (Fig 1C). During the face period, subjects were presented with one of the induction faces. In the induction period, subjects were instructed to somehow regulate their brain activity to make the size of a solid green disk (presented in the subsequent feedback period) as large as possible. Subjects were encouraged to enlarge the disk size so that they would receive a payment bonus proportional to the mean disk size. Subjects were given no further instructions. The size of the disk presented in the feedback period served as a feedback signal and reflected an estimated preference rating from the CC, which was calculated by applying the preference decoder to the activation pattern of the CC obtained in the preceding induction period of the trial (see Main Experiment in Materials and Methods for details). However, the computation of the disk size was opposite in its direction between the two groups, although the instructions given to the two groups were exactly the same. For the higher-preference group, the disk size was proportional to the estimated rating from the CC activation pattern. That is, if the CC activation became more similar to the patterns corresponding to higher preference, the disk size became larger. In contrast, for the lower-preference group, a lower estimated rating made the disk larger. This made the instruction and the range of feedback signals to both groups identical. Note that all other information, including the intended preference direction, the purpose of the induction stage, and the meaning of the disk size, was withheld from subjects so that

knowledge of the purpose of the experiment would not influence subjects' rating criteria in the post-test stage.

Changes in Behavioral Preference Rating as a Result of fMRI DecNef

To confirm that subjects' behavioral preference ratings to the induction faces changed as a result of associations of the induction faces with CC activation patterns of higher (higher-preference group) or lower (lower-preference group) preference ratings during the induction stage, the following three criteria have to be satisfied. First, subjects' preference ratings as behavioral measures for originally neutrally rated induction faces must be significantly higher with the higher-preference group and lower with the lower-preference group in the post-test stage than in the pre-test stage. Second, to rule out the effect of mere exposure to faces on preferences to the faces [26], the subjects' behavioral preference ratings must be unchanged simply by repeated exposures to the faces during the fMRI decoder construction and induction stages. Thus, a new group of subjects as a control group ($n = 6$) underwent an experiment in which visual presentations were identical to those for the higher- and lower-preference groups while no fMRI DecNef was administered (see Main Experiment in Materials and Methods for details). Third, changes in subjects' behavioral preference ratings after the induction stage must occur specifically for the induction faces but not for the baseline faces, which were originally neutrally rated but were not used during the induction stage.

To test if the results of the main experiment met these criteria, a three-way mixed-model ANOVA with factors being test stage (pre- versus post-test stages), face type (induction versus baseline faces), and group (higher-preference, lower-preference versus control groups) was applied to the behavioral preference ratings (Fig 2). The main effects of test stage ($F_{1,27} = 4.61$, $p = 0.04$) and group ($F_{2,27} = 3.56$, $p = 0.04$) were significant. Significant interactions were obtained between test stage and group ($F_{2,27} = 8.31$, $p < 10^{-2}$), between face type and group ($F_{2,27} = 10.84$, $p < 10^{-3}$), between test stage and face type ($F_{1,27} = 4.63$, $p = 0.04$), and among the three factors ($F_{2,27} = 13.10$, $p = 10^{-4}$). Post hoc t-tests revealed that, in the post-test stage, subjects' behavioral preference ratings to the induction faces were significantly higher for the higher-preference group (Fig 2, red; paired two-tailed t-test, $t_{11} = 4.78$, $p < 10^{-3}$; Bonferroni

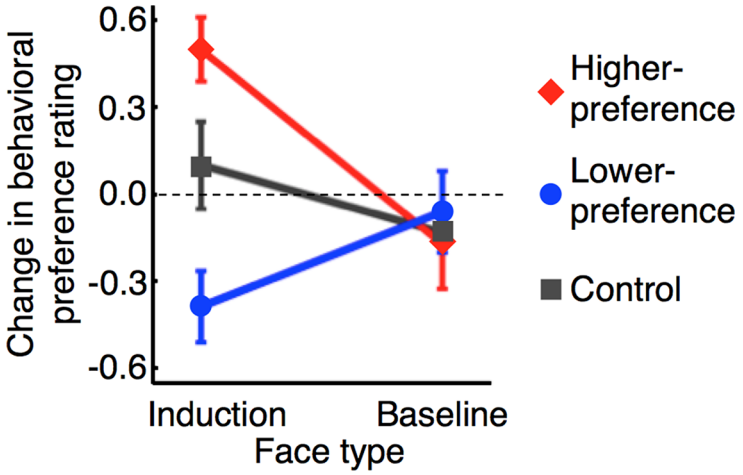

Fig 2. Changes in the behavioral preference ratings. The mean (± s.e.m.) changes in subjects' behavioral preference ratings to the induction faces and baseline faces in the post-test stage in comparison to the pre-test stage. Red diamonds, blue circles, and gray squares represent the higher-preference ($n = 12$), lower-preference ($n = 12$), and control ($n = 6$) groups, respectively.

doi:10.1371/journal.pbio.1002546

corrected) and significantly lower for the lower-preference group (Fig 2, blue; $t_{11} = 3.31$, $p < 10^{-2}$, Bonferroni corrected) than in the pre-test stage. The results meet the first criterion. Moreover, post hoc *t*-tests showed no significant changes in subjects' behavioral preference ratings between the two test stages for the control group (Fig 2, gray; $t_5 = 0.69$, $p = 0.52$), meeting the second criterion. For all of the three groups, no significant change in subjects' behavioral preference ratings was observed for the baseline faces, which were neutrally rated in the pre-test stage but not presented during the induction stage, meeting the third criterion (Fig 2, baseline faces; $t_{11} = 1.15$, $p = 0.27$ for the higher-preference group; $t_{11} = 0.45$, $p = 0.66$ for the lower-preference group; $t_5 = 0.72$, $p = 0.50$ for the control group). From all of these results, we conclude that association of originally neutrally rated faces with covert induction of activity patterns in the single brain region, the CC, led to changes in facial preference specifically for those faces and in a specific preference (positive or negative) direction.

Subjects Were Unaware of What the Disk Size Represented

It is important that subjects were unaware of our manipulation and what the disk size represented, because knowledge or suspicion of what the disk size represented could have significantly influenced subjects' rating criteria. In the interview stage, which was held right after the post-test stage, subjects from the higher- and lower-preference groups were asked whether they knew or suspected what the disk size represented and what, if anything, they tried to do to increase the disk size. None of their responses indicated even a slightest understanding of the true workings of the experiment (see S1 Data for details). Subjects were then debriefed on how the disk size was computed and were asked to guess whether they had been assigned to the higher- or lower-preference group. The accuracies of their guesses were indistinguishable from chance for the higher-preference (Chi-square test, $\chi = 0.17$, $p = 0.68$) and lower-preference ($\chi = 0.00$, $p = 1.00$) groups (Fig 3). These results of the interview stage suggest that subjects remained unaware of what the disk size represented. That is, it was beyond subjects' will that

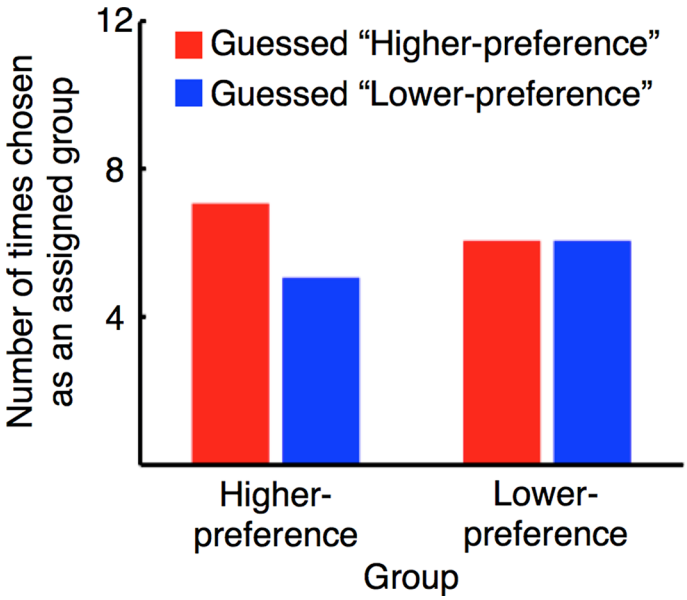

Fig 3. Results of the interview stage. The two left bars and two right bars show subjects' guesses as to which group they had been assigned for the higher- ($n = 12$) and lower- ($n = 12$) preference groups, respectively.

doi:10.1371/journal.pbio.1002546.g003

induction of specific activation patterns in the single region changed facial preference in a specific (positive or negative) direction.

Effects of Inductions of a Pattern of Activation on the Changes in the Behavioral Preference Ratings

The present study tested whether subjects' behavioral preference ratings to the neutrally rated faces can be altered by repetitive inductions of a certain pattern of activation within the same region. We found that the behavioral preference ratings to the neutrally rated faces were indeed changed in the post-test stage from the pre-test stage (Fig 2). If activation patterns in the CC during the induction stage resulted in the changes in the behavioral preference ratings, there should be a quantitative relationship between the induced shifts in activation patterns in the CC during the induction stage and the observed changes in subjects' behavioral preference ratings in the post-test stage.

To examine this relationship, we first quantified the activation patterns in the CC during the induction stage as described in (1), and called the metric induced CC-activation shifts. Then, we tested the prediction that the degree of the induced CC-activation shift should be correlated with the degree of the behavioral preference rating change as described in (2).

(1) Calculating CC-activation shifts during the induction stage. The CC-activation shift was calculated using the preference decoder in the following ways. As mentioned above, the preference-related component of an activation pattern in the CC in the induction stage was represented as an estimated preference rating by the preference decoder (see Main Experiment in Materials and Methods for details). The preference decoder was constructed based on the individual behavioral preference ratings obtained during the fMRI decoder construction stage for each subject. The average behavioral preference rating varied with a different subject during the fMRI decoder construction stage. Therefore, to appropriately evaluate and compare the induced shifts in activation patterns in the CC during the induction stage across subjects or groups of subjects, we subtracted the abovementioned average behavioral preference rating during the fMRI decoder construction stage from the estimated preference rating from the CC during the induction stage for each subject. The resultant value of the subtraction is called an induced CC-activation shift. That is, an induced CC-activation shift represents how far an induced activation pattern in the CC during the induction stage was shifted away from the activation pattern for the average behavioral preference rating for each subject (see Main Experiment in Materials and Methods for detailed description of the induced CC-activation shift). For instance, a "0" induced CC-activation shift indicates that the activation pattern of the CC was biased in neither the positive nor negative preference direction.

If the induced shifts in the activation patterns in the CC during the induction stage led to the observed changes in the behavioral preference ratings between the pre- and post-test stages, the induced CC-activation shifts should have a quantitative relationship to the behavioral preference rating changes.

(2) Testing of the prediction. The following results support the abovementioned prediction that there is a quantitative relationship between induced CC-activation shifts and changes in the behavioral preference ratings. We found a significant correlation between the changes in subjects' behavioral preference rating and the induced CC-activation shifts averaged over the three-day induction stage (Fig 4; Pearson's correlation test, $r_{22} = 0.78$, $p < 10^{-4}$). Importantly, the correlation coefficient remained significant if the group effect (higher- or lower-preference group) was removed by reversing the sign of the data for the lower-preference group (S2 Fig, $r_{22} = 0.57$, $p < 10^{-2}$). Moreover, as described below, the close-to-zero intercept and positive slope of the regression line (Fig 4) suggest a quantitatively close one-to-one relationship of the

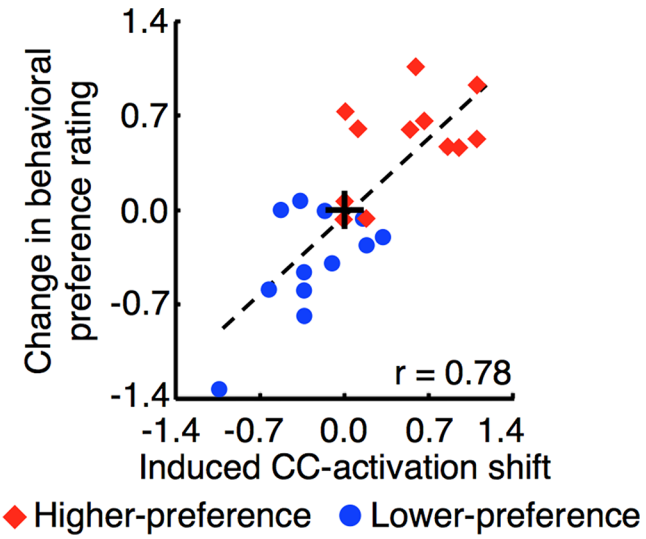

Fig 4. Significant correlation between the induced CC-activation shifts and the degree of changes in the behavioral preference ratings. Here shown is the scatter plot of the induced CC-activation shift during the three-day induction stage versus the change in subjects' behavioral preference rating for the higher- (red diamonds; $n = 12$) and lower- (blue circles; $n = 12$) preference groups. There was no significant outlier in the plot (Grubbs' test, $p > 0.05$). The black cross and broken line indicate the origin (0, 0) and the ordinary least-square regression, respectively. See S2 Fig for more details regarding the removal of the group effect.

doi:10.1371/journal.pbio.1002546.g004

induced CC-activation shifts to the changes in the behavioral preference ratings. The zero intercept of the regression line indicates that if an induced CC-activation shift was zero (in other words, activation patterns corresponding to higher or lower preference ratings did not occur in the CC), then no behavioral change occurred. The positive slope of the regression line means that if the induced CC-activation shift is positive (or negative), the behavioral change is also positive (or negative). Third, the slope was close to one. This indicates that the magnitude of the behavioral preference change was almost equivalent to that of the induced CC-activation shift. These results suggest that a different pattern of activation in the CC changed facial preference in a positive or negative direction.

The Differential Activation Patterns in the CC Can Explain the Bi-directional Changes in the Behavioral Preference Rating While Mere Pairing of Faces with Monetary Reward Cannot Account for the Bi-directional Behavioral Changes

We found significant bi-directional changes in the subjects' behavioral preference ratings to the induction faces in the post-test stage from the pre-test stage (Fig 2). In addition, the results of the quantitative analysis (Fig 4 and S2 Fig) suggested that a different activation pattern induced in the CC during the induction stage determined a direction and degree of the changes in subjects' behavioral preference ratings to the induction faces. However, is there any possibility that pairings of the induction faces with the monetary reward during the induction stage directly led to the subjects' behavioral preference rating changes in the post-test stage without relying on the differential activation patterns induced in the CC? This possibility arises because payment bonus (monetary reward) was given to subjects in proportion to the size of the green feedback disk during the induction stage. However, the following lines of evidence do not support this possibility.

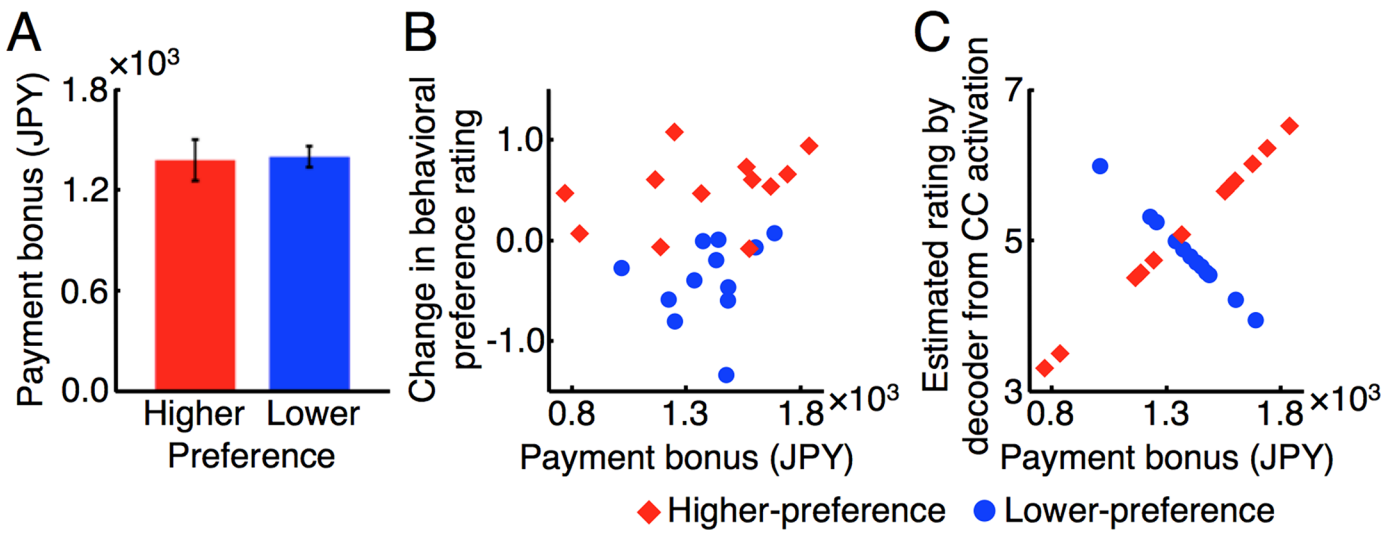

Fig 5. Monetary reward, activation patterns of the CC during the induction stage, and the behavioral preference rating changes. (A) The mean (± s.e.m.) amounts of payment bonus (monetary reward) for the higher- (red; $n = 12$) and lower- (blue; $n = 12$) preference groups. (B) Scatter plot of the amount of payment bonus versus the change in subjects' behavioral preference rating for the higher- (red diamonds) and lower- (blue circles) preference groups. (C) Scatter plot of the amount of payment bonus versus the estimated rating from activation patterns of the CC during the three-day induction stage for the higher- (red diamonds) and lower- (blue circles) preference groups.

doi:10.1371/journal.pbio.1002546.g005

First, the lack of difference in the amounts of monetary reward between the higher- and lower-preference groups is inconsistent with the model that the amount of reward should simply increase the preference. Assume that monetary reward was a direct deterministic factor to make faces more preferred. Then the amount of the payment bonus given to the higher-preference group should have been larger than the lower-preference group, because the behavioral preference rating increased in the higher-preference group and decreased in the lower-preference group (Fig 2). However, there was no significant difference in the amounts of the payment bonus between the higher- and lower-preference groups (Fig 5A; two-sample two-tailed t-test, $t_{22} = 0.19$, $p = 0.85$). Second, if monetary reward was a direct deterministic factor to make faces more preferred, the amount of the payment bonus should have been correlated with the degree of the behavioral preference rating change across the two groups. However, no significant correlation was found between them (Fig 5B; $r_{22} = 0.15$, $p = 0.49$). Third, if hypothesis for the direct role of reward is true, then the amount of the payment bonus should have been positively correlated with the estimated preference rating by the decoder from the CC activation patterns in each of the higher- and lower-preference groups. Indeed, the estimated rating from the CC activation pattern was positively correlated with the amount of the payment bonus in the higher-preference group (Fig 5C; red diamonds). However, the estimated rating was negatively correlated with the amount of payment bonus in the lower-preference group (Fig 5C; blue circle). These correlations reflect the experimental procedure for the induction stage (see Main Experiment in Materials and Methods). These lines of evidence clearly deny the possibility that monetary reward during the induction stage was the direct factor to induce both positive and negative preferences. They also support the conclusion that it is the differential activation patterns induced in the CC that led to the bi-directional changes in the behavioral preference ratings.

Induction of the Preference-Related Activation Patterns Was Largely Confined to the CC

The present results suggest that the different activation patterns in the CC resulted in different directions of facial preference. How can we know that the CC mainly determines the facial preference? The size of the feedback disk provided to subjects in the induction stage was based on activation patterns only in the CC. However, this procedure alone does not assure that inductions of the preference-related activation patterns were confined to the CC. It is possible that, in concert with the successful induction of specific activation patterns in the CC, similar activation patterns occurred in some other regions during the induction stage, which might also contribute to the behavioral preference rating changes. If the activation patterns in the CC that represented higher (or lower) preference ratings "leaked out" and induced similar preference-related activation patterns in other regions, the activation patterns in these regions should reconstruct the estimated ratings based on the CC activation patterns on a trial-by-trial basis.

To test the possibility that preference-related activation patterns in the CC leaked out of the CC to other regions, we conducted two leak analyses. In the first analysis (see Leak analysis using an ROI-based method for details), we anatomically divided the whole brain into a total of 38 regions (the CC and 37 other regions) and compared the amounts of leakage in the 38 regions. In the second analysis (see Leak analysis using a searchlight method for details), using a searchlight method [27], we mapped the amount of leakage by moving a searchlight sphere throughout the whole brain during the fMRI decoder construction stage and induction stage and compared these maps. We describe details of these two leak analyses below.

In the first leak analysis, we tested whether the estimated ratings based on the CC activation patterns were reconstructed using activation patterns measured in each of the aforementioned 37 other brain regions during the induction stage, as well as activation patterns in the CC itself as a control. We defined reconstruction performance as a correlation coefficient between the reconstructed values and the estimated ratings based on the CC activation patterns (see Leak analysis using an ROI-based method in Materials and Methods for details). A high correlation coefficient would indicate that there was leakage of the CC activation patterns to the other brain regions. However, the Fisher-transformed correlation coefficients for the other 37 regions (Fig 6, gray) were significantly and markedly smaller than that for the CC itself (Fig 6, red; paired two-tailed t-test on z-scores of the correlation coefficients after permutation [28], $t_{23} > 13.26$, $p < 10^{-11}$, Bonferroni corrected; see Permutation Test in Materials and Methods for details of calculation of the z-scores). If activation patterns in other brain regions had been closely linked to those in the CC, the correlation coefficients should have been as high as those in the CC. Therefore, our results indicate that it is unlikely that the activation patterns induced in the CC leaked out to other regions.

In the second leak analysis, using a searchlight method (see Leak analysis using a searchlight method for details), we examined how well activation patterns of voxels in a moving searchlight sphere can reconstruct the estimated ratings in the CC (1) while subjects were conducting the preference-rating task in the fMRI decoder construction stage and (2) while they were going through the induction stage. A comparison of the results between the fMRI decoder construction stage and induction stage should clarify that preference-related activation patterns were confined to the CC during the induction stage. During the fMRI decoder construction stage, the CC and several other regions showed significant and high reconstruction performances on the reconstruction of the estimated ratings based on activation patterns in the CC (Fig 7A). This result demonstrated that the searchlight method possesses a sufficient power of sensitivity for detecting leaking of preference-related activation patterns out of the CC to other regions. However, during the induction stage, significant and high reconstruction performances were found only for voxels

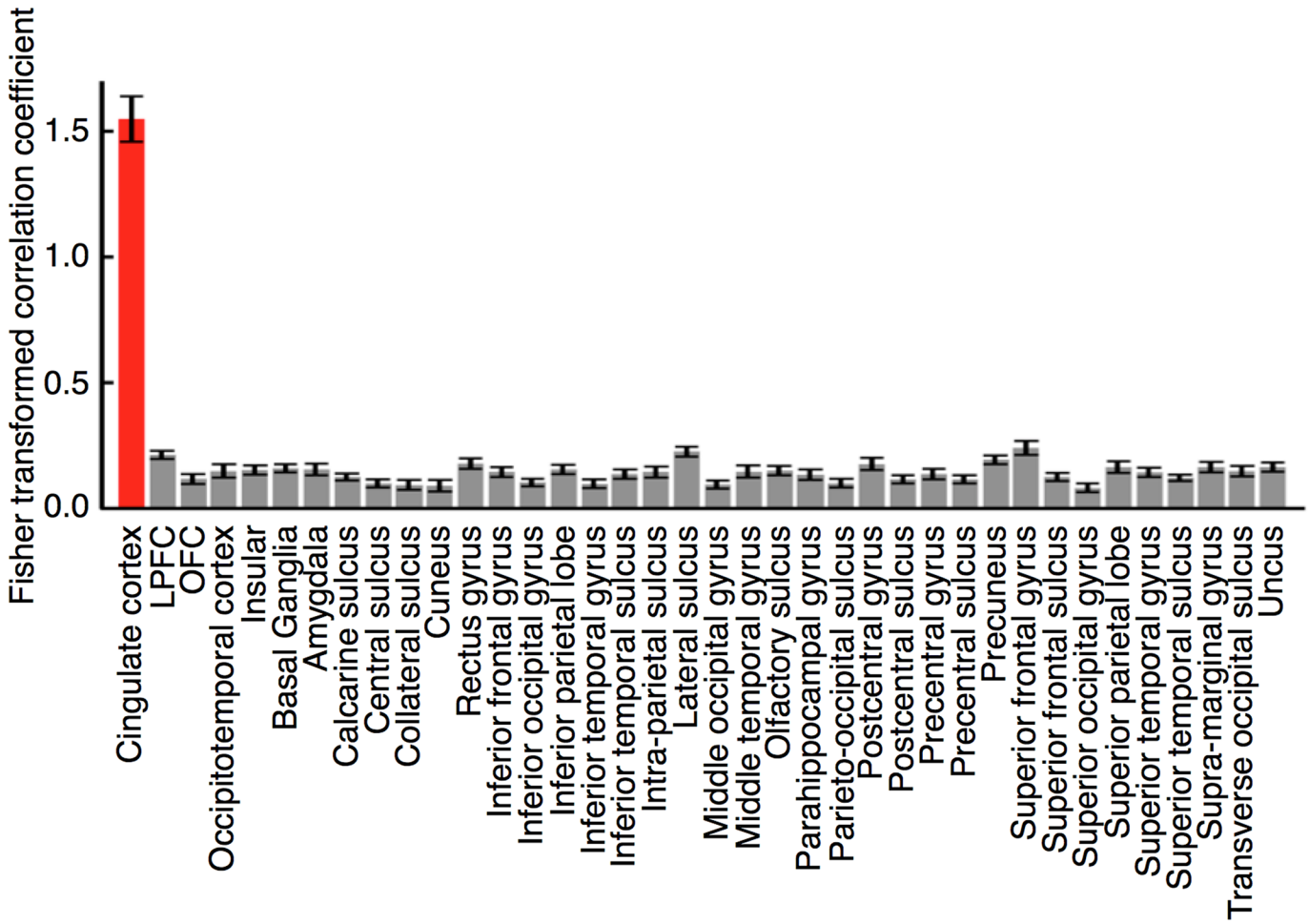

Fig 6. Results of a leak analysis using an ROI-based method. Each bar represents the mean (± s.e.m.) Fisher-transformed correlation coefficients between the estimated ratings from the CC activation patterns during the induction stage and the reconstructed values of the estimated ratings from activation patterns in each of the CC (red) and other regions (gray). Subjects from the higher- and lower-preference groups were combined (n = 24).

doi:10.1371/journal.pbio.1002546.g006

within the CC (Fig 7B). Although a few regions in the prefrontal and parietal cortices showed significant reconstruction performance in reconstruction, these performances were much lower than those in the voxels within the CC. These results obtained by the searchlight analysis indicate the following two points. First, when subjects conducted the preference-rating task in the fMRI decoder construction stage, preference-related activation patterns in the CC leaked out of the CC to several other regions. Second, during the induction stage, there was little leakage of preference-related activation patterns out of the CC to other regions.

These two leak analyses collectively and consistently demonstrate that the preference-related activation patterns were indeed largely confined to the CC during the induction stage. We conclude that the behavioral rating changes due to fMRI DecNef ware mainly induced by activation patterns within the CC.

Discussion

To our knowledge, this is the first study in which manipulation of a different brain activation pattern in a single region successfully changed facial preferences in the positive or negative

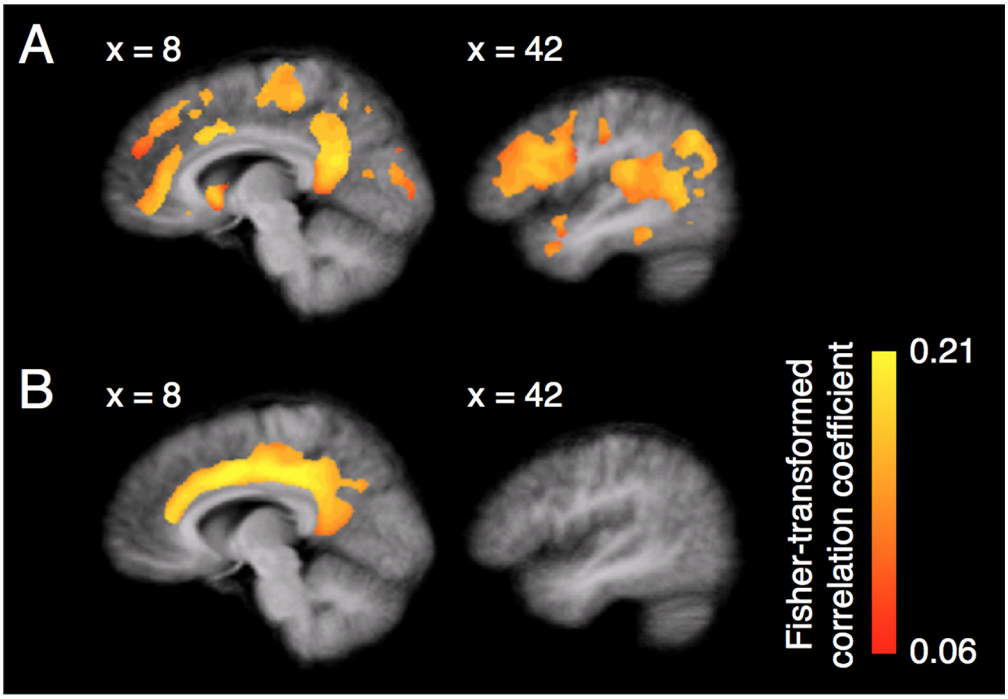

Fig 7. Results of a leak analysis using a searchlight method. (A) Mean map of Fisher-transformed correlation coefficients between the estimated preference ratings in the CC and its reconstructed values based on activation patterns of voxels in each searchlight sphere across subjects during the fMRI decoder construction stage. (B) Mean map of the Fisher-transformed correlation coefficients during the induction stage. Subjects from the higher- and lower-preference groups were combined ($n = 24$). In both (A) and (B), only the voxels showing significant correlation coefficients are highlighted in colors here (one-sample two-tailed t-test on z-scores of the correlation coefficients after permutation, $p < 0.05$, Bonferroni corrected; see Permutation Test in Materials and Methods for details of calculation of z-scores).

doi:10.1371/journal.pbio.1002546.g007

direction. Previous neuroscientific approaches have revealed that a different region or a group of regions is involved in positive or negative facial preference [14–20] and developed the general consensus that positive and negative facial preferences are represented by different brain regions. Contrary to this general consensus, the present results using fMRI DecNef [24] indicated a case in which highly selective activity patterns in the single brain region represent and suffice to determine both positive and negative facial preferences.

Additionally, our control analyses provide two important insights. First, the facial preference was not necessarily increased by a larger amount of monetary reward. Especially in the lower-preference group, the higher the monetary reward was, the less liked faces were. Second, when the CC was solely manipulated by fMRI DecNef during the induction stage, the effect of the manipulation on other brain regions was minimal or negligible. Together, the present results suggest that the induced activation pattern within the CC played a crucial role in changing the facial preference.

Based on the abovementioned results, we argue that a different activation pattern mainly in the CC as a single region represents positive or negative facial preference. However, one may raise the possibility that the activation patterns for positive and negative facial preference are spatially distributed in different subregions of the CC. We tested whether this is the case. We found that there is no significant difference in the extent of spatial distributions in the CC between positive and negative facial preferences (see S3 Fig for details). Thus, it is unlikely that a different subregion of the CC represents each of the positive and negative facial preferences.

These results are in accord with the hypothesis that positive and negative facial preferences are represented by spatially overlapped but different activation patterns in the CC.

Are changed activations in the CC during the induction period specific to facial preference? If so, activation changes during the induction period should be more related to the facial preference than any other period of a trial. To test this possibility, we analyzed activation patterns in the CC during the fixation period as well as the inter-trial period of the induction stage and compared them with activation patterns in the induction period (see Fig 1C for time course of a trial in the induction stage). In particular, we applied the preference decoder to the activation patterns during the fixation as well as inter-trial periods and calculated the induced CC-activation shifts as in the induction period (S4 Fig). During the induction period, the induced CC-activation shifts for the higher- and lower-preference groups were significantly different from zero (S4A Fig). In contrast, during the fixation period (S4B Fig) or inter-trial period (S4C Fig), the induced CC-activation shifts were not significantly different from zero. These results are consistent with the possibility that changed activations in the CC during the induction period were specific to facial preference rather than related to other functions or states including general emotion.

One may also wonder whether the significant shifts in induced CC-activations during the induction period (Fig 4 and S4A Fig) were simply due to changes in the overall mean activation amplitudes in the CC rather than induced multi-voxel activation patterns associated with higher or lower facial preferences. To test this possibility, we calculated the overall mean amplitudes across voxels in the CC during the induction period for both the higher- and lower-preference groups over the three-day induction stage (see S5 Fig for details) and compared the mean amplitudes between the two groups. There was no significant difference in the mean amplitudes between the groups. We also found that in neither group the overall mean activation amplitude was significantly different from zero on any day of the induction stage. These results suggest that the behavioral rating changes toward higher- and lower-preference resulted from induction of certain activation patterns in the CC rather than changes in the overall mean activation amplitudes in the CC.

It has been found that a number of brain regions are involved in facial preferences [14–20]. Why was the CC in particular selected as the target region for fMRI DecNef to change subjects' facial preferences? An important purpose of the current study was to test whether a different pattern of activity in a single brain region changes facial preference in an opposite direction. In the aforementioned pilot experiment, we found that the CC is the best region that codes both positive and negative facial preferences depending on the pattern of activation within the CC. Although the prefrontal cortex plays an important role in facial preferences in studies using a univariate analysis, it has been found that a different part of the prefrontal cortex is involved in each of the positive and negative facial preference coding [18]. Thus, the criterion for the selection of the CC was not based on whether the CC is the most important region for facial preference but rather was based on the fact that the CC was found to code both positive and negative facial preferences in the current study.

In summary, the results of the present study indicate that highly selective activity patterns for higher or lower preference within the CC that were repeatedly paired with facial stimuli lead to changes in subjects' facial preferences. Although these results do not deny important roles of other regions in facial preference, our study clearly demonstrates that inductions of different patterns of activation within the CC suffice to determine changes of facial preference in opposite directions beyond subjects' will [29].

Materials and Methods

The study was approved by the Institutional Review Board of the Advanced Telecommunications Institutes International (ATR) in which the current study was conducted.

Subjects

Thirty-three naïve subjects (19 to 29 years old; 23 males and 10 females) with normal or corrected-to-normal vision participated in the study. All experiments and data analyses were conducted in ATR. All subjects gave written informed consents.

Pilot Experiment

The purpose of the pilot experiment was to determine a target region of interest (ROI) to be used in the main experiment. Three subjects participated in the pilot experiment. The complete experiment consisted of two stages: pre-test (1 d) and fMRI decoder construction (1 d). The two stages were separated by at least 24 h.

Only behavioral data were collected from the pre-test stage, during which subjects' behavioral preference ratings to face pictures were measured. Subjects performed a preference-rating task (Fig 1B) for a total of 400 trials across 20 runs. During each run, subjects were asked to fixate on a white bull's-eye presented at the center of the display. A brief break period was provided after each run upon a subject's request.

We used a pool of 400 face pictures (200 males, 200 females, of a variety of races and ages) collected from several open databases [30–35]. A stimulus primarily consisted of a face and usually included some body parts, including hair, a neck, and shoulders, as well as a background scene. Each face picture was 4° square in size. The order of presentation of faces was randomized for each subject.

Each trial (Fig 1B) consisted of a face period (0.5 s), a rating period (6 s), and a reporting period, in the respective order. During the face period, a face picture was presented for 0.5 s at the center of the display. During the rating period, only a fixation point was presented at the center. Subjects were instructed to rate their preference to the previously presented face on a ten-point scale (one for the lowest preference, ten for the highest preference). During the reporting period, subjects were asked to report the preference rating by pressing two buttons (left and right) on a keyboard using the index and middle fingers of their right hand. At the beginning of the reporting period, a random number from one to ten was selected and presented at the center of the display. Subjects were asked to adjust its value to their preference rating, pressing the left button to increment it. Values were "wrapped" so that when subjects attempted to increment past a value of ten, values would start over at one. Subjects completed the reporting period by pressing the right button. After completion of the reporting period, the next trial began.

For each subject, 240 of the 400 face pictures were selected to be used for the subsequent fMRI decoder construction stage. Selections for pictures were based on subjects' individual behavioral preference ratings in the pre-test stage: 100 highest-rated faces, 100 lowest-rated faces, and 40 neutrally rated faces.

In the fMRI decoder construction stage, we measured subjects' blood-oxygen-level dependent (BOLD) signal patterns (see MRI Measurements and Parameters) while they once again conducted the preference-rating task on the 240 face pictures selected from the pre-test stage (the 100 highest-rated faces, 100 lowest-rated faces, and 40 neutrally rated faces). The measured BOLD signal patterns and behavioral preference ratings were in turn used to compute parameters for a preference decoder for each of the different ROIs (see below). Task procedures were identical to those of the pre-test stage except that an inter-trial period was added to the end of each trial, in which only a white fixation point was presented at the center of the display (Fig 1B). Subjects were asked to report their ratings within the reporting period (maximum of 5.5 s). The duration of the inter-trial period varied across trials, depending on subjects' reporting time, so that the total duration of the reporting and inter-trial periods would be equal to 5.5 s.

Each fMRI run for the fMRI decoder construction stage consisted of 20 task trials (one trial = 12 s) plus a 10 s fixation period before the trials and a 2 s fixation period after the trials (one run = 252 s). The fMRI data for the initial 10 s were discarded to allow the longitudinal magnetization to reach equilibrium. Subjects conducted a total of 240 trials in 12 fMRI runs. A presentation order for the abovementioned 240 face pictures was randomized for each subject. Throughout each fMRI run, subjects were asked to fixate on a white bull's-eye presented at the center of the display. A brief break period was provided after each fMRI run upon a subject's request.

Recorded fMRI data were preprocessed using the BrainVoyager QX software [36]. All functional images underwent 3-D motion correction. No spatial or temporal smoothing was applied. Rigid-body transformations were performed to align the functional images to the structural image for each subject. A gray matter mask was used to extract BOLD signals only from gray matter voxels for further analysis.

We specified seven ROIs implicated in facial preference [14–20] according to anatomical data for each subject: the cingulate cortex (CC), lateral prefrontal cortex (LPFC), orbitofrontal cortex (OFC), occipitotemporal cortex, insular cortex, basal ganglia, and amygdala. LPFC was defined as the middle frontal gyrus plus the inferior frontal sulcus. OFC was defined as the orbital gyrus plus the orbital sulci. The occipitotemporal cortex was defined as the lateral occipitotemporal gyrus, the medial occipitotemporal gyrus, plus the occipitotemporal sulcus. The basal ganglia were defined as the caudate, the pallidum, the putamen, plus the nucleus accumbens. Voxels from the left and right hemispheres were merged for each ROI. The cortical regions were specified using an atlas on the BrainVoyager QX software [36]. A cortical surface for each subject was spatially normalized into a standard cortical surface using a cortex-based alignment method [37]. Then, the specified regions were projected into a native space for each subject. The subcortical regions were specified for each subject using an automated brain parcellation method [38] on the Freesurfer software (http://surfer.nmr.mgh.harvard.edu).

A time-course of BOLD signal intensities was extracted from each voxel in each ROI and shifted by 4 s to account for the hemodynamic delay using the Matlab software. A linear trend was removed from the time-course. The time-course was z-score normalized for each voxel using all time points except for those for the initial 10 s in each fMRI run. This normalization was aimed to minimize baseline differences in time-courses of BOLD signal intensities across the fMRI runs. The data samples for computing the decoder were created by averaging the BOLD signal intensities of each voxel for three volumes corresponding to the 6 s rating period.

To construct a preference decoder for each ROI, we used a sparse linear regression algorithm [25], which automatically selected the relevant voxels within an ROI for decoding. Note that the behavioral preference ratings measured in this study were non-linear. Although they ranged from one to ten, preference measurement on the Likert-type scale cannot be considered strictly linear. Thus, before applying the sparse linear regression for each ROI, the behavioral preference ratings were linearized using an arc hyperbolic tangent function. An estimated rating $R_{decoded}$, that is, the decoder output calculated based on an activation pattern for a trial, was obtained in each ROI by

$$R_{decoded} = W_{voxel}^{T} \cdot A_{voxel} + b.$$

Here, A_{voxel} represents the activation pattern of voxels in the ROI for the trial. W_{voxel} indicates linear weights for the voxels, which were optimized by the sparse linear regression algorithm based on fMRI data, which was used for training the decoder. b corresponds to the decoder's constant term, which was determined for each subject as his/her average behavioral preference rating across all faces in the preference-rating task during the fMRI decoder construction stage.

A_{voxel} and W_{voxel} are denoted as n-dimensional column vectors with n as the number of voxels in each ROI. T denotes matrix transpose. The inputs to the decoder were subjects' moment-to-moment brain activations in each ROI, whereas the outputs from the decoder represented the decoder's best estimate of the corresponding behavioral preference ratings.

Decoder performance for each ROI was defined as the correlation coefficient between actual subjects' behavioral preference ratings in the preference-rating task and the estimated ratings calculated from activation patterns of the ROI and evaluated by a leave-one-run-out cross validation procedure. In the cross-validation procedure, the pairs of the actual subjects' behavioral preference ratings and the activation patterns for the ROI measured on one fMRI run were treated as the test data (20 samples), whereas those measured on the remaining runs (220 samples) were used for training the decoder to estimate subjects' trial-by-trial behavioral preference ratings. Thus, 12 cross-validation sets were generated per subject. For each voxel, activation amplitudes of the training and test data were normalized by mean and variance of activation amplitudes of the training data so that mean and variance of voxel activation amplitudes corresponding to the induction period would be zero and one, respectively. The correlation coefficients for each ROI were first standardized using Fisher's transformation, averaged over the cross-validation sets, and then averaged across subjects, as shown in S1A Fig.

The result of the fMRI decoder construction stage showed that the highest decoder performance was obtained from the CC (S1A Fig). Consistent with this result, previous neuroimaging studies have reported that the CC is highly involved in facial preference [15,18,39] as well as preferential decision-making in general [21–23]. Thus, we selected the CC as the target region for the main experiment. Note that the highest decoder performance was also found in the CC when we evaluated decoder performance in the same way using the fMRI signals obtained in the fMRI decoder construction stage of the main experiment (S1B Fig). These results indicate robustness of the tendency in which the CC most accurately reflects subjects' facial preference in the preference-rating task in this study.

Main Experiment

Thirty subjects participated in the main experiment. The main experiment consisted of five stages: pre-test (1 d), fMRI decoder construction (1 d), induction (fMRI DecNef, 3 d), post-test (20 min after the induction stage), and interview (immediately after the post-test stage) stages, in this order (Fig 1A). The pre-test, fMRI decoder construction, and induction stages were separated by at least 24 h. Thirty subjects in the main experiment were randomly assigned to one of the higher-preference (n = 12), lower-preference (n = 12), and control (n = 6) groups. They were not informed about their assigned group.

The procedures of the pre- and post-test stages in the main experiment were identical to those of the pre-test stage in the pilot experiment. As in the pilot experiment, the 100 highest-rated faces, the 100 lowest-rated faces, and 40 neutrally rated faces were selected to be used in the subsequent fMRI decoder construction stage. In addition, out of the 40 neutrally rated faces, 15 were randomly selected for use in the induction stage ("induction faces") and another set of 15 was also randomly selected as a preference-matched control against the induction faces ("baseline faces"). The baseline faces were not shown during the subsequent induction stage.

For the higher- and lower-preference groups, the procedures of the fMRI decoder construction stage in the main experiment were identical to those in the pilot experiment (see Pilot Experiment). A preference decoder for the CC was computed for each subject for use in the subsequent induction stage. To train the decoder, we used 240 data samples obtained from the 240 trials in the 12 fMRI runs. For each voxel, activation amplitudes of the training data were

normalized by the mean and the variance of activation amplitudes of the training data so that the mean and the variance of voxel activation amplitudes would be zero and one, respectively. The mean (± s.e.m.) numbers of voxels selected by the sparse linear regression algorithm to decode the subjects' preference ratings were 219.8 ± 0.5 across subjects. Note that because the decoder was built based on data samples from all the trials in the fMRI decoder construction stage, a unique set of voxels was selected for each subject.

For the control group, the visual presentations in the fMRI decoder construction stage were identical to those in the pilot experiment while the experiment was conducted outside the MRI scanner without fMRI measurements.

In the induction stage, which consisted of three daily sessions, subjects from the higher- and lower-preference groups were instructed to regulate the activation of their brains, which were controlled by an online fMRI technique [24,40–43]. On each day, subjects participated in up to 12 fMRI runs. The mean (± s.e.m.) number of runs per day was 10.8 ± 0.2. Each fMRI run for the induction stage consisted of 15 trials (one trial = 20 s) preceded by a 30 s fixation period (one run = 330 s). The fMRI data for the initial 10 s were discarded to allow the longitudinal magnetization to reach equilibrium. During each run, subjects were instructed to fixate on a white bull's-eye presented at the center of the display. After each fMRI run, a brief break period was provided upon a subject's request.

Each trial in the induction stage (Fig 1C) consisted of a face period (0.5 s), an induction period (6 s), a fixation period (6 s), a feedback period (2 s), and an inter-trial period (5.5 s), in that order. During the face period, one of the 15 induction faces described above was presented for 0.5 s at the center of the display. The order of presentation of the 15 induction faces was randomized for each fMRI run. During the induction period, the color of the fixation point changed from white to green, and no visual stimulus except for the fixation point was presented. Subjects were instructed to regulate activation of their brain, with the goal of making the size of a solid green disk presented in the later feedback period as large as possible. The experimenters provided no further instructions or strategies. During the fixation period, subjects were asked simply to fixate on the central white point. This period was inserted between the induction period and the feedback period to compensate for the known hemodynamic delay, which we assumed lasted 4 s, during which activation patterns in the CC were calculated in time for a green disk to be shown in the subsequent feedback period. The feedback period presented the green disk for 2 s. The size of the disk was determined based on the estimated rating (see below), which is the decoder output value based on the BOLD signal pattern of the CC measured in the prior induction period. The green disk was always enclosed by a larger green concentric circle (10° in diameter), which indicated the disk's possible maximum size. The feedback period was followed by the inter-trial period, during which subjects were asked to fixate on the central white point. This period was followed by the start of the next trial.

The size of the disk presented during the feedback period was based on the estimated rating from the CC activation pattern, which was computed during the fixation period, as follows. First, measured functional images underwent 3-D motion correction using the Turbo BrainVoyager software. Second, a time-course of BOLD signal intensities was extracted from each of the voxels in the CC identified in the fMRI decoder construction stage and was shifted by 4 s to account for the hemodynamic delay. Third, a linear trend was removed from the time-course for each voxel using a linear regression algorithm based on all time points except for those for the initial 10 s in each fMRI run, and the BOLD signal time-course was z-score normalized for each voxel using BOLD signal intensities measured for 20 s starting from 10 s after the onset of each fMRI run. Fourth, the data sample used to calculate the size of the disk was created by averaging the BOLD signal intensities of each voxel for three volumes corresponding to the 6 s induction period. Finally, the estimated rating was calculated from the data sample using the

decoder constructed in the fMRI decoder construction stage. For the higher-preference group, the size of the disk was proportional to the estimated rating (ranging from one to ten). For the lower-preference group, the size of the disk was proportional to 11 minus the estimated rating so that a lower estimated rating resulted in a larger disk. In addition to the fixed amount of the compensation for participation in the experiment, a bonus of up to 3,000 JPY was paid to subjects based on the mean size of the disk during each day.

Induced CC-activation shifts shown in Fig 4 and S4A Fig indicate how far the activation patterns of the CC during the induction stage are from the activation patterns in the CC corresponding to the average behavioral preference rating. The induced CC-activation shift was calculated as follows. First, the estimated rating $R_{decoded}$ from an activation pattern of the CC for a trial was computed by the same method as in the pilot experiment, but only from the CC. That is, $R_{decoded}$ was computed by

$$R_{decoded} = W_{voxel}^{T} \cdot A_{voxel} + b.$$

Here, A_{voxel} represents the activation pattern of voxels in the CC in the induction period. W_{voxel} indicates linear weights for the voxels, which had been computed for each subject in the fMRI decoder construction stage. b corresponds to the decoder's constant term and had been determined for each subject as her/his average behavioral preference rating in the preference-rating task during the fMRI decoder construction stage. This constant term varied across subjects. A_{voxel} and W_{voxel} are denoted as n-dimensional column vectors with n as the number of voxels in the CC. T denotes matrix transpose.

The induced CC-activation shift for a trial was defined by

$$R_{decoded} - b = W_{voxel}^{T} \cdot A_{voxel}.$$

Based on the following computation, the induced CC-activation shift represents how far the activation patterns in the CC during the induction stage are away from the activation pattern in the CC corresponding to the average behavioral preference rating initially for each subject. As described above, the constant term b was determined for each subject as her/his average behavioral preference rating in the preference-rating task during the fMRI decoder construction stage. The set of 240 faces (the 100 highest-rated faces, the 100 lowest-rated faces, and 40 neutrally rated faces) used in the fMRI decoder construction stage was selected for each subject according to her/his behavioral preference ratings to the 400 faces presented in the pre-test stage. That is, the constant term b represents the subject's average behavioral preference rating for the population of faces. Thus, the induced CC-activation shift, which is calculated by subtracting the constant term b from the estimated rating $R_{decoded}$, represents how far the activation pattern in the CC is from the activation patterns in the CC corresponding to the average behavioral preference rating for each subject.

It is necessary to obtain the induced CC-activation shift in order to appropriately evaluate and compare induced activation patterns in the CC across subjects and groups during the induction stage. Note that an induced CC-activation shift being "0" indicates that the activation pattern of the CC was biased in neither the high nor low preference direction. A positive (or negative) value of the induced CC-activation shift indicates that the activation pattern of the CC was biased toward a positive (or negative) preference direction, compared to the activation pattern corresponding to the average behavioral preference rating for each subject. Thus, when the mean-induced CC-activation shift is significantly higher (or lower) than 0, this means that subjects accomplished significant learning to induce the preference-related activation patterns in the CC that correspond to higher (or lower) preference rating.

With the control group, during the induction stage, the induction faces were presented in the same way as with the higher- and lower-preference groups. On the other hand, unlike with the higher- and lower-preference groups, the experiment was conducted outside the MRI scanner without fMRI measurements for the control groups. Subjects from the control group conducted a fixation task (see below) instead of the task given to those from the higher- and lower-preference groups during the induction stage.

In the fixation task for the control group in the "induction" stage, during the 6 s "induction" period, the luminance of the central fixation point slightly decreased (from green to dark green), returning to its original luminance 300 ms later. This luminance change occurred several times in an unpredictable manner during the 6 s period. Subjects from the control group were asked to count the number of luminance changes and report whether the number of the changes was even or odd by pressing one of two buttons using the index or middle finger of their right hand during the fixation period. The task difficulty was controlled by using an adaptive staircase method, so that the overall task difficulty was kept constant throughout the induction stage; the degree of fixation luminance change was slightly increased in the trial following an incorrect answer, and slightly decreased after two consecutive correct answers. Otherwise, luminance was kept around the same. The mean (± s.e.m.) task accuracy for the fixation task was 67.4% ± 5% across subjects. The green "feedback" disk was presented during the 2 s "feedback" period. The size of the disk was determined randomly for each trial. Subjects were instructed to fixate on the center of the display during the feedback period.

Leak Analyses

We tested whether preference-related activation patterns in the CC leaked out of the CC to other regions and induced similar activation patterns in these regions. We conducted the leak analysis using an ROI-based method and a searchlight method.

Leak analysis using an ROI-based method. We first anatomically divided the whole brain into a total of 38 regions (the CC and 37 other regions; see S1 Table for the number of voxels in each region). The 37 other regions were specified using an atlas on the BrainVoyager QX software [36], as shown in Fig 6 and S1 Table.

Second, for each subject we used the sparse linear regression algorithm [25] to reconstruct the estimated ratings in the CC based on activation patterns measured in each of the aforementioned 37 other brain regions during the induction stage, as well as based on those in the CC itself as a control. A reconstructed value $R_{reconstructed}$ for a trial in each region was obtained by

$$R_{reconstructed} = W_{voxel}^{T} \cdot A_{voxel} + b.$$

Here, A_{voxel} represents an activation pattern of voxels in a region for a trial. W_{voxel} indicates linear weights for the voxel values, which were optimized for each subject by the sparse linear regression algorithm, and b corresponds to the constant term, which was determined as the average of the estimated ratings from the CC activation patterns during the induction stage for the subject. We defined reconstruction performance for each of the 38 regions as a correlation coefficient between the reconstructed values and the estimated ratings from the CC activation patterns and evaluated by a leave-one-run-out cross-validation procedure (Fig 6). Statistical comparisons were conducted on z-scores of the correlation coefficients after permutation (see Permutation Test for details).

Leak analysis using a searchlight method. We employed a searchlight method [27] by moving a searchlight sphere (radius = 15 mm) across the whole gray matter (~35,000 voxels). Using the searchlight method, we calculated whole-brain maps, which represent leaking of

preference-related activation patterns within the CC to other regions both during the fMRI decoder construction stage (Fig 7A) and induction stage (Fig 7B).

Here are details of these calculations of the two whole-brain leak maps. For the fMRI decoder construction stage, the leak map was obtained as follows. We calculated reconstruction performance of the estimated rating in the CC. This calculation was conducted based on activation patterns of voxels in a searchlight sphere measured during the fMRI decoder construction stage. In each sphere, the reconstruction performance was evaluated by a "nested cross-validation"-like procedure, as described below. In the inner cross-validation loop, we calculated estimated ratings based on activation in the entire CC as reconstruction targets. This calculation was conducted using the sparse linear regression [25]. We trained an estimated preference regressor using 220 data samples from 11 out of the 12 fMRI runs in the fMRI decoder construction stage. Then, we calculated 20 reconstruction targets using 20 data samples from the remaining fMRI run. Thus, there were 12 cross-validation sets in this inner cross-validation loop. In the outer cross-validation loop, reconstructed values of the targets were calculated based on activation patterns of voxels within the sphere using L_1-regulalized least-square regression. Specifically, we trained a regression model using 19 out of the 20 data samples for reconstruction. Then, we calculated one reconstructed value using one remaining data sample. Thus, there were 20 cross-validation sets in the outer cross-validation loop. This "nested cross-validation"-like procedure allowed us to calculate 240 reconstructed values of the estimated ratings from activation patterns in the CC for all 240 trials in the fMRI decoder construction stage. We defined reconstruction performance as the correlation coefficient between the reconstructed values and the estimated ratings from the CC activation patterns during the fMRI decoder construction stage.

For the induction stage, the leak map was obtained as follows. First, we calculated estimated ratings based on activation patterns in the entire CC using 240 data sampled from the fMRI decoder construction stage. This calculation was conducted using the sparse linear regression algorithm. Second, we trained a regression model to reconstruct the estimated ratings in the entire CC based on activation patterns of voxels in each searchlight sphere during the decoder construction stage. This regression model was trained using an L_1-regulalized least-square regression algorithm. Then, this regression model was applied to data sampled from all trials during the induction stage. Note that the estimated ratings in the entire CC during the induction stage had been obtained in the experiment with fMRI DecNef. Thus, we defined the reconstruction performance as the correlation coefficient between the estimated ratings in the entire CC obtained during the induction stage and its reconstructed values calculated by the above-mentioned regression model.

The above analyses using the searchlight method were conducted on the native coordinates of subjects. The results were spatially smoothed by a Gaussian filter (FWHM = 4 mm) and projected to a Talairach space. For each voxel, a statistical test was conducted on z-scores of correlation coefficients after permutation (see Permutation Test for details of calculation of the z-scores).

As mentioned above, to shorten computation time, we used the L_1-regulalized least-square regression in the leak analysis using the searchlight method. The L_1-regulalized least-square regression was implemented by using the LASSO algorithm, which has two hyper-parameters: α and λ. The hyper-parameter α was fixed to one for all subjects. The hyper-parameter λ determines the balance between the residual and L_1 norm of weights for voxels in each sphere. λ was determined for each subject by an independent analysis using the L_1-regulalized least-square regression. In this analysis, based on activation patterns during the fMRI decoder construction stage across all searchlight spheres, λ was determined to maximize the mean decoding performance of the subject's behavioral preference ratings.

Permutation Test

To appropriately evaluate the statistical significance of a correlation coefficient between two vectors of values, we performed a permutation test in the following ways. (1) We permutated a relationship between the two vectors 1,000 times and obtained a permutation distribution of correlation coefficients between the two values. (2) We tested whether a correlation coefficient between the two values obtained from the original relationship was ranked within top 5% in the distribution. If so, the correlation coefficient was regarded as significant.

In the case of between-subject statistics, we calculated a z-score of the original correlation coefficient by comparing the original correlation coefficient with the distribution obtained by the permutation [28].

Apparatus

Visual stimuli were presented on an LCD display (1,024 × 768 resolution, 60 Hz refresh rate) during the pre- and post-test stages and via an LCD projector (1024 × 768 resolution, 60 Hz refresh rate) during fMRI measurements in a dim room. All visual stimuli were made using the Matlab software and Psychtoolbox 3 [44] on Mac OS X.

MRI Measurements and Parameters

Subjects were scanned in a 3T MR scanner (Siemens MAGNETOM Verio) with a 12-channel head matrix coil in the ATR Brain Activation Imaging Center. FMRI signals were measured using a gradient echo-planar imaging sequence. In the fMRI experiments, 33 contiguous slices (repetition time = 2 s, voxel size = 3 × 3 × 3.5 mm^3, 0 mm slice gap, field of view = 192 mm × 192 mm, echo time = 26 ms, matrix size = 64 × 64, bandwidth = 2,367 Hz/pixel, phase encoding direction: from anterior to posterior, slice order: interleaved) oriented parallel to the AC-PC plane were acquired, covering the entire brain. For an automated parcellation method [38], T1-weighted MR images (magnetization-prepared rapid gradient-echo or MP-RAGE; 256 slices, the number of partition = 208, voxel size = 1 × 1 × 1 mm^3, 0 mm slice gap, repetition time = 2,250 ms, inversion time = 900 ms, echo time = 3.06 ms, flip angle = 9 deg, field of view = 256 mm, matrix size = 256 × 256, bandwidth = 230 Hz/pixel, phase encoding direction: from anterior to posterior, partition (2nd phase) encoding direction: from right to left) were also acquired during the fMRI decoder construction stage.

Supporting Information

S1 Data. Subjects' reports in the interview stage.
(DOCX)

S2 Data. Changes in the behavioral preference ratings to the induction and baseline faces. The Excel spreadsheet shows the mean (± s.e.m.) changes in the behavioral preference ratings to the induction and baseline faces in the higher- and lower-preference groups. Fig 2 is based on values shown in this sheet.
(XLSX)

S3 Data. Amount of monetary reward, induced CC-activation shifts, and behavioral preference rating changes. The Excel spreadsheet shows the amounts of monetary reward during the induction stage, induced CC-activation shifts during the induction stage, and changes in behavioral preference ratings for each subject in the higher- and lower-preference groups. Fig 4, Fig 5, S2 Fig, and S4 Fig are based on values shown in this sheet.
(XLSX)

S4 Data. Results of the leak analysis using a ROI-based method. The Excel spreadsheet shows the mean (± s.e.m.) Fisher-transformed correlation coefficients between the estimated ratings from the CC activation patterns during the induction stage and the reconstructed values of the estimated ratings from activation patterns in each of the 38 regions across subjects. Fig 6 is based on values shown in this sheet.
(XLSX)

S5 Data. Results of the leak analysis for the fMRI decoder construction stage using a searchlight method. The.mat file contains the mean Fisher-transformed correlation coefficients between the estimated ratings from the CC activation patterns during the fMRI decoder construction stage and its reconstructed values across subjects in the Talairach coordinates. Fig 7A is based on values shown in this file.
(MAT)

S6 Data. Results of the leak analysis for the induction stage using a searchlight method. The.mat file contains the mean Fisher-transformed correlation coefficients between the estimated ratings from the CC activation patterns during the induction stage and its reconstructed values across subjects in the Talairach coordinates. Fig 7B is based on values shown in this file.
(MAT)

S7 Data. Decoding performance for seven ROIs in the fMRI decoder construction stage. The Excel spreadsheet shows the mean (± s.e.m.) Fisher-transformed correlation coefficients between the subjects' behavioral preference ratings and the estimated ratings from the CC activation patterns during the fMRI decoder construction stage for the seven ROIs across subjects. S1 Fig is based on values shown in this sheet.
(XLSX)

S8 Data. Spatial distribution of the positive weight values of the preference decoder in the CC. The.mat file contains the mean positive weight values of the preference decoder across subjects in the CC in the Talairach coordinates. S3A Fig is based on values shown in this file.
(MAT)

S9 Data. Spatial distribution of the negative weight values of the preference decoder in the CC. The.mat file contains the mean negative weight values of the preference decoder across subjects in the CC in the Talairach coordinates. S3B Fig is based on values shown in this file.
(MAT)

S10 Data. The induced CC-activation shifts in the induction, fixation, and inter-trial period. The Excel spreadsheet shows the mean (± s.e.m.) induced CC-activation shifts in the induction, fixation, and inter-trial periods in each of the three days during the induction stage for the higher- and lower-preference groups. S4 Fig is based on values shown in this sheet.
(XLSX)

S11 Data. Overall mean BOLD signal amplitudes in the CC during the three-day induction stage. The Excel spreadsheet shows the overall mean (± s.e.m.) BOLD signal amplitudes in the CC in the three days during the induction stage for the higher- and lower-preference groups. S5 Fig is based on values shown in this sheet.
(XLSX)

S1 Fig. Comparison of decoding performance among seven regions in the fMRI decoder construction stage in the pilot and main experiments. (A) The mean (± s.e.m.) Fisher-transformed correlation coefficient in the pilot experiment (n = 3). The CC showed the highest decoding performance. (B) The mean (± s.e.m.) Fisher-transformed correlation coefficient in

the main experiment ($n = 24$). The pilot experiment (see Pilot Experiment in Materials and Methods for details) was aimed to select a single region that would be used as a target for fMRI DecNef in the main experiment. The selection was made based on comparison of performance of preference decoders among regions of interests (ROIs) that were implicated in facial preference [14–20]: the amygdala, basal ganglia, insular cortex, occipitotemporal cortex, orbit frontal cortex (OFC), lateral prefrontal cortex (LPFC), and cingulate cortex (CC) (see Pilot Experiment in Materials and Methods for the definition of the ROIs). Decoder's performance for each ROI was defined as a correlation coefficient between subjects' behavioral preference ratings and the estimated rating by the decoder from activation patterns of the ROI (see Pilot Experiment in Materials and Methods for details). In the main experiment (shown in B), we found that the decoder successfully predicted the subjects' behavioral preference ratings based on the fMRI datasets measured in the fMRI decoder construction stage in the basal ganglia (one-sample two-tailed t-test on z-scores of the correlation coefficients after permutation, $t_{23} = 3.20$, $p < 10^{-2}$; Bonferroni corrected; see Permutation Test in Materials and Methods for details of calculation of the z-scores), insular cortex ($t_{23} = 3.07$, $p < 10^{-2}$, Bonferroni corrected), occipitotemporal cortex ($t_{23} = 4.17$, $p < 10^{-3}$, Bonferroni corrected), OFC ($t_{23} = 4.11$, $p < 10^{-3}$, Bonferroni corrected), LPFC ($t_{23} = 4.46$, $p < 10^{-3}$, Bonferroni corrected), and CC ($t_{23} = 12.52$, $p < 10^{-4}$, Bonferroni corrected), but not in the amygdala ($t_{23} = 0.64$, $p = 0.53$). The mean performance in the CC was significantly higher than that in the other ROIs (paired two-tailed t-test on z-scores of the correlation coefficients after permutation, $t_{23} > 4.73$, $p < 10^{-4}$; Bonferroni corrected; see Permutation test in Materials and Methods for details of calculation of the z-scores). The results of the main experiment (shown in B) showed the same tendency as in the pilot experiment (shown in A), indicating that activation patterns of the CC most accurately reflected subjects' behavioral preference ratings in the preference-rating task.
(TIF)

S2 Fig. Significant correlation between the induced CC-activation shifts and the changes in the behavioral preference ratings. The same scatter plot of the induced CC-activation shift during the three-day induction stage versus the change in subjects' behavioral preference rating for the higher- (red diamonds; $n = 12$) and lower- (blue circles; $n = 12$) preference groups as in Fig 4, except that the sign of the data for the lower-preference group was reversed. The broken line indicates the ordinary least-square regression. For each of the higher- and lower-preference groups, we found a significant correlation between the CC-activation shifts and changes in subjects' behavioral preference rating ($r = 0.62$ for the higher-preference group, $r = 0.50$ for the lower-preference group, permutation test, $p < 0.05$; see Permutation Test in Materials and Methods for details of the permutation test).
(TIF)

S3 Fig. Spatial distributions of weights of the preference decoder in the CC. (A) The mean spatial distribution of the positive weight values across subjects. (B) The mean spatial distribution of the negative weight values across subjects. We tested whether the activation patterns for positive and negative facial preferences are spatially dissociable in the CC. We also tested whether the spatial activation pattern is consistent across subjects. In particular, we compared spatial structures of activation patterns associated with positive and negative facial preferences in two ways. The first analysis was aimed to examine spatial distributions of weights of the preference decoder in the CC. The preference decoder was constructed based on measured activation patterns in the CC in the fMRI decoder construction stage, during which subjects reported their preference rating to faces (see Materials and Methods for details). The preference decoder was constructed to estimate the subjects' preference ratings based on the activation patterns in the CC. Thus, the CC voxels that have positive weight values contribute to positive

facial preferences, whereas those that have negative weight values contribute to negative facial preferences. Therefore, examination of the spatial distributions of the weight values in the CC should tell us whether the activation patterns associated with positive and negative facial preferences are distributed differently in the CC. Positive (shown in A) and negative (shown in B) weight values uniformly distributed and their distributions were largely overlapped in the CC. Second, we tested whether positive and negative weight patterns are consistent across subjects. If these patterns are consistent across subjects, the mean weight value across subjects for a given voxel should be significantly different from zero. In contrast, if these patterns are not consistent across subjects, the mean weight value for a given voxel would be around zero. We found that the mean weight value across subjects was not significantly different from zero in any of voxels in the CC (one-sample two-tailed t-test, $p > 0.05$, Bonferroni corrected). These results suggest that spatial activation patterns for positive and negative facial preferences are spatially overlapped and that these activation patterns vary across subjects. These lines of evidence are at odds with the possibility that activation patterns for positive and negative facial preferences are distributed in different subsections in the CC.
(TIF)

S4 Fig. Induced CC-activation shifts during the induction, fixation, and inter-trial periods in each day of the induction stage. (A) Induction period. (B) Fixation period. (C) Inter-trial period. The mean (± s.e.m.) induced CC-activation shifts for the higher- (red) and lower- (blue) preference groups. To test whether similar activation patterns were induced in the CC during the fixation and inter-trial period to during the induction period, we calculated the induced CC-activation shifts over the three-day induction stage using the preference decoder and compared them among the induction, fixation, and inter-trial periods. If activation patterns were similar among these periods, the same pattern of induced CC-activation shifts should be observed for all three periods. A three-way mixed-model ANOVA with factors being period (induction, fixation, versus inter-trial periods), induction day (Day 1, Day 2, versus Day 3), and group (higher- versus lower-preference groups) was applied to the induced CC-activation shifts. The main effects of period ($F_{2,44} = 3.48$, $p = 0.04$) and group ($F_{1,22} = 5.51$, $p = 0.03$) were significant. Significant interactions were obtained between period and group ($F_{2,44} = 15.09$, $p < 10^{-4}$) and between induction day and group ($F_{2,44} = 3.48$, $p = 0.04$). These results are not in accord with the possibility that the same pattern of induced CC-activation shifts should be observed for all three periods. As shown in (A), for the induction period, on Day 3, the induced CC-activation shifts were significantly greater than zero for the higher-preference group (one-sample two-tailed t-test, $t_{11} = 5.15$, $p < 10^{-3}$; Bonferroni corrected) and significantly smaller than zero for the lower-preference group ($t_{11} = 4.09$, $p < 10^{-2}$; Bonferroni corrected). These results demonstrate that by Day 3 subjects successfully learned to induce activation patterns in the CC that correspond to higher (higher-preference group) or lower (lower-preference group) preference ratings during the induction stage. During the fixation period (shown in B), for neither group the induced CC-activation shifts were significantly different from zero on any day (one-sample two-tailed t-test, $t_{11} < 1.43$, $p > 0.18$). During the inter-trial period (shown in C) for neither group, the induced CC-activation shifts were significantly different from zero on any day (one-sample two-tailed t-test, $t_{11} < 1.35$, $p > 0.20$). These results are not consistent with the possibility that the neurofeedback affected subjects' general emotional state, which in turn influenced activation patterns throughout the overall experimental session in the induction stage.
(TIF)

S5 Fig. Overall mean (± s.e.m.) BOLD signal amplitudes in the CC during the three-day induction stage for the higher- (red) and lower- (blue) preference groups. To test whether

the overall mean amplitude in the CC changed during the induction period, we conducted an amplitude-based analysis using the general linear model on the BrainVoyager QX software. The amplitude of each voxel in the CC was calculated based on the contrast between the induction and inter-trial periods, and the calculated amplitudes were averaged across the voxels. To test if the overall amplitudes differ between the higher- and lower-preference groups, a two-way mixed-model ANOVA with factors being group (higher- versus lower-preference groups) and induction day (day 1, day 2, versus day 3) was applied to the overall amplitudes in the CC. None of a main effect of group ($F_{1,22} = 1.01$, $p = 0.33$), a main effect of induction day ($F_{2,44} = 1.23$, $p = 0.30$), or an interaction between the two factors ($F_{2,44} = 0.45$, $p = 0.64$) was significant. In addition, for neither group, the amplitudes were significantly different from zero on any day (one-sample two-tailed t-test, $t_{11} < 2.17$, $p > 0.05$).
(TIF)

S1 Table. The mean (± s.e.m.) number of voxels in each of 38 regions across subjects.
(XLSX)

Acknowledgments

We thank O. Yamashita for helps in mathematical analysis, J. Dobres for editing an early draft, B. Seymour for helpful comments on the work, and Y. Oshima and ATR BAIC for technical assistances.

Author Contributions

Conceived and designed the experiments: KS TW MK YS.

Performed the experiments: KS.

Analyzed the data: KS MK YS.

Contributed reagents/materials/analysis tools: KS.

Wrote the paper: KS TW MK YS.

References

1. Janak PH, Tye KM. From circuits to behaviour in the amygdala. Nature. 2015; 517(7534):284–92. Epub 2015/01/17. doi: 10.1038/nature14188 PMID: 25592533; PubMed Central PMCID: PMC4565157.
2. Young MP, Yamane S. Sparse population coding of faces in the inferotemporal cortex. Science. 1992; 256(5061):1327–31. Epub 1992/05/29. PMID: 1598577.
3. Chiao JY, Adams RB, Tse PU, Lowenthal L, Richeson JA, Ambady N. Knowing Who's Boss: fMRI and ERP Investigations of Social Dominance Perception. Group Processes & Intergroup Relations. 2008; 11(2):201–14. Epub 2008/04/01. doi: 10.1177/1368430207088038 PMID: 19893753; PubMed Central PMCID: PMC2773146.
4. Cowen AS, Chun MM, Kuhl BA. Neural portraits of perception: reconstructing face images from evoked brain activity. NeuroImage. 2014; 94:12–22. Epub 2014/03/22. doi: 10.1016/j.neuroimage.2014.03.018 PMID: 24650597; PubMed Central PMCID: PMC4028096.
5. Fogelson SV, Kohler PJ, Miller KJ, Granger R, Tse PU. Unconscious neural processing differs with method used to render stimuli invisible. Frontiers in Psychology. 2014; 5:601. Epub 2014/07/02. doi: 10.3389/fpsyg.2014.00601 PMID: 24982647; PubMed Central PMCID: PMC4058905.
6. Haberman J, Whitney D. Rapid extraction of mean emotion and gender from sets of faces. Current Biology: CB. 2007; 17(17):R751–3. Epub 2007/09/07. doi: 10.1016/j.cub.2007.06.039 PMID: 17803921; PubMed Central PMCID: PMC3849410.
7. Korn HA, Johnson MA, Chun MM. Neurolaw: Differential brain activity for black and white faces predicts damage awards in hypothetical employment discrimination cases. Social Neuroscience. 2012; 7 (4):398–409. Epub 2011/11/09. doi: 10.1080/17470919.2011.631739 PMID: 22059860.

8. Leib AY, Puri AM, Fischer J, Bentin S, Whitney D, Robertson L. Crowd perception in prosopagnosia. Neuropsychologia. 2012; 50(7):1698–707. Epub 2012/04/24. doi: 10.1016/j.neuropsychologia.2012.03.026 PMID: 22521873.
9. Little AC, Jones BC, DeBruine LM. Facial attractiveness: evolutionary based research. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences. 2011; 366(1571):1638–59. Epub 2011/05/04. doi: 10.1098/rstb.2010.0404 PMID: 21536551; PubMed Central PMCID: PMC3130383.
10. Post RB, Haberman J, Iwaki L, Whitney D. The frozen face effect: why static photographs may not do you justice. Frontiers in Psychology. 2012; 3:22. Epub 2012/03/01. doi: 10.3389/fpsyg.2012.00022 PMID: 22363302; PubMed Central PMCID: PMC3282501.
11. Sherman A, Sweeny TD, Grabowecky M, Suzuki S. Laughter exaggerates happy and sad faces depending on visual context. Psychonomic Bulletin & Review. 2012; 19(2):163–9. Epub 2012/01/05. doi: 10.3758/s13423-011-0198-2 PMID: 22215467; PubMed Central PMCID: PMC3307857.
12. Shibata K, Watanabe T. Preference suppression caused by misattribution of task-irrelevant subliminal motion. Proceedings Biological Sciences / The Royal Society. 2012; 279(1742):3443–8. Epub 2012/06/15. doi: 10.1098/rspb.2012.0797 PMID: 22696526; PubMed Central PMCID: PMC3396910.
13. Sweeny TD, Suzuki S, Grabowecky M, Paller KA. Detecting and categorizing fleeting emotions in faces. Emotion. 2013; 13(1):76–91. Epub 2012/08/08. doi: 10.1037/a0029193 PMID: 22866885; PubMed Central PMCID: PMC3758689.
14. Aharon I, Etcoff N, Ariely D, Chabris CF, O'Connor E, Breiter HC. Beautiful faces have variable reward value: fMRI and behavioral evidence. Neuron. 2001; 32(3):537–51. Epub 2001/11/16. PMID: 11709163.
15. Chatterjee A, Thomas A, Smith SE, Aguirre GK. The neural response to facial attractiveness. Neuropsychology. 2009; 23(2):135–43. Epub 2009/03/04. doi: 10.1037/a0014430 PMID: 19254086
16. Iaria G, Fox CJ, Waite CT, Aharon I, Barton JJ. The contribution of the fusiform gyrus and superior temporal sulcus in processing facial attractiveness: neuropsychological and neuroimaging evidence. Neuroscience. 2008; 155(2):409–22. Epub 2008/07/02. doi: 10.1016/j.neuroscience.2008.05.046 PMID: 18590800; PubMed Central PMCID: PMC2605709.
17. Liang X, Zebrowitz LA, Zhang Y. Neural activation in the "reward circuit" shows a nonlinear response to facial attractiveness. Social Neuroscience. 2010; 5(3):320–34. Epub 2010/03/12. doi: 10.1080/17470911003619916 PMID: 20221946; PubMed Central PMCID: PMC2885490.
18. O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ. Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia. 2003; 41(2):147–55. Epub 2002/12/03. doi: S0028393202001458 [pii]. PMID: 12459213.
19. Said CP, Haxby JV, Todorov A. Brain systems for assessing the affective value of faces. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences. 2011; 366(1571):1660–70. Epub 2011/05/04. doi: 10.1098/rstb.2010.0351 PMID: 21536552; PubMed Central PMCID: PMC3130375.
20. Senior C. Beauty in the brain of the beholder. Neuron. 2003; 38(4):525–8. Epub 2003/05/27. PMID: 12765605.
21. Izuma K, Matsumoto M, Murayama K, Samejima K, Sadato N, Matsumoto K. Neural correlates of cognitive dissonance and choice-induced preference change. Proc Natl Acad Sci U S A. 2010; 107(51):22014–9. Epub 2010/12/08. doi: 1011879108 [pii] doi: 10.1073/pnas.1011879108 PMID: 21135218; PubMed Central PMCID: PMC3009797.
22. Kuhn S, Gallinat J. The neural correlates of subjective pleasantness. NeuroImage. 2012; 61(1):289–94. Epub 2012/03/13. doi: S1053-8119(12)00243-1 [pii] doi: 10.1016/j.neuroimage.2012.02.065 PMID: 22406357.
23. Lebreton M, Jorge S, Michel V, Thirion B, Pessiglione M. An automatic valuation system in the human brain: evidence from functional neuroimaging. Neuron. 2009; 64(3):431–9. Epub 2009/11/17. doi: S0896-6273(09)00751-X [pii] doi: 10.1016/j.neuron.2009.09.040 PMID: 19914190.
24. Shibata K, Watanabe T, Sasaki Y, Kawato M. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. Science. 2011; 334(6061):1413–5. Epub 2011/12/14. doi: 334/6061/1413 [pii] doi: 10.1126/science.1212003 PMID: 22158821; PubMed Central PMCID: PMC3297423.
25. Toda A, Imamizu H, Kawato M, Sato MA. Reconstruction of two-dimensional movement trajectories from selected magnetoencephalography cortical currents by combined sparse Bayesian methods. NeuroImage. 2011; 54(2):892–905. Epub 2010/10/05. doi: S1053-8119(10)01257-7 [pii] doi: 10.1016/j.neuroimage.2010.09.057 PMID: 20884361.
26. Zajonc RB. Attitudinal effects of mere exposures. J Person Soc Psychol. 1968; 9:1–27.

27. Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. Proc Natl Acad Sci U S A. 2006; 103(10):3863–8. doi: 10.1073/pnas.0600244103 PMID: 16537458; PubMed Central PMCID: PMCPMC1383651.

28. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput Biol. 2011; 7(1):e1001057. doi: 10.1371/journal.pcbi.1001057 PMID: 21283776; PubMed Central PMCID: PMCPMC3024255.

29. Tse PU. The neural basis of free will. Cambridge: The MIT Press; 2013.

30. Chen L, Man H, Nefian AV. Face recognition based multi-class mapping of Fisher scores. Pattern Recognition. 2005; 38:799–811.

31. Genki-4d database. http://mplab.ucsd.edu.

32. http://pics.psych.stir.ac.uk/.

33. Phillipes PJ, Wechsler H, Huang J, Rauss P. The FERET database and evaluation procedure for face recognition algorithms. Image and Vision Computing J. 1998; 16(5):295–306.

34. Phillips PJ, Moon H, Rizvi SA, Rauss PJ. The FERET evaluation methodology for face recognition algorithms. IEEE Trans Pattern Analysis and Machine Intelligence. 2000; 22:1090–104.

35. Solina F, Peer P, Batagelj B, Juvan S, Kovac J. Color-based face detection in the '15 seconds of frame' art installation. Mirage 2003; INRIA, France2003. p. 38–47.

36. Goebel R, Esposito F, Formisano E. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Human Brain Mapping. 2006; 27(5):392–401. Epub 2006/04/06. doi: 10.1002/hbm.20249 PMID: 16596654.

37. Frost MA, Goebel R. Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. NeuroImage. 2012; 59(2):1369–81. Epub 2011/08/31. doi: 10.1016/j.neuroimage.2011.08.035 PMID: 21875671.

38. Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, et al. Automatically parcellating the human cerebral cortex. Cereb Cortex. 2004; 14(1):11–22. Epub 2003/12/05. PMID: 14654453.

39. Winston JS, O'Doherty J, Kilner JM, Perrett DI, Dolan RJ. Brain systems for assessing facial attractiveness. Neuropsychologia. 2007; 45(1):195–206. Epub 2006/07/11. doi: 10.1016/j.neuropsychologia.2006.05.009 PMID: 16828125.

40. Amano K, Shibata K, Kawato M, Sasaki Y, Watanabe T. Learning to associate orientation with color in early visual areas by associative decoded fMRI neurofeedback. Current Biology. 2016; 26(14): 1861–6. Epub 2016/6/30. doi: 10.1016/j.cub.2016.05.014 PMID: 27374335.

41. deBettencourt MT, Cohen JD, Lee RF, Norman KA, Turk-Browne NB. Closed-loop training of attention with real-time brain imaging. Nature Neuroscience. 2015; 18(3):470–5. Epub 2015/02/11. doi: 10.1038/nn.3940 PMID: 25664913.

42. Ruiz S, Lee S, Soekadar SR, Caria A, Veit R, Kircher T, et al. Acquired self-control of insula cortex modulates emotion recognition and brain network connectivity in schizophrenia. Human Brain Mapping. 2013; 34(1):200–12. Epub 2011/10/25. doi: 10.1002/hbm.21427 PMID: 22021045.

43. Weiskopf N, Mathiak K, Bock SW, Scharnowski F, Veit R, Grodd W, et al. Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). IEEE Trans Biomed Eng. 2004; 51(6):966–70. Epub 2004/06/11. doi: 10.1109/TBME.2004.827063 PMID: 15188865.

44. Brainard DH. The Psychophysics Toolbox. Spat Vis. 1997; 10(4):433–6. Epub 1997/01/01. PMID: 9176952.