Neural correlates of cognitive dissonance and choice-induced preference change

Keise Izuma[a,1,2], Madoka Matsumoto[a], Kou Murayama[b], Kazuyuki Samejima[a], Norihiro Sadato[c], and Kenji Matsumoto[a]

[a]Brain Science Institute, Tamagawa University, Machida, Tokyo 194-8610, Japan; [b]Department of Psychology, University of Munich, 80802 Munich, Germany; and [c]Department of Cerebral Research, Division of Cerebral Integration, National Institute for Physiological Sciences, Okazaki, Aichi 444-8585, Japan

According to many modern economic theories, actions simply reflect an individual's preferences, whereas a psychological phenomenon called "cognitive dissonance" claims that actions can also create preference. Cognitive dissonance theory states that after making a difficult choice between two equally preferred items, the act of rejecting a favorite item induces an uncomfortable feeling (cognitive dissonance), which in turn motivates individuals to change their preferences to match their prior decision (i.e., reducing preference for rejected items). Recently, however, Chen and Risen [Chen K, Risen J (2010) *J Pers Soc Psychol* 99:573–594] pointed out a serious methodological problem, which casts a doubt on the very existence of this choice-induced preference change as studied over the past 50 y. Here, using a proper control condition and two measures of preferences (self-report and brain activity), we found that the mere act of making a choice can change self-report preference as well as its neural representation (i.e., striatum activity), thus providing strong evidence for choice-induced preference change. Furthermore, our data indicate that the anterior cingulate cortex and dorsolateral prefrontal cortex tracked the degree of cognitive dissonance on a trial-by-trial basis. Our findings provide important insights into the neural basis of how actions can alter an individual's preferences.

attitude | cognitive control | conflict | neuroeconomics | value

In Aesop's Fable "The Fox and the Grapes," a fox tries to get some grapes that are hanging on a high, unreachable vine. After failing to reach them, the fox decides that the grapes were probably sour anyway. An interesting aspect of this story is the idea that actions (e.g., giving up on the grapes) can change preferences. Because the dissonance-induced preference change indicates that behaviors can create, not just reflect, people's preferences, it challenges a vital assumption in neoclassical economics that preference or "hedonic utility" determines people's behavior (1).

Since Brehm's original study in 1956 (2), this sort of preference change (i.e., the increase in ratings for chosen goods and/or the decrease in ratings for rejected goods) has been repeatedly observed under the "free-choice paradigm" (3–6). In a typical free-choice study design, participants are asked to: (*i*) rate their preference for a set of goods (e.g., art prints, CDs, and so forth), (*ii*) choose between two of the goods, and (*iii*) rate them again. After making a difficult choice between two equally preferred items at stage *ii*, individuals tend to like the selected item more and the rejected item less than they originally did (2). This tendency happens because when making a choice between two equally highly preferred items, individuals have to give up either of the two items. According to cognitive dissonance theory (7), simultaneously holding two or more contradictory cognitions (e.g., "I like the item" and "I rejected it") causes a psychological discomfort called "cognitive dissonance," and individuals are motivated to reduce this discomfort by changing their original preferences. Therefore, in this particular case, individuals reduce their preferences for the rejected item so that two cognitions are consistent with each other (e.g., "I don't like the item" and "I rejected it"; similarly, when a choice was made between two equally unpreferred items, choosing a disliked item induces cognitive dissonance, and individuals in-

crease their preferences for the chosen item to reduce it). The existence of choice (dissonance)-induced preference change is further supported by a neuroimaging study showing that the activity in the anterior striatum, a region implicated in reward processing, is correlated with the choice-induced preference change (8).

However, a serious methodological problem in the free-choice paradigm has been recently pointed out by Chen and Risen (9). Their critique raises the possibility that the choice-induced preference change reported in studies that have used the free-choice paradigm over the past 50 y, including a recent neuroimaging study (8; see also ref. 10), might be just a methodological artifact.

Chen and Risen (9) argue that the free-choice paradigm measures ostensible preference change, even if the individual's true preferences remain completely unchanged (i.e., even in the absence of cognitive dissonance). Their arguments are based on two assumptions. First, individuals' ratings contain at least some noise, and the ratings are not perfect measures of one's true preferences. Second, individuals' choices are at least partially guided by their true preferences, and thus their choices reveal some information about their true preferences. Thus, even if two goods are rated equally in self-reports, it does not necessarily mean that true preferences for these two goods are also exactly the same (first assumption; similarly, even if one good is rated higher than the other, true preference might be the opposite). Accordingly, when choices were made between two goods that have the same (or similar) initial ratings, true preference for a chosen good is likely to be higher than true preference for a rejected good (second assumption). Then, when participants are asked to rate the same goods again, it is more likely that their rating for a chosen good would increase and that for a rejected good would decrease, on average, than vice versa (for complete discussion and mathematical proof, see ref. 9).

Chen and Risen (9) also provided experimental data to support their claims. Whereas the typical free-choice paradigm follows a "rate, choose, and rate" order, the new Chen and Risen experiments include a control condition in which participants made a choice after they had made two preference ratings (thus, they followed a "rate, rate, and choose" order). In both conditions, items were categorized as "chosen" or "rejected" according to the subject's choice at each choice stage, and preference changes for chosen items and rejected items from the first to the second ratings were calculated. Because subjects made their choice at the end of the experiment in the rate-rate-choose condition, any changes between the two preference ratings can never be attrib-

uted to choice-induced cognitive dissonance. Nevertheless, both conditions are expected to produce preference change in a manner that is consistent with the past cognitive dissonance literature (i.e., increased preference for chosen and decreased preference for rejected items), because participants' choices contain information about their true preferences, regardless of when the choice is made. In other words, because the effect of the information revealed by choice is equalized between the rate-choose-rate and rate-rate-choose conditions, and the only difference between the conditions is whether subjects made their choice before or after the second rating session, the effect of choice-induced cognitive dissonance can be isolated by comparing preference changes in these two conditions. Consistent with their prediction, Chen and Risen (9) found that there was a significant preference change even in the rate-rate-choose condition, indicating that the preference change measured in a typical free-choice paradigm can occur in the absence of cognitive dissonance. When compared with this rate-rate-choose condition, subjects who followed the typical free-choice procedure (i.e., the rate-choose-rate condition) showed a tendency for more preference change, but it did not reach statistical significance (9). Thus, after more than 50 y of research, the evidence for choice-induced preference change still remains inconclusive.

Thus, the primary purpose of the present study was to investigate the existence of choice-induced preference change with proper control to eliminate this methodological artifact (9). We used two independent measures of preference: self-report and brain activity. By using a neural measure of preference, it was possible to further test whether choice-induced preference change is just a superficial phenomenon seen only in the self-report, or whether choices can alter not only self-report preference but also its neural representation. Toward this goal, as in a previous neuroimaging study (8), we focused on the striatum as our region of interest (ROI). It is well established that the striatum, especially its anterior part (e.g., caudate head and nucleus accumbens), tracks individuals' reported preference for various types of stimuli (8, 11–14).

In the current experiment, subjects inside the functional MRI (fMRI) scanner were first asked to rate their preference for 160 food items (Preference task 1) (Fig. 1A). Following Preference task 1, subjects made choices between pairs of foods (Self trials in the Choice task) (Fig. 1B, Upper) which varied systematically so that choices were sometimes made between two equally liked items (Self-Difficult trials), and other times between one liked item and one disliked item (Self-Easy trial). In still other trials, choices were made randomly by a computer between two equally liked items (Computer trials) (Fig. 1B, Lower). Following the Choice task, subjects rated their preferences for the same 160 foods again. Importantly, in this Preference task 2 (Fig. 1C), subjects were also presented with information about the previous decisions made in the Choice task (e.g., "You rejected it"; Fig. 1C). This process was intended to make subjects perceive the inconsistency (or consistency) between their preferences and past decisions for each food item. Our critical trials were those in which subjects perceived that they had rejected their liked foods (rejected items in Self-Difficult trials of the Choice task); according to the cognitive dissonance theory, their preferences for these foods were predicted to be reduced (7). In other trials in which subjects rejected disliked foods or chose preferred foods, less dissonance would be induced (therefore, little or no preference change was expected) because choosing preferred foods and rejecting unpreferred foods are more consonant than rejecting liked foods. Moreover, for items chosen or rejected randomly by a computer, no dissonance was expected because subjects were not responsible for the choices (2). Finally, to control information about subjects' preferences as revealed by their choices and to isolate the effect of choice-induced preference change, we had subjects perform a Post-Experimental
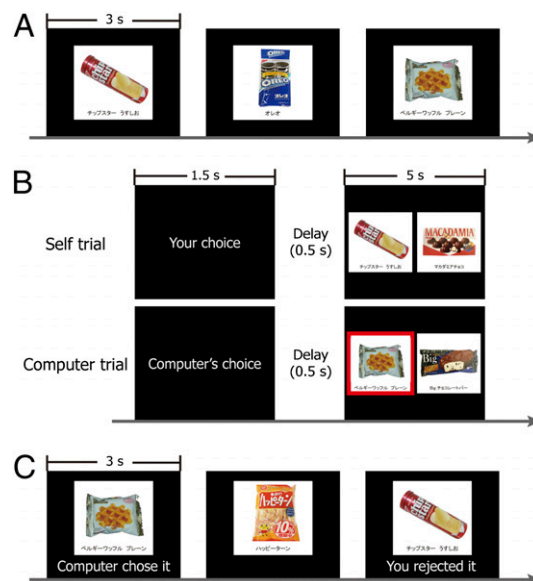
Fig. 1. Experimental tasks. (A) Preference task 1. In each trial, subjects were presented with a food item for 3 s and asked to rate their preference for it on an eight-point scale. (B) Examples of a Self trial and a Computer trial during the Choice task. In a Self trial, subjects were asked to choose one of two foods which they preferred. In a Computer trial, they were asked to simply pick the item that a computer randomly chose (highlighted by a red square). A cue was presented for 1.5 s followed by a brief delay (500 ms), and two foods were presented for 5 s. (C) Preference task 2. Subjects were presented with the same food stimuli as in Preference task 1 and asked to rate their preference again. However, during Preference task 2, past decisions by themselves or a computer (e.g., "You rejected it", etc.) were also displayed under each picture. For items not used in the Choice task, only pictures of foods were presented.

Choice (PostEx-Choice) task outside the fMRI scanner. This PostEx-Choice task creates an analogous condition to the rate-rate-choose condition used in Chen and Risen (9). During this task, subjects were presented with the same food pairs used in the Computer trials of the Choice task and asked to select the one that they preferred. As subjects had not been responsible for the choices during the Computer trials, these choices provided additional information about their true preferences for the items. Because food pairs for the Computer trials were determined similarly to the Self-Difficult trials (for both conditions, pairs consist of two equally preferred foods; Methods), choices subjects made in the Self-Difficult and the PostEx-Choice trials were equally informative about how much preference change could be expected, even in the absence of cognitive dissonance.

In addition to the investigation of choice (cognitive dissonance)-induced preference change, the present study also investigated the neural correlates of cognitive dissonance itself. Despite its significant influences in the field of psychology (15), the neural processes underlying cognitive dissonance are still not fully understood. We chose the anterior cingulate cortex (ACC) and anterior insula as our ROIs. The involvement of the ACC in cognitive dissonance has been predicted by Harmon-Jones (16, 17). Recently, van Veen et al. (18) adapted the induced compliance paradigm (19) to the fMRI scanner and found that the ACC and anterior insula are involved in cognitive dissonance. In the present study, we sought to extend their findings (18) and test whether the ACC and anterior insula track subjects' perceived dissonance on a trial-by-trial basis. In the current free-choice paradigm, it is possible to quantify the degree of dissonance in each trial by calculating the discrepancy between subjects' past decisions (i.e., chose or rejected) and their preferences for each food item (Methods).

Results

Behavioral Results. First, our data revealed that preference for those foods which were rejected in the PostEx-Choice condition significantly decreased compared with rejected foods in the Self-Easy condition [$t(19) = 7.84$, $P < 0.001$] or those rejected in the Computer condition [$t(19) = 4.60$, $P < 0.001$] (Fig. 2). This finding indicates that a typical free-choice paradigm would produce preference change even in the absence of cognitive dissonance, as suggested by Chen and Risen (9). However, our data also showed that preference for rejected foods in the Self-Difficult condition dropped even further, and the change was significantly lower than preference change observed for rejected items in the PostEx-Choice condition [$t(19) = 1.91$, $P = 0.036$] (Fig. 2), thus providing evidence for choice-induced preference change. On the other hand, for the chosen foods, while preference changes for chosen foods in Self-Difficult conditions were significantly higher than those in Self-Easy and Computer trials (both $P < 0.01$), there was no significant difference in preference change between chosen foods in Self-Difficult and PostEx-Choice conditions ($P = 0.41$, n.s.). Preference changes observed in chosen items in PostEx-Choice task were also significantly higher than chosen items in Self-Easy and Computer conditions (both $P < 0.01$). Although we did not find a significant choice-induced preference increase for chosen items, these results are not surprising, given that the act of choosing preferred items does not induce strong dissonance (see *SI Results* and Fig. S1 for additional behavioral results).

Imaging Results. We first sought to identify areas within the anterior striatum that encode subjective preferences for foods. We conducted a parametric modulation analysis using subjects' preference and familiarity ratings as covariates [GLM 1; see *Methods* for details about all general linear models (GLMs)]. This model was estimated using the data from Preference task 1 only so that any changes in self-report preferences from the first to second Preference tasks remain independent from changes in striatal activity. The result revealed that there were three clusters within the anterior striatum in which activity was parametrically modulated by subjects' reported preference (Fig. 3A). Other than the striatum, activations were seen along almost the entire cingulate cortex (from anterior to posterior) (Fig. S2), and all activated regions are listed in Table S1. Because some activated clusters in the striatum extended to the outside of the striatum [e.g., ventromedial prefrontal cortex (vmPFC) and globus pallidus], the anatomical mask [the WFU PickAtlas toolbox for SPM (20)] was used to limit our ROI within the anterior part of
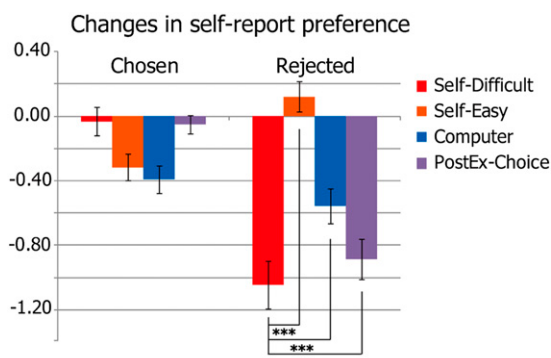
Fig. 2. Behavioral result for preference change. Bars indicate the change in preference for foods from Preference task 1 to Preference task 2 (preference ratings in Preference task 2 minus those in Preference task 1) in each condition. *$P < 0.05$; ***$P < 0.001$ (paired t test, one tailed). Error bars indicate the SEM.
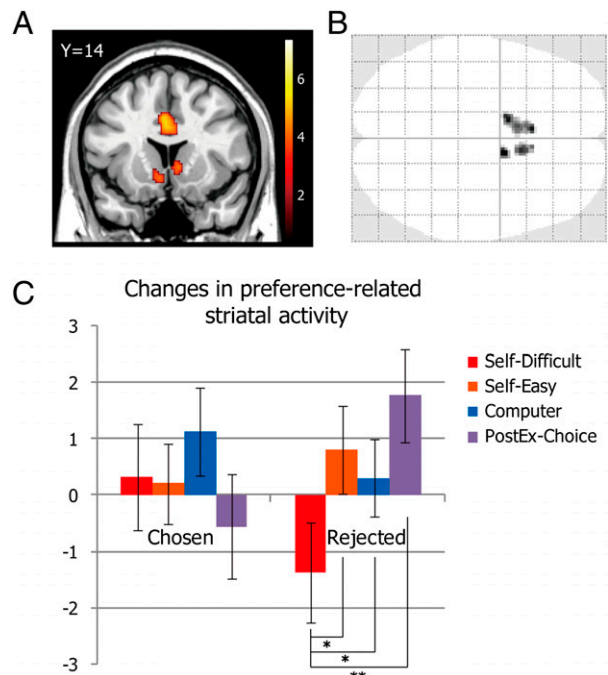
Fig. 3. (A) A coronal slice showing areas in the anterior striatum significantly correlated with subjects' self-reported preference during Preference task 1 ($P < 0.001$ uncorrected). The scale shows the t values. (B) An axial grass brain showing the ROI in the anterior striatum. The striatal ROI consists of three separate clusters, a total of 189 voxels. (C) Percent-signal changes averaged over the whole striatum ROI (189 voxels) were extracted, and activation changes in the striatal ROI (striatal activations in Preference task 2 minus those in Preference task 1) were plotted for each condition. *$P < 0.05$; **$P < 0.01$ (paired t test, one-tailed). Error bars indicate SEM.

the striatum (see *SI Methods* for detail). Our striatum ROI, then, consists of three clusters in the anterior ventral striatum with a volume of 189 voxels in total (Fig. 3B).

To test whether choice-induced preference changes were also observed as changes in the preference-related brain activity, we averaged the percent-signal changes in response to food presentation over the whole striatum ROI (189 voxels) (Fig. 3B) for all conditions (GLM 2; note that the data for chosen and rejected items in the PostEx-Choice task were extracted from GLM 3). Fig. 3C plots the activity changes (Preference task 2 minus Preference task 1) in the striatum ROIs for all conditions. Consistent with the behavioral results, the change in striatal activity for rejected items in Self-Difficult trials were significantly lower than all three control conditions [vs. the Rejected-Self-Easy condition, $t(19) = 1.75$, $P = 0.048$; vs. the Rejected-Computer condition, $t(19) = 2.33$, $P = 0.016$; vs. the Rejected-PostEx-Choice condition, $t(19) = 2.56$, $P = 0.01$]. In contrast to the behavioral results, the change in striatal activity for rejected items in the PostEx-Choice condition was actually positive. Although this preference change in the Rejected-PostEx-Choice condition was greater than rejected items in both the Self-Easy and Computer conditions, the differences were not significant (both $P > 0.11$, n.s.). The difference between Rejected-Self-Difficult and Rejected-PostEx-Choice conditions remains significant even when the β-values for both conditions were estimated from the same regression model [GLM 3; $t(19) = 2.54$, $P = 0.01$]. Percent-signal changes (β-values) for seven conditions each for Preference tasks 1 and 2 (a total of 14 conditions) are plotted in Fig. S3.

For chosen items, striatal activity changes in the Chosen-Self-Difficult condition did not differ significantly from the other three control conditions (all $P > 0.28$, n.s.). In addition, among

other areas correlated with subjects' reported preferences (Table S1), no area showed the predicted patterns of preference change as seen in the striatum.

Next, we investigated the neural correlates of cognitive dissonance. To test whether the activity in the ACC and anterior insula is parametrically modulated by perceived cognitive dissonance on a trial-by-trial basis (GLM 4), we computed the degree of cognitive dissonance for each trial as the Cognitive Dissonance Index (CDI). The CDI is defined as the function of the discrepancy between subjects' past decisions (i.e., chose or rejected) and their preferences for each food item (see *Methods* and Table S2 for the detail about the CDI). A higher CDI indicates greater discrepancy and thus higher cognitive dissonance. Using CDI, preference, familiarity, and reaction time (RT) for each trial as covariates, we conducted a parametric modulation analysis on data from both Preference tasks 1 and 2. To exclude the possibility that CDIs simply reflect lower-level perceptual features of stimuli (e.g., size, colorfulness, and so forth), we explored the areas positively correlated with CDIs during Preference task 2 within the areas not correlated with CDIs during Preference task 1 (i.e., outside the areas correlated with CDIs during Preference task 1, with a more liberal threshold of $P < 0.05$, uncorrected). Results showed, as predicted, that the dorsal ACC (dACC) (Fig. 4A) was significantly positively correlated with the CDIs only after choices were made (i.e., during Preference task 2), but not before choices (i.e., during Preference task 1) (Fig. 4B). Other activated areas (Fig. S4 and Table S3) included bilateral dorsolateral prefrontal cortex (DLPFC), which is an area also reported to be involved in cognitive dissonance (21, 22). In contrast to the dACC, anterior insula activity was found only when we lowered the threshold to uncorrected $P < 0.005$ (left anterior insula; $x = -32, y = 20, z = -2$, 39 voxels). There was no area whose activity was negatively correlated with CDIs during Preference task 2 (see *SI Results*, Fig. S5, and Table S4 for results of brain-behavior correlations).

We also confirmed that, even when the CDIs for chosen and rejected items were modeled separately (GLM 5), the same dACC area was significantly correlated with CDIs for both categories (both $P < 0.01$), indicating that the dACC activation depends on the discrepancy per se, regardless of what the subject's past behavior was. All other areas listed in Table S3 were also significantly correlated with both CDIs for chosen (all $P < 0.05$) and rejected items (all $P < 0.01$), except for the right primary visual cortex, which was not correlated with CDIs for chosen items ($P = 0.061$, n.s.).
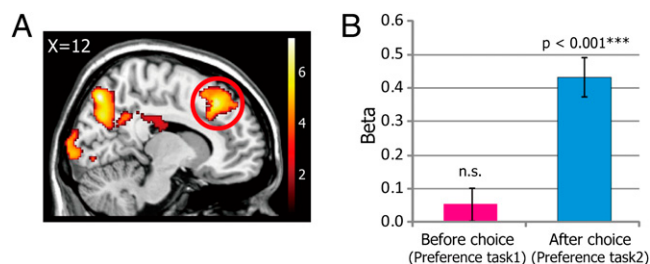
Fig. 4. (A) A sagittal slice showing an area in the dACC significantly correlated with the degree of cognitive dissonance (CDI) during Preference task 2, but not during Preference task 1 ($P < 0.001$ uncorrected). The scale shows the *t* values. (B) β-Values extracted from the peak voxel at the dACC ($x = 12$, $y = 30$, $z = 46$) for Preference task 1 and Preference task 2 (from GLM 4). Note that exactly the same CDIs for the same food items were modeled in both Preference tasks 1 and 2, but CDIs were significantly correlated with the dACC activity only during Preference task 2. ***$P < 0.001$ (one-sample *t* test, one-tailed). Error bars indicate SEM.

Discussion

Although the point raised by Chen and Risen (9) casts serious doubt on the existence of choice-induced preference changes, our data demonstrate that the mere act of rejecting favorite goods actually reduces preferences for them. While our results first confirmed the validity of Chen and Risen (9) and showed that the typical free-choice paradigm can measure preference changes even in the absence of cognitive dissonance, our results also showed that preferences for those items rejected during difficult choices were significantly reduced, even after controlling for the effect of the information about subjects' preference revealed by their choices. Moreover, in the present study, this choice-induced preference change was observed not only in the self-report measure of preference, but also in the brain-imaging measure of preference, namely the activity in the anterior ventral striatum. Thus, the present study extends previous behavioral studies in psychology and indicates that choice-induced preference change is not a superficial phenomenon seen only in self-report measures, but is observable as changes in brain activation. Taken together, our results provide strong evidence that actions not only reflect, but indeed create, preferences.

Past behavioral research on the cognitive dissonance has generated unique findings that cannot be easily explained by economic theory (23), and the current neural evidence supporting the theory highlights this point even further. Because brain activation measurements are much more difficult for subjects to fake (e.g., less susceptible to a demand characteristic effect) compared with self-report measures of preference, the present study provides strong evidence that choice can actually alter an individual's hedonic utility. This result poses a challenge to the neoclassical economics view, which assumes that hedonic utility determines an individual's behavior, but not vice versa (1), suggesting that the relationship between hedonic utility and behaviors is more complex than assumed in economics. Also, it should be noted that while there were other regions correlated with subjects' self-reported preferences (Fig. S2 and Table S1), the striatum was the only region showing the predicted pattern of preference change. This finding might be because other regions, such as the vmPFC, are also related to other variables inherent to stimuli (e.g., arousal), as suggested by Sharot et al. (8). However, as exact roles played by striatum and vmPFC in a valuation process are still debated (24), it would be interesting to see if, for example, preference change seen in other paradigms of cognitive dissonance is reflected not only in the striatum but also in other regions. Perhaps future research could address this question.

We also demonstrated the involvement of the dACC in cognitive dissonance, which supports the hypothesis by Harmon-Jones (16, 17). Although a previous study reported the involvement of the dACC in cognitive dissonance (18), the present study extended it in a different paradigm. We showed that dACC activity reflects the degree of cognitive dissonance (CDI) on a trial-by-trial basis, thus providing more solid evidence that dACC activity is a neural correlate of cognitive dissonance. Our result demonstrated that dACC activation in each trial during Preference task 2 depends on the degree of discrepancy between subjects' past behaviors and their reported preferences for foods. This result matches well with Festinger's original idea that the magnitude of cognitive dissonance depends on the degree of discrepancy between behavior and belief (or cognition) (7). In contrast to the dACC, the evidence for anterior insula involvement in cognitive dissonance is limited in the present study. We found left insula activity only when lowering the threshold to uncorrected $P < 0.005$, while other brain areas showed strong correlations with cognitive dissonance.

Among other areas showing dissonance-related activity, especially notable is the DLPFC. The DLPFC is known to be involved

in the implementation of control or behavioral adjustment following the experience of conflict (25–30). Interestingly, the DLPFC, especially on the left side, has been reported to be involved in cognitive dissonance, especially its reduction, in previous EEG studies (21, 22). The idea that the DLPFC is involved in conflict resolution rather than conflict monitoring seems to be consistent with the fact that we found stronger DLPFC activations in the present study than was reported previously (18). Whereas subjects had a chance to resolve perceived cognitive dissonance by explicitly stating their new preferences in the present study, there was no such chance during the fMRI scanning in the previous study. In other words, it might be the case that while van Veen et al. (18) identified brain areas involved in passive emotional reaction to perceived cognitive dissonance (i.e., dACC and anterior insula), brain areas identified in the present study (i.e., dACC and DLPFC) are involved in the active dissonance reduction process following dissonance perception. Nonetheless, these two studies converge to show that the dACC plays an important role in cognitive dissonance; the exact roles played by other areas in the cognitive dissonance process should be further investigated in future research.

In conclusion, the present study identified the neural mechanism underlying the psychological processes explained by a classic psychological theory of cognitive dissonance (7). We found the evidence that the mere act of making choices can alter individuals' self-report preferences as well as its neural representation. Our data also suggest that this dissonance-induced preference change recruits the same neural network underlying conflict monitoring and subsequent implementation of control. Whereas past research in neuroscience has focused on the neural mechanisms underlying bottom-up modulation of value by prediction error signal based on choice outcome (reinforcement learning) (31, 32), choice-induced preference change in cognitive dissonance is an example of top-down cognitive modulation of value based on choice itself. The present results suggest that in cognitive dissonance, the ACC and DLPFC play important roles in modulating value signals in the striatum.

Methods

Subjects. The reported analyses were based on 20 subjects (10 male; age range, 18–24 y). None of the subjects had a history of neurological or psychiatric illness. All subjects gave written informed consent for participation, and the study was approved by the Ethics Committee of Tamagawa University, Japan.

Experimental Tasks and Procedures. The present experiment consisted of four parts: (*i*) Preference task 1, (*ii*) Choice task, (*iii*) Preference task 2, and (*iv*) Post-Experimental Choice task. Except for the Post-Experimental Choice task, all tasks were performed inside the fMRI scanner (see *SI Methods* for a more detailed description of the experimental tasks and procedures).

During Preference task 1 (Fig. 1*A*), subjects were instructed to rate their preference for the food presented on the screen using an eight-point scale (1 = do not like it at all, 8 = like it very much). In the Choice task, there were two types of trials: the Self trial and the Computer trial (Fig. 1*B*). During a Self trial (Fig. 1*B, Upper*), two foods were presented on the screen at the same time and subjects were instructed to choose one food that they prefer. Unknown to the subjects, food pairs were determined by the Matlab program such that about half of Self trials included two preferred food items (i.e., preference ratings of 5 or more) and rated similarly in Preference task 1 (i.e., difference in preference ratings was either 0 or 1) (these were Self-Difficult trials). The other half of the Self trials included one preferred food and one unpreferred food (i.e., preference ratings of 4 or less), and the two foods differed in preference by at least three points in ratings (these were Self-Easy trials). During a Computer trial (Fig. 1*B, Lower*), subjects were asked to simply press the button corresponding to the food a computer had randomly chosen (highlighted by a red square). For the Computer trials, food pairs were determined similarly to the Self-Difficult trials so that the Computer trials always included two equally preferred foods. Each food appeared in only one pair.

In Preference task 2, the same 160 foods were presented on the screen, and subjects were asked to rate their preference for each food again. However,

one important difference was that for the foods included in pairs presented during the Choice task, subjects' (or computer's) past decisions during the Choice task (e.g., "You chose it," "You rejected it," "Computer chose it," or "Computer rejected it") were presented under each food (Fig.1*C*). For food stimuli that were not used in the Choice task, only a picture of the food was presented on the screen the same as in Preference task 1.

After Preference task 2, subjects performed the PostEx-Choice task outside the fMRI scanner. Subjects were asked to choose the one they preferred from the same pairs of foods that had appeared during the Computer trials of the Choice task. Finally, subjects rated the familiarity of each of 160 foods using the eight-point scale.

Functional MRI Data Acquisition and Analysis. Functional MRI data acquisition and preprocessing were carried out using standard procedure described in *SI Methods*. Because we were interested in how preference-related striatal activity might change from Preference task 1 and 2 and whether the activity in the ACC and anterior insula tracks the degree of cognitive dissonance during Preference task 2, we analyzed the fMRI data from Preference tasks 1 and 2 only.

We used five GLMs to analyze the fMRI data. The first three GLMs were used to test the hypotheses that dissonance-induced preference changes are seen as changes in striatal activity. The remaining two GLMs were used to test the involvement of the ACC and anterior insula in cognitive dissonance.

GLM 1 was intended to identify regions especially in the anterior striatum whose activity was positively correlated with subjects' reported preferences for each food. The model was estimated only on Preference task 1. The model included three regressors: (*i*) each food stimulus onset, (*ii*) stimulus onset modulated by subjects' reported preference for each food stimulus, and (*iii*) stimulus onset modulated by reported familiarity for each food stimulus.

GLM 2 was designed to test the hypotheses about preference change in each condition. This model was estimated using the data from both Preference tasks 1 and 2. For both Preference tasks, trials were classified into seven conditions according to subjects' choices during the Choice task: (*i*) Chosen items in Self-Difficult trials (Chosen-Self-Difficult), (*ii*) Rejected items in Self-Difficult trials (Rejected-Self-Difficult), (*iii*) Chosen items in Self-Easy trials (Chosen-Self-Easy), (*iv*) Rejected items in Self-Easy trials (Rejected-Self-Easy), (*v*) Chosen items in Computer trials (Chosen-Computer), (*vi*) Rejected items in Computer trials (Rejected-Computer), and (*vii*) Not Used trials (i.e., trials in which items not included in food pairs during the Choice task were presented). On rare occasions (6.54%), subjects chose unpreferred food (i.e., the one that had the lower preference rating during Preference task 1) over the preferred one in Self-Easy trials of the Choice task. Because underlying psychological processes are unclear for these items, these trials were also modeled as Not Used trials (note that all food items in Self-Difficult trials were categorized into "Chosen" or "Rejected" items, even if subjects chose the food that had a lower preference rating than the alternative). These seven conditions were modeled separately for the first and second Preference tasks, generating 14 different regressors in total.

GLM 3 was the same as GLM 2, except that food items in Computer trials were now classified according to subjects' own choices during the PostEx-Choice task. Thus, GLM 3 included 14 regressors, seven regressors each for Preference tasks 1 and 2: (*i*) Chosen-Self-Difficult, (*ii*) Rejected-Self-Difficult, (*iii*) Chosen-Self-Easy, (*iv*) Rejected-Self-Easy, (*v*) Chosen items in the PostEx-Choice task (Chosen-PostEx-Choice), (*vi*) Rejected items in the PostEx-Choice task (Rejected-PostEx-Choice), and (*vii*) Not Used items (the same as GLM 2).

GLM 4 was designed to identify the regions whose activity was parametrically related to perceived cognitive dissonance. We first computed the CDI for each food (Table S2). According to the cognitive dissonance theory, the magnitude of cognitive dissonance depends on the degree of discrepancy between the two cognitions (7, 23). In the present study, subjects' perceived dissonance in each trial of Preference task 2 depended on their decisions during the choice task (i.e., chosen or rejected by themselves or a computer) and their preferences for the food. For example, for the foods they had rejected, the higher the preference for the food, the more dissonance would be induced because subjects' behavior of rejecting contradicted more with their cognition (high preference for it). In other words, rejecting liked foods is more dissonant than rejecting disliked foods. Therefore, the degree of cognitive dissonance (CDI) for the Rejected items in Self-Difficult and Self-Easy trials is considered to be positively proportional to preference for the items as follows:

$$CDI(\text{Rejected-Self items}) = \text{preference of the item}$$

On the other hand, for the foods they had chosen, the higher the preference for the food, the less dissonance would be induced because choosing favorite foods is more consonant than choosing unfavorite foods. Thus, CDI for the

Chosen items in Self-Difficult and Self-Easy trials is considered to be negatively proportional to preference for the item as follows:

$$\mathrm{CDI}(\text{Chosen-Self items}) = 9 - \text{preference of the item}$$

Therefore, both CDIs for chosen and rejected items have the possible range of 1 to 8 (Table S2). Also, if subjects were not responsible for the choice of a food (i.e., Computer trials), no dissonance would be induced. Thus, CDI for the foods in Computer trials and those not presented during the choice task (Not Used items) was equal to 0, regardless of subjects' preferences for them.

This model was estimated using the data from both Preference tasks 1 and 2. The model included one regressor for each food stimulus onset and the following four parametric modulators: (*i*) CDI for each food, (*ii*) preference for each food, (*iii*) familiarity for each food, and (*iv*) RT in each trial. These five regressors were modeled separately for Preference tasks 1 and 2, generating 10 regressors in total. It should be noted that we used preference ratings during Preference task 1 to compute each subject's CDIs, and therefore, exactly the same CDI was assigned for each food item between Preference tasks 1 and 2. In addition, because RTs tend to correlate with response conflict (27, 28, 33), subjects' RTs in each trial were included in the GLM as a parametric regressor to partial out the potential effect of such lower-level conflicts.

We created another model (GLM 5) to test the robustness of the ACC's involvement in cognitive dissonance. Although in GLM 4 it was assumed that CDIs for chosen and rejected items have the same sensitivity to self-reported preference (e.g., CDIs for chosen items increase by 1 when preference ratings decrease by 1, and similarly, CDIs for rejected items increase by 1 when preference ratings increase by 1), they might differ in such sensitivity (e.g., CDIs for rejected items decrease by 2 when preference ratings increase by 1).

Also, there is no definite reason to assume that rejecting items with a preference rating of 4 induces the same magnitude of dissonance as choosing items with a preference rating of 5 (in both cases, CDIs = 4) (Table S2). GLM 5 was the same as GLM 4, except that chosen foods and rejected foods are now modeled separately. Because there are no chosen or rejected trials for Not Used items, these items were randomly divided into halves and then allocated to either "chosen" or "rejected."

For all five GLMs, the regressors were calculated using a box-car function convolved with a hemodynamic-response function. Regressors that were of no interest, such as the session effect, and high-pass filtering (128 s) were also included. Furthermore, for GLMs 1, 4, and 5, missed trials (1.95% and 2.25% for Preference tasks 1 and 2, respectively) were modeled separately as a regressor of no interest. For GLMs 2 and 3, missed trials were classified into Not-Used items.

For a priori ROIs (anterior striatum, ACC, and anterior insula), a statistical threshold was set at $P < 0.001$ (uncorrected for multiple comparisons) with an extent threshold of 20 contiguous voxels. Activations outside the ROIs were reported if they exceeded a threshold of $P < 0.001$ (uncorrected) and cluster $P < 0.05$ (corrected for multiple comparisons).

1. Ariely D, Norton MI (2008) How actions create—not just reveal—preferences. *Trends Cogn Sci* 12:13–16.
2. Brehm JW (1956) Post-decision changes in the desirability of choice alternatives. *J Abnorm Soc Psychol* 52:384–389.
3. Coppin G, Delplanque S, Cayeux I, Porcherot C, Sander D (2010) I'm no longer torn after choice: How explicit choices implicitly shape preferences of odors. *Psychol Sci* 21: 489–493.
4. Gerard HB, White GL (1983) Post-decisional reevaluation of choice alternatives. *Pers Soc Psychol Bull* 9:365–369.
5. Lieberman MD, Ochsner KN, Gilbert DT, Schacter DL (2001) Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychol Sci* 12:135–140.
6. Shultz TR, Leveille E, Lepper MR (1999) Free choice and cognitive dissonance revisited: Choosing "lesser evils" versus "greater goods". *Pers Soc Psychol Bull* 25:40–48.
7. Festinger L (1957) *A Theory of Cognitive Dissonance* (Stanford University Press, Stanford).
8. Sharot T, De Martino B, Dolan RJ (2009) How choice reveals and shapes expected hedonic outcome. *J Neurosci* 29:3760–3765.
9. Chen MK, Risen JL (2010) How choice affects and reflects preferences: Revisiting the free-choice paradigm. *J Pers Soc Psychol* 99:573–594.
10. Jarcho JM, Berkman ET, Lieberman MD (2010) The neural basis of rationalization: Cognitive dissonance reduction during decision-making. *Soc Cogn Affect Neurosci*, in press.
11. Berns GS, Capra CM, Moore S, Noussair C (2010) Neural mechanisms of the influence of popularity on adolescent ratings of music. *Neuroimage* 49:2687–2696.
12. Knutson B, Rick S, Wimmer GE, Prelec D, Loewenstein G (2007) Neural predictors of purchases. *Neuron* 53:147–156.
13. Knutson B, et al. (2008) Neural antecedents of the endowment effect. *Neuron* 58: 814–822.
14. Lebreton M, Jorge S, Michel V, Thirion B, Pessiglione M (2009) An automatic valuation system in the human brain: Evidence from functional neuroimaging. *Neuron* 64: 431–439.
15. Harmon-Jones E, Mills J (1999) *Cognitive Dissonance: Progress on a Pivotal Theory in Social Psychology* (Braun-Brumfield, Washington, DC).
16. Harmon-Jones E (2004) Contributions from research on anger and cognitive dissonance to understanding the motivational functions of asymmetrical frontal brain activity. *Biol Psychol* 67:51–76.
17. Harmon-Jones E, Harmon-Jones C (2008) Action-based model of dissonance: A review of behavioral, anterior cingulate, and prefrontal cortical mechanisms. *Soc Pers Psychol Compass* 2:1518–1538.
18. van Veen V, Krug MK, Schooler JW, Carter CS (2009) Neural activity predicts attitude change in cognitive dissonance. *Nat Neurosci* 12:1469–1474.
19. Festinger L, Carlsmith JM (1959) Cognitive consequences of forced compliance. *J Abnorm Psychol* 58:203–210.
20. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19:1233–1239.
21. Harmon-Jones E, Gerdjikov T, Harmon-Jones C (2008) The effect of induced compliance on relative left frontal cortical activity: A test of the action-based model of dissonance. *Eur J Soc Psychol* 38:35–45.
22. Harmon-Jones E, Harmon-Jones C, Fearn M, Sigelman JD, Johnson P (2008) Left frontal cortical activation and spreading of alternatives: Tests of the action-based model of dissonance. *J Pers Soc Psychol* 94:1–15.
23. Cooper J (2007) *Cognitive Dissonance: 50 Years of a Classic Theory* (Sage Publications, London).
24. Kable JW, Glimcher PW (2009) The neurobiology of decision: Consensus and controversy. *Neuron* 63:733–745.
25. Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. *Psychol Rev* 108:624–652.
26. Botvinick MM, Cohen JD, Carter CS (2004) Conflict monitoring and anterior cingulate cortex: An update. *Trends Cogn Sci* 8:539–546.
27. Kerns JG, et al. (2004) Anterior cingulate conflict monitoring and adjustments in control. *Science* 303:1023–1026.
28. MacDonald AW, 3rd, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288:1835–1838.
29. Mansouri FA, Tanaka K, Buckley MJ (2009) Conflict-induced behavioural adjustment: A clue to the executive functions of the prefrontal cortex. *Nat Rev Neurosci* 10: 141–152.
30. Matsumoto K, Tanaka K (2004) Neuroscience. Conflict and cognitive control. *Science* 303:969–970.
31. Dayan P, Niv Y (2008) Reinforcement learning: The good, the bad and the ugly. *Curr Opin Neurobiol* 18:185–196.
32. Kawato M, Samejima K (2007) Efficient reinforcement learning: Computational theories, neuroscience and robotics. *Curr Opin Neurobiol* 17:205–212.
33. Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD (1999) Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* 402:179–181.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

NEUROSCIENCE