

A Causal Role for Posterior Medial Frontal Cortex in Choice-Induced Preference Change

Keise Izuma,^{1,2} Shyam Akula,³ Kou Murayama,⁴ Daw-An Wu,¹ Marco Iacoboni,⁵ and Ralph Adolphs¹

¹Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125, ²Brain Science Institute, Tamagawa University, Machida, Tokyo 194-8610, Japan, ³Departments of Genetics and Psychiatry, Washington University in St. Louis, St. Louis, Missouri 63110, ⁴Department of Psychology, University of Reading, Whiteknights Reading RG6 6AL, United Kingdom, and ⁵Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California 90095

After a person chooses between two items, preference for the chosen item will increase and preference for the unchosen item will decrease because of the choice made. In other words, we tend to justify or rationalize our past behavior by changing our attitude. This phenomenon of choice-induced preference change has been traditionally explained by cognitive dissonance theory. Choosing something that is disliked or not choosing something that is liked are both cognitively inconsistent and, to reduce this inconsistency, people tend to change their subsequently stated preference in accordance with their past choices. Previously, human neuroimaging studies identified posterior medial frontal cortex (pmFC) as a key brain region involved in cognitive dissonance. However, it remains unknown whether the pmFC plays a causal role in inducing preference change after cognitive dissonance. Here, we demonstrate that 25 min, 1 Hz repetitive transcranial magnetic stimulation applied over the pmFC significantly reduces choice-induced preference change compared with sham stimulation or control stimulation over a different brain region, demonstrating a causal role for the pmFC.

Key words: attitude; choice justification; cognitive dissonance; medial prefrontal cortex; preference change; TMS

Introduction

Contrary to a basic economic view that our behavior is driven by our preferences, psychological studies have demonstrated that past behaviors (e.g., making a choice) can also affect our preferences. The phenomenon of choice-induced preference change has been traditionally demonstrated by the “free-choice paradigm” (Brehm, 1956) in which individuals are first asked to rate several items for their preference and then make choices between pairs of equally attractive items (e.g., posters, CDs, etc.) so that they inevitably have to reject items they may like. When asked to rate the same items a second time, people typically report increased preference for the items they chose and decreased preference for the items they rejected. This phenomenon is usually explained as an effect of cognitive dissonance. According to cognitive dissonance theory (Festinger, 1957), any inconsistency between beliefs, preferences, or choices (e.g., not choosing a preferred item) results in an aversive feeling (cognitive dissonance) that in turn motivates people to reconcile the inconsis-

teny (i.e., to justify their past behavior by changing their preference).

More recently, social cognitive neuroscientists have started to investigate the neural mechanisms underlying cognitive dissonance and subsequent preference change and found that a posterior part of medial frontal cortex (pmFC) activates when individuals detect inconsistency between their attitude and behavior (van Veen et al., 2009; Izuma et al., 2010). However, these neuroimaging results still left it unknown whether the pmFC activity merely reflects an epiphenomenon of cognitive dissonance such as passive emotional reactions to inconsistency or if it plays a more active role in inducing preference change to resolve inconsistency.

In the present study, we attempted to answer this question of causality by using 25 min, 1 Hz repetitive transcranial magnetic stimulation (rTMS) to temporarily downregulate neuronal activity within pmFC to determine whether this reduces the cognitive dissonance effect. To measure choice-induced preference change, we used a modified free-choice paradigm using a “choice-blindness” procedure (Johansson et al., 2008; Hall et al., 2012; Hall et al., 2013; see Materials and Methods for details). This procedure controls for an important artifact inherent in the original free-choice paradigm (Chen and Risen, 2010; for review, see Izuma and Murayama, 2013), thus allowing for a methodologically rigorous test of the hypothesis.

Materials and Methods

Participants

A total of 61 individuals participated in the study. One participant decided to withdraw from the study after the TMS procedure was ex-

Received Nov. 5, 2014; revised Jan. 6, 2015; accepted Jan. 16, 2015.

Author contributions: K.I., S.A., K.M., M.I., and R.A. designed research; K.I., S.A., and D.-A.W. performed research; K.I. and S.A. analyzed data; K.I., S.A., D.-A.W., and R.A. wrote the paper.

This work was supported by the Japan Society for the Promotion of Science Fellows (to K.I.) and a Caltech Summer Undergraduate Research Fellowship (to S.A.). We thank Kevin Yang for help with pilot data collection and Ryuta Aoki for helpful comments on the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Keise Izuma, Department of Psychology, University of York, Heslington, York, YO10 5DD, UK. E-mail: keise.izuma@york.ac.uk.

DOI:10.1523/JNEUROSCI.4591-14.2015

Copyright © 2015 the authors 0270-6474/15/353598-09\$15.00/0

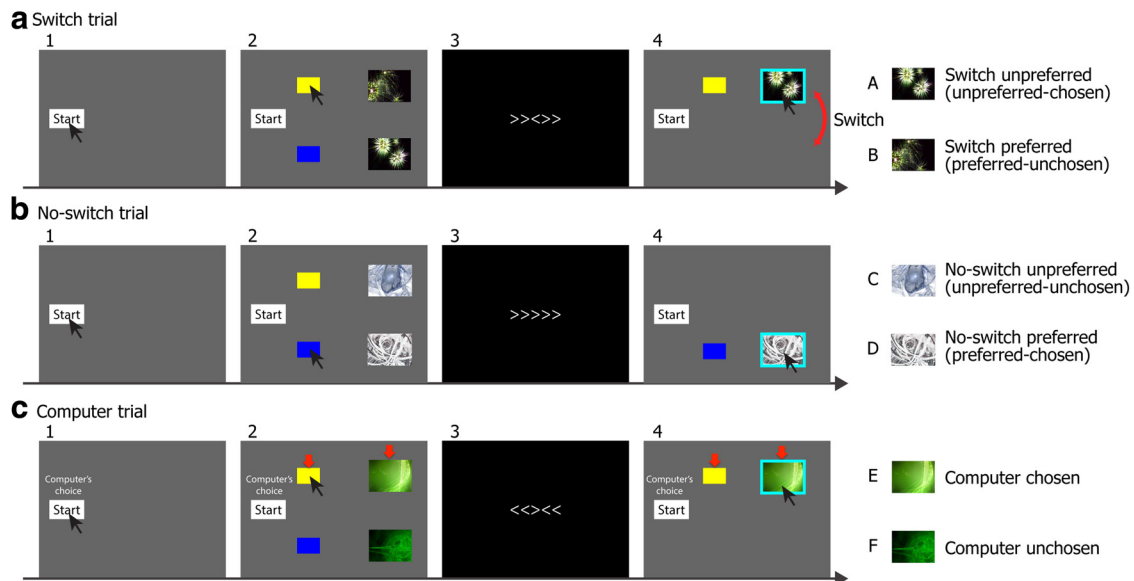


Figure 1. The illustration of a switch trial (**a**), a no-switch trial (**b**), and a computer trial (**c**) during the choice task. During switch and no-switch trials, participants were first instructed to move the mouse cursor to the box corresponding to the image that they preferred (e.g., a yellow box if they preferred the image on the top). After performing 4 trials of the Flanker task that serves as a distractor (stage 3), they were asked to click the image on the screen (stage 4; presumably the one they preferred at stage 2). Note that the locations of two images were switched between stages 2 and 4 only during a switch trial so that the participants were led to believe that they had chosen the image that they actually preferred less than the alternative. Each image of switch and no-switch trials was categorized into 1 of the 4 conditions (images A–D) depending on the condition (switch or no-switch) and participant’s action during stage 2 (preferred or unpreferred). During a computer trial, participants were asked to simply pick the image selected by a computer (indicated by red arrows). Each image of computer trials was categorized into either computer-chosen or computer-unchosen conditions (image E or F).

plained. Another participant was also withdrawn from the study because the participant reported feeling dizzy during the motor thresholding procedure (see below for more details). Seven additional participants were also excluded from the analyses: four detected our critical switch manipulation during the choice task (see below for details), two were excluded due to technical problems with TMS, and one showed a significantly negative correlation between two preference ratings (first and second rating tasks described below; $r_{(48)} = -0.23$, which is >3 SDs below the group average of $r = 0.48$ [SD = 0.21]), suggesting very poor compliance with the task instructions. Accordingly, the final analyses were based on 52 participants. Participants were randomly assigned to one of three TMS groups (between-subject design): (1) the pmFC group ($n = 17$, 8 female, mean age = 22.4), (2) the sham-pmFC group ($n = 17$, 7 female, mean age = 22.5), or (3) the posterior parietal cortex (PPC) group ($n = 18$, 8 female, mean age = 21.6). There was no significant difference in age across the three TMS groups ($p = 0.84$). The participants were all right-handed with no history of neurological or psychiatric illness. All participants gave written informed consent for participation and the study was approved by the Institutional Review Board of the California Institute of Technology.

Stimuli

The experimental stimuli chosen for this experiment were abstract computer desktop images freely available on the Internet (Fig. 1). There were 50 images used in the study, paired up in 25 pairs of similar but distinct images (the pairs were fixed among all participants). This was done to allow the switch manipulation used in the choice task to be undetected by participants, but still provide stimuli that could be preferred differentially.

Experimental tasks

In the present study, we used a modified free-choice paradigm using a “choice-blindness” procedure (Johansson et al., 2008; Hall et al., 2012; Hall et al., 2013) during the choice task. The procedure (detailed below) allowed us to address an important methodological flaw in the original paradigm (Chen and Risen, 2010; Izuma and Murayama, 2013). Ratings are not perfect measures of true preference (i.e., there is some “noise” in ratings). During the choice task of the paradigm, an individual’s choice

reveals additional information about their true preference. Because items chosen are more likely to have higher true preferences than rejected items, when measured a second time, chosen items are more likely to be rated higher than rejected items (i.e., regression toward true preference). Because the effects of true preference and choice are confounded in such a manner, the original “free-choice” paradigm could inadvertently measure preference change without the effect of choice on true preference. To address this methodological issue, we used a modified choice task in our experimental paradigm in which individuals were occasionally led to believe that they had chosen the item that they actually preferred less than the alternative. Therefore, although preference and choice often go together, our paradigm allows us to dissociate one from the other, which made it possible to dissociate the effects of choice from the effect of true preference. Before the current TMS study, we ran a behavioral pilot study (without TMS; $n = 26$) to verify that we could measure choice-induced preference change over and above the confound (data not shown). We also verified that the detection rate of our critical switch manipulation (see below) was sufficiently low (3 of 26 participants).

In the present study, participants performed the four following tasks in fixed order: (1) the first rating task, (2) the choice task, (3) the second rating task, and (4) the memory task. Because our prior fMRI study demonstrated pmFC activation in response to the discrepancy between behavior and attitude during the second rating task (Izuma et al., 2010), the rTMS treatment was administered right before participants performed the second rating tasks (an offline rTMS approach). After completing the second rating task, all participants took a 35–40 minute break before the memory task to ensure that the effects of TMS had subsided before continuing.

In the first rating task, participants were presented with images serially and were instructed to rate how much they liked each image using an 8-point scale where 1 = not at all and 8 = very much labeled on the keyboard. There was no time limit, but participants were encouraged to give their first impression. The intertrial interval was 1 s.

In the choice task, images were presented in pairs and participants were asked to indicate the image they preferred in each trial using a computer mouse with their right hand. Therefore, participants made 25 binary choices in total (50 images). There were four experimental condi-

tions during the choice task: (1) switch trials (five trials), (2) no-switch trials (five trials), (3) computer trials (five trials), and (4) other trials (10 trials; not included in preference change analysis).

Whereas the 25 pairs were fixed among all participants, which pairs were used in each of the four conditions (switch, no-switch, computer, and other) in the choice task was determined based on each participant's rating during the first rating task. Before the choice task, 15 image pairs (30 images) used in the switch, no-switch, and computer trials were automatically matched by a MATLAB program in terms of the difference in the first ratings between two images in each pair. As discussed previously (Chen and Risen, 2010; Izuma and Murayama, 2013), the strength of the artifact in the traditional free-choice paradigm depends on the rating difference between two images in a choice pair (i.e., the smaller the within-pair rating difference, the bigger the preference change by the artifact). Moreover, pairs with larger difference in the first ratings were preferentially allocated to the three conditions used in analysis. This was done to induce stronger cognitive dissonance (i.e., larger discrepancy between preference and choice) in the switch condition. That is, assuming that participants choose the image from a pair with the higher first rating most of the time, when their choice is reversed by the switch manipulation, they would believe that they chose the image they disliked and did not choose the image they liked. The remaining 20 images were assigned to the "other" condition and excluded from preference change analysis because items in the other trials were not matched on within-pair rating difference. Therefore, the comparisons between the other and the three critical conditions would not be informative. Although not used in the analyses, these additional 20 images were necessary to have more options, from which we chose 15 pairs for the three critical conditions that were sufficiently matched on the within-pair difference.

In each trial of the switch condition, participants were first asked to click the "start" button located on the left side of the screen (Fig. 1*a*, stage 1). As soon as they clicked the button, two images and two boxes (a yellow box on the top and a blue box on the bottom, both of which were aligned with each of two images) appeared on the screen (Fig. 1*a*, stage 2). Participants were asked to look at each of the two images carefully and to move the mouse cursor to the box corresponding to the image they preferred. For example, if they preferred the image on the top, they moved the cursor to the yellow box. The two images remained on the screen until a preference was indicated or for a maximum of 3 s. If the participants did not decide after 3 s, the 2 images disappeared, and participants then had to make their choice from memory.

As soon as the cursor touched one of the two boxes, the screen automatically changed, and participants performed four trials of a flanker task (Fig. 1*a*, stage 3), which was intended to distract participants' memory for the locations of the two images. During instructions, the choice task was referred to as the "choice and attention" task and each participant was told that the task was intended to test decision making and attention at the same time, providing an explanation for the flanker task. During the four trials of the flanker task, there were always two congruent (">>>>>" or "<<<<<") and two incongruent ("<<<<<" or ">><<>>") trials, and participants were asked to answer the direction of a central target arrow (left or right). Participants responded by pressing 1 of 2 keys labeled on the keyboard with their left hand (index and middle fingers) and were told that they would get 1 point every time they responded accurately within 1 s. They were encouraged to try to get at least 60 points (of 100) in total and aim for 80 points if possible. There was no monetary incentive for the task. Stimuli were presented once every second (0.5 s for stimulus presentation plus 0.5 s for interstimulus interval). After 4 trials (4 s), the total points accumulated were displayed on the screen for 1 s.

After the flanker task (5 s in total), the screen showed only one image, which was presumably the one they had preferred before the flanker task, along with the start button and the box touched by the participants (Fig. 1*a*, stage 4). Finally, participants were instructed to click the image on the screen (presumably the one they had preferred) and the image was highlighted by a light blue square for 1.5 s. Unknown to the participants, during the switch trials, the locations of two images were switched after the flanker task so that they were led to believe that they had chosen the image they actually preferred less than the alternative (i.e., the image that

was not preferred at the stage 2 of the choice task). The switch manipulation always occurred on 5th, 10th, 16th, 20th, and 24th trials. No participant reported the switches during the task, but four of them reported it during the postexperimental interview (see below).

In the no-switch trials (and other trials), the task proceeded in the same way except that no switch was made (Fig. 1*b*, stage 4). In the computer trials, the words "Computer's choice" were displayed above the start button at the beginning of the trial (Fig. 1*c*, stage 1). After clicking the start button, two images and two boxes were displayed on the screen, just like in the other trials. However, the participants were asked to wait for a computer to choose one of two images. After 1 s, the computer's choice (red arrows) was displayed on the screen (Fig. 1*c*, stage 2) and the participants were asked to simply move the cursor to the box (thus the wallpaper) selected by a computer. The rest of the task proceeded in the same way as the no-switch and other trials and no switch manipulation was introduced during the computer trials.

Importantly, unknown to the participants, each computer trial occurred shortly after each of the switch trials (specifically, the computer trials always occurred on 6th, 12th, 17th, 22nd, and 25th trials) and, by doing so, the computer's choice automatically mimicked the participant's choice made in a previous switch trial in terms of consistency between their first preference rating and choice (note that more specifically, "choice" here refers to participant's preference revealed at the stage 2 of the choice task). That is, suppose a participant was presented with two images that were rated 5 and 7 on the 8-point scale during the first preference rating task. Typically, they preferred the image with a higher first rating (i.e., the one rated 7) during stage 2 of the choice task. However, because behavior is noisy, they would occasionally prefer the image with a lower rating (i.e., the one rated 5) during the stage 2 of the choice task. Because this behavior is known to contribute to an important artifact (Chen and Risen, 2010; Izuma and Murayama, 2013), the computer trials were programmed to select the image with a lower first rating whenever the participant preferred (i.e., based on their action in the stage 2 of Fig. 1) the image with a lower first rating in a previous switch trial. Similarly, the computer selected the image with a higher first rating when the participant preferred the image with a higher first rating in a previous switch trial. The order of no-switch and other trials was determined randomly.

At the end of the choice task, each of the 30 images used in the switch, no-switch, and computer trials was categorized into the six conditions (five images each) depicted in Figure 1 (images A–F). Based on participant's action during the stage 2 of the switch and no-switch trials, each image was categorized as either "preferred" or "unpreferred." Similarly, based on participant's final choice during the stage 4, each image was categorized as either "chosen" or "unchosen." Because of the switch manipulation, preference and choice are inconsistent in the switch condition (e.g., preferred-unchosen or unpreferred-chosen). Therefore, the choice-blindness paradigm makes it possible to dissociate the effect of choice from the effect of preference.

After the completion of the choice task, the participants were escorted to the TMS room, where rTMS was administered as described below. Immediately after rTMS, the participants were asked to perform the second rating task in the TMS room. All participants started the task within 2 min of the TMS completion.

The second rating task was similar to the first one, with the exception that the images were now accompanied with information from the choices the participants had made previously (as similarly done in the previous fMRI study; Izuma et al., 2010). Below each image, they were presented with the information about whether they (or a computer) had chosen or unchosen the image during the choice task (e.g., "You chose this picture," "Computer rejected this picture," etc.). For items in the switch condition, this information was based on what participants believed to have chosen at the stage 4 of the choice task (Fig. 1). More specifically, those images clicked on after the Flanker task in the stage 4 were accompanied with "You chose this picture," whereas alternative images were accompanied with "You rejected this picture." This information was intended to make the effect of cognitive dissonance stronger by making the discrepancy between attitude and past behavior salient. It should be noted that, because of the explicit information, preference

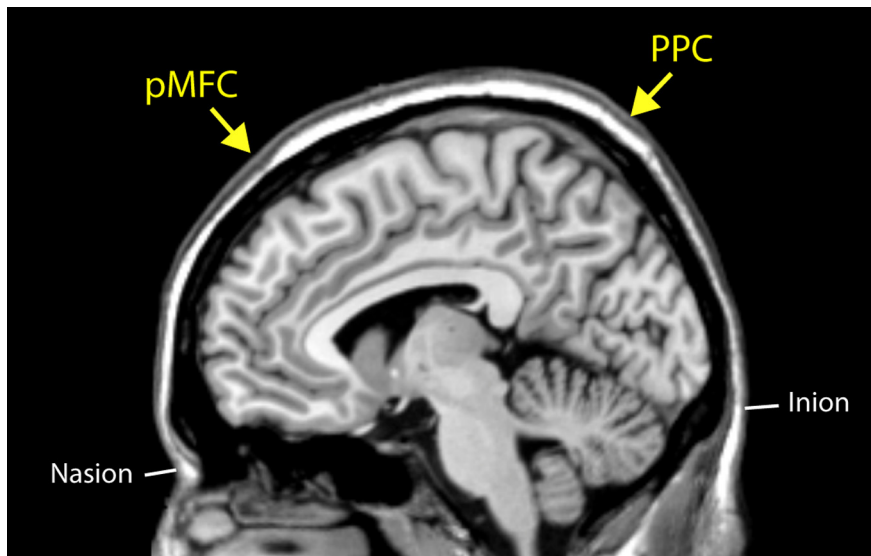


Figure 2. The location of rTMS target sites. The pMFC target region was determined as one-third of the distance from nasion to inion and the PPC region was determined as two-thirds of the same distance. Our two simulated regions approximately correspond to the MNI coordinates of $x = 10, y = 48, z = 70$ (pMFC) and $x = 10, y = -76, z = 85$ (PPC).

change found in the present study should, if anything, be inflated. However, the aim of the present study is not to find a minimum condition for choice-induced preference change to occur, but to test the effect of TMS on pMFC when it does occur (see Salti et al., 2014, which formally tests the effect of this explicit information). The participants were told that this information about their past choices was related to a task they were going to perform afterward, but not related to the rating task, and they were asked to focus on reporting their preference for each image one more time. As was the case in the first rating task, they rated the same 50 images using the 8-point scale.

After the second rating task, the participants were also asked to answer the Positive and Negative Affect Schedule (PANAS) questionnaire (10 items each for positive and negative moods; Watson et al., 1988) using an 8-point scale to measure whether different rTMS conditions had differential effect on their mood. All participants finished the second rating task and the PANAS questionnaire within 8 min after the TMS, presumably well under the influence of the 25 min rTMS (Brückner et al., 2013). After completing the questionnaire, they took a 35–40 min break in the original experimental room to ensure that the effects of the TMS had subsided before continuing.

After the break, the participants completed the memory task. Importantly, the memory task was completely unanticipated by the participants and they did not know that memory about their past choices would be tested later on. During the task, they were presented images serially (as in the first and second rating tasks) and were instructed to answer whether they “chose” or “rejected” (i.e., unchosen) the item during the choice task by pressing one of two keys on the keyboard. They were instructed that images presented in the computer trials of the choice task would not be presented. Therefore, a total of 40 images were presented during the task. Participants were also asked to make best guess if they were not certain.

The memory task was intended to test whether the rTMS had any differential effect on the participants’ ability to retrieve memory about their past choices. Although, during the second rating task, the participants’ past choice for each presented image was displayed explicitly, if the participants were able to remember their (correct) indicated preferences in the switch trials, then the intended effect of cognitive dissonance induced by discrepancy between behavior and attitude would have been likely been reduced. It should be noted that, during the memory task, the correct response was defined based on participants’ action at the stage 2 of the choice task. Accordingly, if the switch manipulation was successful, and participants believed the explicit information about their past choice during the second rating task, memory performance for the switch con-

dition should be lower than that for the no-switch condition (i.e., they mistakenly remembered that they chose the item that they actually did not prefer at the stage 2 due to the switch manipulation).

After finishing the memory task, participants were asked to complete the Preference for Consistency Scale (Cialdini et al., 1995) using an 8-point scale. This 18-item scale was intended to measure individuals’ disposition to perceive themselves as predictable, stable, and thus consistent (e.g., “it is important to me that my actions are consistent with my beliefs”).

Finally, as similarly done in the original choice-blindness study (Johansson et al., 2008), the participants were asked the three following questions during the postexperimental interview to ensure that they did not detect our critical switch manipulation: (1) “do you have any comments on the experiment?”, (2) “do you find anything strange with the wallpaper stimuli during the tasks?”, and (3) “some participants mentioned that, during some trials of the choice task, the locations of two images changed, do you think it happened to you?” The three questions were asked always in this order of increasing specificity. Four participants reported in response to the last question that the switch happened to them also, so they were excluded from the analyses. The remaining participants showed no sign of detection.

TMS procedures

To downregulate the activity in pMFC, we used 1 Hz rTMS, a noninvasive technique that can produce after-effects on cortical excitability, and a low frequency (e.g., 1 Hz) rTMS known to decrease cortical excitability in stimulated areas (Ridding and Rothwell, 2007). Each participant received 1 Hz rTMS for 25 min. According to previous reports (Brückner et al., 2013), the after effect of the 25 min 1 Hz rTMS is considered to last 20 min or more. A Magstim Super Rapid with a 70 mm air-cooled figure-8 coil was used.

The experimenter administering TMS was blind to the task and study design except for knowledge of the three stimulation conditions and was in the same room as the participants only during the TMS phase and an initial portion of the second rating task. The experimenter administering the participants’ tasks was absent during TMS positioning and delivery and was not informed of the TMS condition until after the completion of the experiment. Participants were blind to the study’s intent, but were aware of the possibility that they could be receiving either real or sham stimulation and that real stimulation would be expected to lower the level of brain activity in the area under the coil for up to 30 minutes after the end of stimulation.

The intensity of the TMS was set for each participant at 80% of the participant’s active motor threshold. To determine threshold, we first adjusted the TMS coil location and orientation to maximize the EMG magnitude in tibialis anterior from single pulse stimulation. The participant was then asked to steadily contract the muscle with moderate intensity. Starting from 50% output, the stimulator was adjusted in 5% steps to find the minimum intensity at which single-pulse TMS elicited a liminal EMG for 3 of 4 stimuli. The main rTMS treatment was conducted at 80% of this value.

Participants were randomly assigned to one of three TMS groups: (1) pMFC, (2) sham-pMFC, and (3) PPC. Based on previous fMRI studies (Izuma et al., 2010; Izuma and Adolphs, 2013), the pMFC target region was determined as one-third of the distance from nasion to inion (i.e., the region approximately corresponding to the areas commonly activated in our two previous fMRI studies) and the control (PPC) region was determined as two-thirds of the same distance (Harmer et al., 2001; Fig. 2). Because activations found in the previous fMRI studies were slightly stronger in the right hemisphere (Izuma et al., 2010; Izuma and Adolphs,

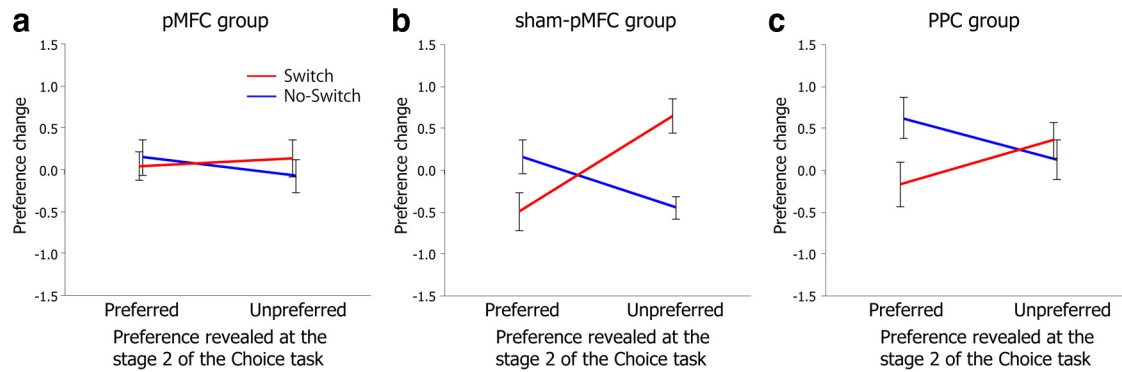


Figure 3. Mean preference changes (second preference ratings minus first preference ratings) in the switch and no-switch conditions across the three TMS groups; (a) pMFC group, (b) sham-pMFC group, and (c) PPC group. Error bars indicate the SEM. Note that “preferred” and “unpreferred” (*x*-axis) are based on participant’s relative preferences of paired images, which was revealed during stage 2 of the choice task (Fig. 1).

2013), both stimulation sites were offset to the right of the nasion-Cz-inion midline by 1 cm. The TMS coil was positioned with handle pointing in the superior direction. During the 25 min rTMS, participants in the pMFC and sham-pMFC group were supine, whereas those in the PPC group were prone. Sham-TMS was administered by tilting the coil 90 degrees off of the scalp so that no stimulation actually reached the cortex.

Coil orientation. TMS effects are maximized when the induced field is perpendicular to the gyral pattern because this aligns the field with cortical columns in the walls of the sulcus (Brasil-Neto et al., 1992; Mills et al., 1992; Fox et al., 2004). Based on the gyral pattern in the area of interest, we chose to orient the field parallel to the midline. Stimulation polarity with respect to this line can modulate the effectiveness of stimulation, but there is no preceding research that would indicate the optimal polarity for this area or function. In addition, the stimulator is biphasic, which makes polarity less critical (Brasil-Neto et al., 1992). Therefore, we positioned the handle in the superior direction to minimize the equipment’s obtrusiveness to the participant. It is possible that the opposite orientation of the coil would have induced a stronger effect.

Adverse effects. TMS was well tolerated by most participants, who generally fell asleep during the main treatment. One participant, however, experienced dizziness after receiving three sham and five real single-pulse stimuli during the search for the site for stimulating tibialis anterior. Dizziness declined, resolving after ~2 min. The participant was examined by emergency medical technicians and released. Dizziness was attributed to a combination of stress/anxiety of the experiment and preexisting conditions of sleep deprivation and dehydration.

Data analyses

Our main analysis is based on the 2 (preference: preferred or unpreferred) × 2 (switch: switch or no switch) × 3 (TMS group: pMFC, sham-pMFC, or PPC) mixed design with preference change (second preference ratings minus first preference ratings) as a dependent variable. It should be noted that, in our choice-blindness paradigm, a choice-induced preference change is indicated by a significant 2 (preference: preferred or unpreferred) × 2 (switch: switch or no switch) interaction. This 2 × 2 interaction indicates that preference change depends on what the participants believed about past choices that they had made previously. Regardless of their initial relative preferences of two images as revealed in the stage 2 of the choice task (preferred or unpreferred; Fig. 1), preferences for the images should increase when participants believe that they had chosen the image (i.e., the preferred-no-switch condition and the unpreferred-switch condition), whereas preferences for images should decrease when images were thought to have been unchosen (i.e., the preferred-switch condition and the unpreferred-no-switch condition).

Furthermore, to compute the size of the TMS effect on choice-induced preference change and to investigate the correlation between individual difference in choice-induced preference change and participant’s tendency to prefer consistency, as measured by the preference for consistency scale, we computed a single preference change score for each participant as following: We computed the 2-way (preference × switch)

interaction term, divided it by 2 to take an average of preference change observed in the preferred and unpreferred conditions, and subtracted preference change observed in the computer condition. This score represents the magnitude of preference change caused by choices participants themselves made. More specifically, a preference change (PC) score was computed as follows:

$$\left[\frac{(PC_{\text{Preferred-No_switch}} - PC_{\text{Preferred-Switch}}) + (PC_{\text{Unpreferred-Switch}} - PC_{\text{Unpreferred-No_Switch}})}{2} - (PC_{\text{Computer_Chosen}} - PC_{\text{Computer_unchosen}}) \right]$$

Results

Preference change

A 2 (preference: preferred or unpreferred) × 2 (switch: switch or no switch) × 3 (TMS group: pMFC, sham-pMFC, or PPC) mixed ANOVA revealed a significant 2-way preference × switch interaction ($F_{(1,49)} = 23.19, p < 0.001$). Importantly, it also revealed a critical 3-way preference × switch × TMS group interaction ($F_{(2,49)} = 3.34, p = 0.044$; Fig. 3). No other significant main effect or interaction was found in the ANOVA (all $p > 0.082$).

To further probe the significant three-way interaction, three separate two-way (preference × choice) repeated-measures ANOVAs were conducted for each TMS group. This revealed a significant preference × choice interaction in the sham-pMFC group ($F_{(1,16)} = 16.36, p < 0.001$; Fig. 3b) and in the PPC group ($F_{(1,17)} = 7.73, p = 0.013$; Fig. 3c). As stated earlier, this two-way interaction indicates the existence of choice-induced preference change and indicates that preference change depends on what the participants believed about past choices they had made previously regardless of their initial relative preferences of two images (preferred or unpreferred). We further conducted two-sample *t* tests comparing the switch and the no-switch conditions in each of the preferred and unpreferred conditions for the sham-pMFC and PPC groups. The results revealed that the difference between the switch and no-switch conditions were all significant ($p < 0.006$) except for the unpreferred condition of the PPC group ($p = 0.21$). In contrast, the same 2-way interaction was not significant for the pMFC group ($F_{(1,16)} = 1.30, p = 0.27, n.s.$; Fig. 3a). Two-sample *t* tests comparing the switch and no-switch conditions in each of the preferred and unpreferred conditions for the pMFC group revealed that there was no significant difference in both preferred and unpreferred conditions ($p > 0.45$). Therefore, rTMS to the pMFC region significantly reduced choice-induced preference change.

To determine whether TMS had any effect on preference change for the items presented in the computer condition, we conducted 2 (computer’s choice: chosen or unchosen) × 3 (TMS

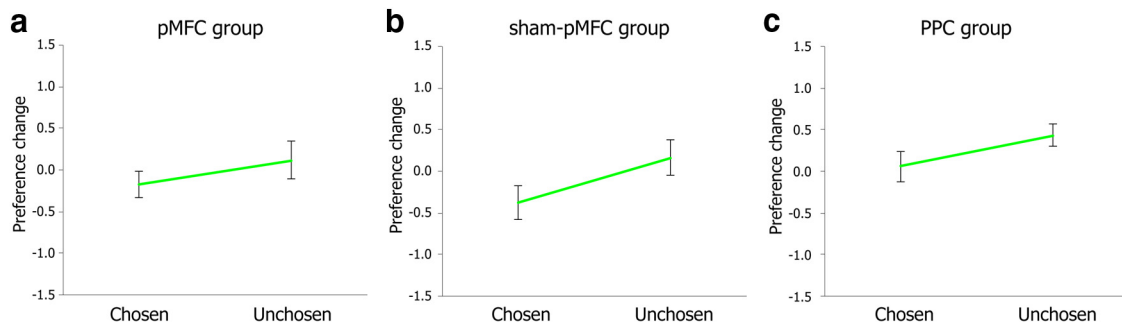


Figure 4. Mean preference changes (second preference ratings minus first preference ratings) in the computer condition across the three TMS groups; (a) pMFC group, (b) sham-pMFC group, and (c) PPC group. Error bars indicate SEM.

group) mixed ANOVA. It revealed a significant main effect of computer's choice ($F_{(1,49)} = 7.65, p = 0.008$). However, a main effect of TMS group and a 2-way interaction were not significant ($p > 0.18$), indicating that TMS had no influence on preference change after choices made by the computer. As seen in Figure 4, across all three TMS groups, participants' preference for the items chosen by computer decreased, whereas their preference for the items unchosen by computer increased. This result seems to suggest the existence of a significant regression-to-the-mean effect. Because the computer's choice was matched with participant's choice made in the switch condition, the items chosen by the computer were more likely to have higher first preference ratings. Accordingly, participant's preference rating for the items chosen by the computer is more likely to decrease during the second rating task due to the regression-to-the-mean effect compared with the items unchosen by computer.

Effect size analysis

To compute the size of the TMS effect on choice-induced preference change, we computed a single preference change score as described in Materials and Methods. One-way ANOVA on this score comparing across the three TMS groups revealed a marginally significant effect of TMS ($F_{(2,49)} = 2.56, p = 0.088$). Because there was no significant difference in mean preference change scores between two control conditions (the sham-pMFC and PPC groups, $p = 0.28$), we combined these two control groups to calculate Cohen's d . A two-sample t test comparing the pMFC group with this combined control group revealed a significant difference ($t_{(50)} = 1.93, p = 0.03$) and the calculation of Cohen's d revealed a moderate effect size of the TMS ($d = 0.59$).

Control analyses

We investigated whether the rTMS had any effect on participants' general attention level as indexed by reaction times (RTs) during the first and second rating tasks, mood as measured by PANAS, and memory about past choices.

A 2 (rating task: first or second) \times 3 (TMS group) mixed ANOVA on RTs revealed no significant main effects or interaction ($p > 0.15$), suggesting that the rTMS did not affect participants' attention level during the second rating task, which they performed immediately after the rTMS. Regardless of the TMS groups, the average RTs (including all 50 trials) during the first rating task was 2.17 s (SD = 0.80), whereas the average RT during the second rating task was 2.07 s (SD = 0.70).

Two 1 \times 3 (TMS group) ANOVAs on positive and negative mood of the PANAS score revealed no significant effect of the rTMS ($p > 0.34$), indicating that the rTMS did not alter participants' mood.

A 2 (switch: switch or no-switch) \times 3 (TMS group) mixed ANOVA on memory about past choices revealed a significant main effect of the switch manipulation ($F_{(1,49)} = 4.14, p = 0.047$). Because, in the switch condition, the participants were led to believe that they chose the image they preferred less than the alternatives, their memory performances were significantly lower than those in the no-switch condition. Regardless of the TMS group, average memory performance of the no-switch condition was 58.5% (SD = 13.6), which was significantly higher than the chance level (50%; $t_{(51)} = 4.48, p < 0.001$), whereas average memory performance of the switch condition was 47.9% (SD = 16.4), which did not differ significantly from the chance level ($p = 0.36$). This result suggests that our switch manipulation was successful and that participants continued to believe that they had chosen the images they actually preferred less than alternatives and had not chosen the images that they preferred more than alternatives. A main effect of the TMS group and a 2-way interaction were not significant ($p > 0.77$), suggesting that our main results reported above cannot be explained by differential memory about past choices across the three TMS groups.

Finally, we also investigated whether the within-pair rating difference and the preference-choice consistency were successfully matched across the three experimental conditions (switch, no-switch, and computer) during the choice task, as we intended.

First, average within-pair rating differences (all TMS group combined, $n = 52$) in the no-switch and switch conditions were both 1.98, and it was 2.01 in the computer condition. A 3 (experimental conditions: switch, no-switch, or computer) \times 3 (TMS group: pMFC, sham-pMFC, or PPC) mixed ANOVA revealed no significant main effects or interaction ($p > 0.19$).

Second, participants generally chose images that had higher first preference rating when they had a chance to make a choice themselves (i.e., the switch and no-switch conditions). The mean preference choice consistency (all TMS group combined, $n = 52$) in the no-switch condition was 62.9%, and it was 65.0% in the switch condition and the computer condition (as previously described, the preference choice consistency was matched between these two conditions by the task program). A 3 (experimental conditions: switch, no-switch, or computer) \times 3 (TMS group: pMFC, sham-pMFC, or PPC) mixed ANOVA revealed no significant main effect or interaction ($ps > 0.66$).

Together, these results indicate that the main findings reported above is highly unlikely to be explained by the different level of the artifact (Chen and Risen, 2010) across three TMS groups.

Preference for consistency scale

Finally, we conducted 1×3 (TMS group) between-participant ANOVA on the preference for consistency scale. It revealed no significant effect ($p = 0.54$), suggesting that individual difference in the tendency to prefer consistency was no different across the three TMS groups.

We also investigated, as exploratory analyses, whether participants' preference for consistency scores were related to their level of choice-induced preference change separately for each TMS group. We used the same single preference change score described above to index an individual's tendency to justify choices they made. Although correlations in all 3 groups did not reach statistical significance largely due to our limited sample size, the sham-pMFC and PPC groups tended to show positive correlation (sham-pMFC $r_{(15)} = 0.27, p = 0.15$; PPC $r_{(16)} = 0.32, p = 0.10$). In contrast, correlation in the pMFC group was virtually zero ($r_{(15)} = 0.03, p = 0.46$).

Discussion

The present study demonstrated that choice-induced preference change was significantly reduced by TMS over the pMFC region compared with the control conditions (TMS to the PPC region and sham-TMS to the pMFC region), providing the evidence that the pMFC plays a causal role in inducing preference change after inconsistency between choice and preference. It is important to note that we used a modified choice-blindness paradigm to control for the artifact inherent in the traditional free-choice paradigm (Chen and Risen, 2010). Therefore, unlike a number of previous behavioral and neuroimaging studies that used the original paradigm (Izuma and Murayama, 2013), the present study provides unequivocal evidence for the existence of choice-induced preference change and the effect of TMS over different brain regions on the phenomenon.

The present findings are consistent with previous studies showing that the pMFC plays a causal role in inducing behavioral or attitude change after a variety of aversive experiences in both social and nonsocial contexts. The pMFC is known to be activated by a variety of aversive outcomes (Shackman et al., 2011). Furthermore, some neurons in pMFC do not simply respond to negative outcomes per se, rather, these neurons respond to negative outcome (e.g., reduced reward) only when the monkey subsequently changed their behavior from the previous trial (Shima and Tanji, 1998). There exist similar neurons in the human dorsal anterior cingulate regions and ablation of this area significantly impaired participants' performance in a reward-based movement selection task, especially when they had to change their movement after reduced reward (Williams et al., 2004). Recently, fMRI studies found that the same region is activated by aversive experiences in social contexts (for review, see Izuma, 2013), such as disagreeing with others (Klucharev et al., 2009; Izuma and Adolphs, 2013) or agreeing with people we dislike (Izuma and Adolphs, 2013). Furthermore, Klucharev et al. (2011) demonstrated that TMS to the pMFC region significantly decreased individual's tendency to conform to others' opinions. Therefore, just like its known role for nonsocial situations (Williams et al., 2004), the pMFC plays a key role not only in detecting a social situation that requires some level of adjustment in behavior or attitude (such as cognitive dissonance, disagreeing with others, etc.), but also in actually inducing that change.

It is important to note that preference change after cognitive dissonance requires at least two processes: detection of discrepancy (conflict) and subsequent implementation of conflict resolution. The present study does not allow us to specify the exact

role played by pMFC. The pMFC could play a role in conflict detection, conflict resolution, or both. Two past TMS studies on response-level conflict showed that pre-SMA plays a causal role in conflict resolution rather than conflict detection (Mars et al., 2009; Soutschek et al., 2013) and it will be an important future direction to investigate whether the pMFC plays a similar causal role in resolving (rather than simply detecting) cognitive conflict such as cognitive dissonance. Furthermore, two past studies (Harmon-Jones et al., 2008b; Mengarelli et al., 2013) suggested that left dorsolateral prefrontal cortex (DLPFC) plays a causal role in the resolution of cognitive dissonance, part of a larger literature from neuroimaging studies that have found correlational evidence for this brain region's role in cognitive dissonance processes (Harmon-Jones et al., 2008a; Harmon-Jones et al., 2008b; van Veen et al., 2009; Izuma et al., 2010). It was further demonstrated that the DLPFC specifically in left hemisphere plays a causal role in choice-induced preference change (Harmon-Jones et al., 2008b; Mengarelli et al., 2013). An EEG study found that, when an individual's left DLPFC activity was decreased through neurofeedback training, choice-induced preference change was reduced significantly (Harmon-Jones et al., 2008b). Conceptually similar results have been reported by a study using transcranial direct current stimulation (tDCS; Mengarelli et al., 2013). That study showed that choice-induced preference change was significantly reduced after individual's left DLPFC was stimulated by cathodal tDCS, which is known to reduce the excitability of the target areas. However, another EEG study (Harmon-Jones et al., 2011) suggests that the DLPFC is not related to cognitive dissonance per se and the commitment to the chosen course of actions can increase the left DLPFC activity regardless of the existence of inconsistency between attitude and actions. Because the DLPFC is known to be involved in cognitive control (Miller and Cohen, 2001), the studies seem to suggest that the actual implementation of choice justification requires general cognitive control processes.

Although this neural model of a discrepancy (dissonance) resolution system is similar to the model of response level conflict resolution (Mansouri et al., 2009), it is probably too simplistic to argue that they share the common neural mechanisms. For example, our previous study (Izuma and Adolphs, 2013) showed that the pMFC region activated by cognitive imbalance, which is conceptually similar to cognitive dissonance (Abelson et al., 1968; Gawronski and Strack, 2012), does not overlap with the region activated by response conflict. The region activated by response conflict was located more posteriorly (i.e., pre-SMA; Izuma and Adolphs, 2013), consistent with two previous TMS studies (Mars et al., 2009; Soutschek et al., 2013) demonstrating that stimulating pre-SMA disrupts response conflict resolution. Although the lack of anatomical resolution in the present study does not allow us to specify the exact location stimulated by TMS, our pMFC target region approximately corresponds to $x = 10, y = 48, z = 70$ (Jurcak et al., 2005), which is more anterior than the previous studies: $y = 18$ (Mars et al., 2009) and $y = 21$ (Soutschek et al., 2013). As previously discussed (Izuma and Adolphs, 2013), we believe that the neural mechanism of a discrepancy resolution system may be similar to that of behavioral adjustment after an aversive outcome. Previously, we found that the areas activated by cognitive imbalance largely overlapped with the areas activated by aversive outcome (posterior dmPFC; Izuma and Adolphs, 2013). This functional distinction within pMFC is consistent with what has been suggested previously (Ridderinkhof et al., 2004; Rushworth et al., 2004; Hikosaka and Isoda, 2010). Furthermore, other additional regions, such as

the insula, seem to play a role in processing social types of conflict (Klucharev et al., 2009; van Veen et al., 2009; Izuma et al., 2010; Izuma and Adolphs, 2013). Comparisons between these multiple systems for resolving discrepancies between expectations and outcomes should be investigated further in future research.

It should be noted that the preference change observed in the present study could be explained by self-perception theory (Bem, 1967), which posits that, just as we come to know another person's preference by observing their behavior, we infer our own preference by observing our own behavior (e.g., I must like it because I chose it). This limitation is inherent in all past studies demonstrating choice-induced preference change. Greenwald (1975, 2012) even argued that these two theories cannot be distinguished. Although not conclusive, our present TMS study, along with the previous fMRI study (Izuma et al., 2010), seem to support cognitive dissonance theory. In our previous fMRI study, the degree of cognitive dissonance was quantified as the discrepancy between a participant's past choice and preference for the item and we found that activity in the pmMFC correlated with this parameter. Because this parameter is irrelevant to self-perception theory, being able to reduce choice-induced preference change by stimulating the pmMFC with TMS suggests that dissonance reduction processes played a major role in the present study. However, we admit that the fact that our data are consistent with one theory does not necessarily mean that the other theory plays no role (i.e., consistency fallacy). Nonetheless, the present study highlights the potential of cognitive neuroscience methods (fMRI, TMS) in distinguishing two social psychological theories. For example, we can test whether different brain regions are activated by psychological processes assumed in cognitive dissonance theory versus self-perception theory. If activated regions are reliably different, we can further test whether TMS to those regions could reduce choice-induced preference change.

In summary, the present study demonstrated that TMS to the pmMFC could reduce choice-induced preference change. Our results inform neural models of the choice justification processes implemented by the pmMFC and provide a clear demonstration of the causal role of the pmMFC. It will be most important in future studies to sharpen our understanding of this causal role by providing direct comparisons among different types of conflict and to link the role of the pmMFC to that of other brain regions with which it is connected.

References

- Abelson RP, Aronson E, McGuire WJ, Newcomb TM, Rosenberg MJ, Tannenbaum PH (1968) Theories of cognitive consistency: a sourcebook. Chicago: Rand McNally.
- Bem DJ (1967) Self-perception: an alternative interpretation of cognitive dissonance phenomena. *Psychol Rev* 74:183–200. [CrossRef Medline](#)
- Brasil-Neto JP, Cohen LG, Panizza M, Nilsson J, Roth BJ, Hallett M (1992) Optimal focal transcranial magnetic activation of the human motor cortex: effects of coil orientation, shape of the induced current pulse, and stimulus intensity. *J Clin Neurophysiol* 9:132–136. [CrossRef Medline](#)
- Brehm JW (1956) Post-decision changes in the desirability of choice alternatives. *J Abnorm Soc Psychol* 52:384–389. [CrossRef](#)
- Brückner S, Kiefer M, Kammer T (2013) Comparing the after-effects of continuous theta burst stimulation and conventional 1 Hz rTMS on semantic processing. *Neuroscience* 233:64–71. [CrossRef Medline](#)
- Chen MK, Risen JL (2010) How choice affects and reflects preferences: revisiting the free-choice paradigm. *J Pers Soc Psychol* 99:573–594. [CrossRef Medline](#)
- Cialdini RB, Trost MR, Newsom JT (1995) Preference for consistency—the development of a valid measure and the discovery of surprising behavioral-implications. *J Pers Soc Psychol* 69:318–328. [CrossRef](#)
- Festinger L (1957) A theory of cognitive dissonance. Stanford: Stanford University.
- Fox PT, Narayana S, Tandon N, Sandoval H, Fox SP, Kochunov P, Lancaster JL (2004) Column-based model of electric field excitation of cerebral cortex. *Hum Brain Mapp* 22:1–14. [CrossRef Medline](#)
- Gawronski B, Strack F, eds (2012) Cognitive consistency: a fundamental principle in social cognition, Ed 1. New York: Guilford.
- Greenwald AG (1975) Inconclusiveness of crucial cognitive tests of dissonance versus self-perception theories. *J Exp Soc Psychol* 11:490–499. [CrossRef](#)
- Greenwald AG (2012) There is nothing so theoretical as a good method. *Perspect Psychol Sci* 7:99–108. [CrossRef](#)
- Hall L, Johansson P, Strandberg T (2012) Lifting the veil of morality: choice blindness and attitude reversals on a self-transforming survey. *PLoS One* 7:e45457. [CrossRef Medline](#)
- Hall L, Strandberg T, Pärnamets P, Lind A, Tärning B, Johansson P (2013) How the polls can be both spot on and dead wrong: using choice blindness to shift political attitudes and voter intentions. *PLoS One* 8:e60554. [CrossRef Medline](#)
- Harmer CJ, Thilo KV, Rothwell JC, Goodwin GM (2001) Transcranial magnetic stimulation of medial-frontal cortex impairs the processing of angry facial expressions. *Nat Neurosci* 4:17–18. [CrossRef Medline](#)
- Harmon-Jones E, Gerdjikov T, Harmon-Jones C (2008a) 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. [CrossRef](#)
- Harmon-Jones E, Harmon-Jones C, Fearn M, Sigelman JD, Johnson P (2008b) Left frontal cortical activation and spreading of alternatives: tests of the action-based model of dissonance. *J Pers Soc Psychol* 94:1–15. [CrossRef Medline](#)
- Harmon-Jones E, Harmon-Jones C, Serra R, Gable PA (2011) The effect of commitment on relative left frontal cortical activity: tests of the action-based model of dissonance. *Pers Soc Psychol Bull* 37:395–408. [CrossRef Medline](#)
- Hikosaka O, Isoda M (2010) Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. *Trends Cogn Sci* 14:154–161. [CrossRef Medline](#)
- Izuma K (2013) The neural basis of social influence and attitude change. *Curr Opin Neurobiol* 23:456–462. [CrossRef Medline](#)
- Izuma K, Adolphs R (2013) Social manipulation of preference in the human brain. *Neuron* 78:563–573. [CrossRef Medline](#)
- Izuma K, Murayama K (2013) Choice-induced preference change in the free-choice paradigm: a critical methodological review. *Front Psychol* 4:41. [Medline](#)
- Izuma K, Matsumoto M, Murayama K, Samejima K, Sadato N, Matsumoto K (2010) Neural correlates of cognitive dissonance and choice-induced preference change. *Proc Natl Acad Sci U S A* 107:22014–22019. [CrossRef Medline](#)
- Johansson P, Hall L, Sikstrom S (2008) From change blindness to choice blindness. *Psychologia* 51:142–155. [CrossRef](#)
- Jurcak V, Okamoto M, Singh A, Dan I (2005) Virtual 10–20 measurement on MR images for inter-modal linking of transcranial and tomographic neuroimaging methods. *Neuroimage* 26:1184–1192. [CrossRef Medline](#)
- Klucharev V, Hytönen K, Rijpkema M, Smidts A, Fernández G (2009) Reinforcement learning signal predicts social conformity. *Neuron* 61:140–151. [CrossRef Medline](#)
- Klucharev V, Munneke MA, Smidts A, Fernández G (2011) Downregulation of the posterior medial frontal cortex prevents social conformity. *J Neurosci* 31:11934–11940. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Mars RB, Klein MC, Neubert FX, Olivier E, Buch ER, Boorman ED, Rushworth MF (2009) Short-latency influence of medial frontal cortex on primary motor cortex during action selection under conflict. *J Neurosci* 29:6926–6931. [CrossRef Medline](#)
- Mengarelli F, Spoglianti S, Avenanti A, di Pellegrino G (2013) Cathodal tDCS over the left prefrontal cortex diminishes choice-induced preference change. *Cereb Cortex*. Advance online publication. doi: 10.1093/cercor/bht314. [CrossRef Medline](#)
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202. [CrossRef Medline](#)
- Mills KR, Boniface SJ, Schubert M (1992) Magnetic brain stimulation with a

- double coil: the importance of coil orientation. *Electroenceph Clin Neurophysiol* 85:17–21. [CrossRef Medline](#)
- Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004) The role of the medial frontal cortex in cognitive control. *Science* 306:443–447. [CrossRef Medline](#)
- Ridding MC, Rothwell JC (2007) Is there a future for therapeutic use of transcranial magnetic stimulation? *Nat Rev Neurosci* 8:559–567. [CrossRef Medline](#)
- Rushworth MF, Walton ME, Kennerley SW, Bannerman DM (2004) Action sets and decisions in the medial frontal cortex. *Trends Cogn Sci* 8:410–417. [CrossRef Medline](#)
- Salti M, El Karoui I, Maillet M, Naccache L (2014) Cognitive dissonance resolution is related to episodic memory. *PLoS One* 9:e108579. [CrossRef Medline](#)
- Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ (2011) The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat Rev Neurosci* 12:154–167. [CrossRef Medline](#)
- Shima K, Tanji J (1998) Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* 282:1335–1338. [CrossRef Medline](#)
- Soutschek A, Taylor PC, Müller HJ, Schubert T (2013) Dissociable networks control conflict during perception and response selection: a transcranial magnetic stimulation study. *J Neurosci* 33:5647–5654. [CrossRef Medline](#)
- van Veen V, Krug MK, Schooler JW, Carter CS (2009) Neural activity predicts attitude change in cognitive dissonance. *Nat Neurosci* 12:1469–1474. [CrossRef Medline](#)
- Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect—the Panas scales. *J Pers Soc Psychol* 54:1063–1070. [CrossRef Medline](#)
- Williams ZM, Bush G, Rauch SL, Cosgrove GR, Eskandar EN (2004) Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat Neurosci* 7:1370–1375. [CrossRef Medline](#)