

Neural activity predicts attitude change in cognitive dissonance

Vincent van Veen^{1,2}, Marie K Krug², Jonathan W Schooler³ & Cameron S Carter²

When our actions conflict with our prior attitudes, we often change our attitudes to be more consistent with our actions. This phenomenon, known as cognitive dissonance, is considered to be one of the most influential theories in psychology. However, the neural basis of this phenomenon is unknown. Using a Solomon four-group design, we scanned participants with functional MRI while they argued that the uncomfortable scanner environment was nevertheless a pleasant experience. We found that cognitive dissonance engaged the dorsal anterior cingulate cortex and anterior insula; furthermore, we found that the activation of these regions tightly predicted participants' subsequent attitude change. These effects were not observed in a control group. Our findings elucidate the neural representation of cognitive dissonance, and support the role of the anterior cingulate cortex in detecting cognitive conflict and the neural prediction of attitude change.

According to cognitive dissonance theory, people tend to strive to keep their knowledge, actions and attitudes consistent (consonant). Inconsistent (dissonant) behavior and attitudes result in a psychologically uncomfortable state that motivates people to reduce the dissonance, often by changing their attitudes to be more consonant with the displayed behavior. Since this theory was first proposed in the 1950s^{1,2}, it has led to a large amount of fruitful research in social psychology³ and is considered to be one of the most influential theories in psychology⁴. However, little is known about how cognitive dissonance is represented in the brain or what the cognitive mechanisms might be that mediate this process.

We used functional magnetic resonance imaging (fMRI) to study how the brain responds to cognitive dissonance in a modified version of the classic 'induced compliance' procedure². In this procedure, participants argue in favor of a position that is counter to their actual attitudes (counter-attitudinal argument). It has consistently been found that participants change their attitudes to be more consistent with the counter-attitudinal behavior. Dissonance has been shown to be a negative emotional state^{5–7} accompanied by autonomic arousal^{5,8}; it has been shown that people change their attitudes and restore consonance to specifically reduce the negative affect^{5–7}. When participants in control groups are able to attribute their counter-attitudinal behavior to payment² or coercion^{3,9–11}, or when the counter-attitudinal behavior has no real-world consequences^{10,12}, conflict between behavior and prior attitudes is reduced, and participants experience less cognitive dissonance and do not change their attitudes (see **Supplementary Discussion**). Notably, as dissonance theory has largely focused on what motivates attitude change rather than how that change comes about, we focused on the neural correlates of the actual dissonance, rather than the attitude change that follows it, which awaits future study.

One candidate region for the detection and processing of cognitive dissonance is the dorsal anterior cingulate cortex (dACC). We and others have proposed that one of the dACC's functions in cognition is to detect conflicts between active, but incompatible, streams of information processing^{13–15}, such as between the color and the meaning of a word in the Stroop task^{16,17}. dACC activation is consistently related to the amount of conflict occurring in such tasks. Computational simulations of conflict in simple speeded response tasks have measured conflict as Hopfield's energy and have shown that dACC activation in such tasks can be well modeled by this measure^{13,18}. Likewise, conflict is an important component of the classic dissonance theory¹, and computational models of cognitive dissonance have measured it as increased energy^{19,20}.

We hypothesized that the dACC's conflict monitoring functions might generalize from detecting conflict in simple speeded-response tasks to detecting conflict between prior attitudes and counter-attitudinal behavior in cognitive dissonance^{11,14}. To test this, we adapted the induced compliance procedure² into an event-related fMRI design. Participants first performed a rather long (45 min) and boring task in the uncomfortable environment of the magnetic resonance scanner. Participants then participated in a second task, during which they had to respond to sentences presented on a screen with their left or right ring, middle, or index finger, as on a 6-point Likert scale (1 = left ring finger, completely agree; 6 = right ring finger, completely disagree). We used two types of sentences: target sentences consisting of attitudes toward the scanner and task, and neutral sentences (**Fig. 1**).

While participants were performing the initial task, they were randomly assigned to one of two groups, dissonance or control. Participants in the control group were told to respond to the target sentences as though they were enjoying the scanner and the task, regardless of whether they

¹Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, California, USA. ²Imaging Research Center, University of California at Davis, Sacramento, California, USA. ³Department of Psychology, University of California at Santa Barbara, Santa Barbara, California, USA. Correspondence should be addressed to V.v.V. (vanveen@berkeley.edu) or C.S.C. (cameron.carter@ucdmc.ucdavis.edu).

Received 29 July; accepted 8 September; published online 16 September 2009; doi:10.1038/nn.2413

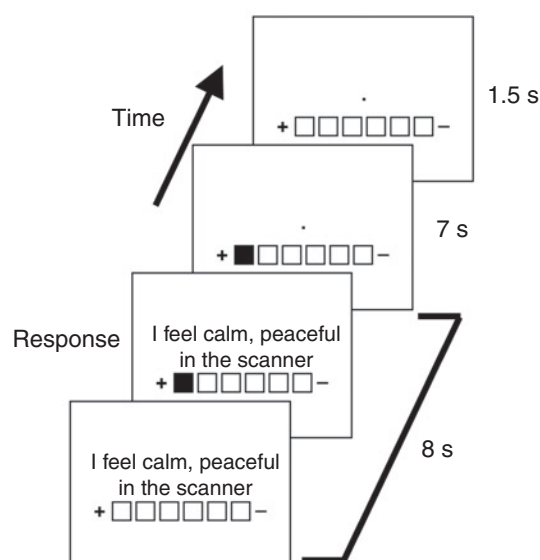


Figure 1 Trial sequence during the task. Throughout the task, the outline of six squares was visible below the fixation point, along with a + and a – sign, which represented agreement and disagreement, respectively. The participant responded by pressing with the left or right ring, middle, or index finger (completely agree, mostly agree, somewhat agree, somewhat disagree, mostly disagree, completely disagree). When a button was pressed, the corresponding square lit up, until 15 s into the trial, followed by a 1.5-s fixation point. Note that the actual stimulus was white on a black background, rather than black on a white background.

were actually enjoying the experience. Furthermore, they were informed that they would receive an additional dollar for each sentence that they responded to in this way. They were instructed to respond honestly to the other (neutral) sentences.

Participants in the dissonance group were also instructed on how to respond to the stimuli. They were then told that a patient had been scheduled to be scanned after them and was to perform a similar task in the scanner. This patient, the participants were told, was now in the scanner control room, watching the screen of the experimental control computer, and was very nervous and uncomfortable about the upcoming scanning session. The participants were then told that several of the sentences were about their attitudes toward the scanner and the task and were asked if they would be willing to respond as though they were enjoying being in the scanner and performing the task, regardless of how they actually felt about the experience. This, they were told, might put the patient's mind at ease, as the patient in the control room could see the responses on screen. We reasoned that this would be analogous to making a counter-attitudinal argument.

After scanning, participants were led into a private waiting room where they filled out a set of forms, again with the target sentences; this time they were asked to respond with how they actually felt about their experience in the scanner. A composite score was then calculated for the participants' overall enjoyment of the scanner and task. After completing the forms, the participants were carefully debriefed; participants who admitted having doubts about the validity of the cover story were not included in the analyses (see **Supplementary Methods and Supplementary Results**).

In previous studies of the induced compliance procedure, attitude change has been measured as a difference in final attitude score between a dissonance and a control group (that is, a post-test–only control group design). The astute reader will realize that attitude change takes place within participants, not between participants. Arguably, a within-subject measure of attitude change, before and after the counter-attitudinal argument, would constitute a more critical measure of attitude change (that is, use a pre- and post-test control group design). However, within-participant measures of attitude change are rarely assessed, as attitude measurements have often been shown to influence subsequent measurements^{21,22}. When participants give their attitude during a pretest, this often influences the subsequent experimental manipulation and the participant's subsequent post-test, either

implicitly or explicitly (pretest sensitization). The way that this issue has been circumvented is by using a Solomon four-group design^{21,22}, in which one group of participants is given a pretest and one group is not, resulting in a two (dissonance, control) by two (with and without pretest) group design. The use of such a design allows us to use measures of attitude change within participant, while also allowing us to test for the presence of pretest sensitization²¹.

We implemented such a design in a way that we thought might minimize pretest sensitization. Following performance of the initial boring task, participants in pretest groups received instructions to rate sentences honestly and we started the experiment. After five sentences, the experiment was interrupted and the participants were instructed to respond counter-attitudinally in the same way as the without-pretest groups were; that is, in the dissonance group, participants were given the patient cover story and, in the control group, participants were told they were paid \$1 extra for each sentence that they responded to as though they were enjoying the scanner. Notably, the five sentences presented before the experimental manipulation included four randomly chosen neutral sentences and the sentence “Overall, I like my experience in the scanner”. Each participant's response to this item was taken as a pretest measure of scanner enjoyment.

We found that participants in the dissonance group changed their attitudes more than participants in the control group following counter-attitudinal behavior. Furthermore, dACC and anterior insula activation during counter-attitudinal behavior predicted the final attitude of participants in the dissonance group, but not in the control group. These results support both dissonance theory^{1,11} and the conflict theory of dACC functioning^{14,15}.

RESULTS

Across all participants, inter-correlation between post-test items was high (Cronbach's $\alpha = 0.897$), with little difference between experimental groups (dissonance group Cronbach's $\alpha = 0.892$, control group Cronbach's $\alpha = 0.904$). Thus, we can assume that the post-test items reflect a single, unidimensional latent psychological construct (attitude toward experience in the scanner) and that treating the composite post-test score as reflecting this construct is valid.

Analysis of composite final attitude scores by means of an experimental group (dissonance, control) by pretest presence (present, not present) ANOVA verified a main effect of experimental group ($F_{1,39} = 12.36$, $P = 0.001$). This showed that scanner enjoyment was greater for the without-pretest dissonance group ($M = 6.3$, $s.d. = 0.8$) and with-pretest dissonance group ($M = 6.0$, $s.d. = 1.1$) than for the without-pretest control group ($M = 5.1$, $s.d. = 1.3$) and the with-pretest control group ($M = 4.5$, $s.d. = 1.3$). Individual t tests verified this effect for both the without-pretests groups ($t_{22} = 2.51$, $P = 0.020$) and the with-pretest groups ($t_{17} = 2.43$, $P = 0.026$). These results verified the basic cognitive dissonance finding.

We did not observe a significant main effect of pretest presence ($F_{1,39} = 0.59$, $P = 0.45$) or a significant interaction ($F_{1,39} = 0.10$, $P = 0.75$), which suggested that the presence of a pretest measure had no

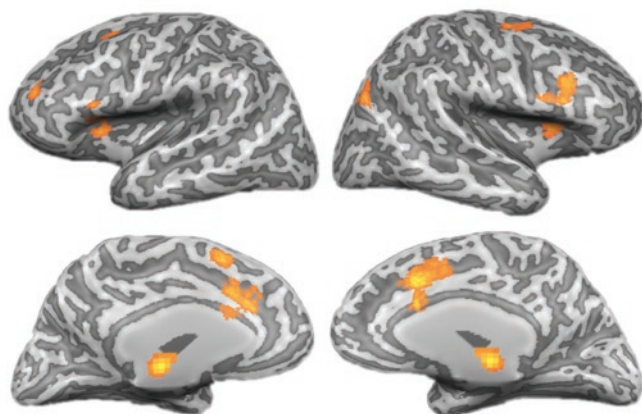


Figure 2 fMRI results. The results of a whole-brain experimental group (dissonance, control) by sentence type (target, neutral) interaction are shown, using participant as random factor ($\alpha = 0.01$, corrected, 297-mm³ voxel contiguity), displayed on an inflated cortical surface.

Table 1 Regions that showed a significant sentence by group interaction ($\alpha = 0.01$, corrected)

| Region | Side | Talairach coordinates | | | | Volume (mm ³) |
|---------------|------|-----------------------|-----|----|----------|---------------------------|
| | | x | y | z | BA | |
| dACC | L | -8 | 19 | 30 | 24/32 | 1,593 |
| dACC | R | 10 | 14 | 37 | 24/32 | 2,565 |
| pre-SMA | L | -7 | 11 | 52 | 6 | 1,134 |
| PMd | L | -26 | 5 | 44 | 6 | 648 |
| PMd | R | 27 | -6 | 47 | 6 | 783 |
| PMv | L | -55 | 12 | 24 | 44/45 | 432 |
| MFG/PMv | L | -27 | 42 | 24 | 46/44/45 | 999 |
| MFG | R | 38 | 21 | 32 | 9 | 1,647 |
| AI | L | -32 | 13 | 19 | – | 1,134 |
| AI | R | 30 | 15 | 15 | – | 351 |
| Angular gyrus | R | 37 | -75 | 28 | 39/19 | 1,431 |
| Midbrain | M | 0 | -18 | -1 | – | 1,944 |

AI, anterior insula; BA, Brodmann area; L, left; M, middle; R, right.

influence on the difference in the final attitude scores between the two experimental groups in this study²¹. When taking pretest scores as a covariate, the two by two ANCOVA continued to show a main effect of experimental group ($F_{1,39} = 14.25$, $P = 0.001$). This was also true for the comparison between the with-pretest groups ($t_{17} = 2.93$, $P = 0.010$). Not surprisingly, pretest scores (rated 1–6) did not differ between the with-pretest dissonance group ($M = 3.9$, s.d. = 1.7) and the with-pretest control group ($M = 3.8$, s.d. = 1.3) ($t_{17} = 0.18$, $P = 0.86$).

This latter analysis also revealed a positive relationship between pretest scores and final attitude scores in the with-pretest groups ($r = 0.64$, $P = 0.004$; controlling for experimental group); further analysis confirmed this positive correlation between pretest and final attitude scores in both the dissonance group (Spearman's $\rho = 0.74$, $P = 0.014$) and the control group (Spearman's $\rho = 0.68$, $P = 0.043$), with no significant between-group difference in the size of the correlation ($F_{1,15} = 1.21$, $P = 0.29$). The significant positive correlation between pretest and final attitudes scores suggests that these different measurements indeed reflected similar psychological constructs.

We analyzed fMRI data using a whole-brain sentence type (target, neutral) by experimental group (dissonance, control) ANOVA, using participant as random factor at $\alpha = 0.001$ (uncorrected). A cluster size threshold of 11 voxels corrected for multiple comparisons at $\alpha = 0.01$. The regions of activation included bilateral dACC, left pre-supplementary motor area (pre-SMA), bilateral anterior insula, bilateral middle frontal gyrus (MFG), bilateral dorsal premotor regions (PMd), right angular gyrus, left ventral premotor area (PMv) and the midbrain (Fig. 2 and Table 1). For each region identified in this analysis, we extracted average time courses of activation as the percentage change from baseline (Fig. 3). We then repeated the analyses, taking the per-participant, per-condition activation of the time course of activation following stimulus onset. The sentence type by experimental group interaction was not significant in the right angular gyrus ($F_{1,41} = 1.26$, $P = 0.27$) and the left vPMC ($F_{1,41} = 0.04$, $P = 0.85$); however, in the rest of these regions, this interaction remained significant ($F_{1,41}$ range = 4.62–14.46, all $P < 0.05$). Between-group t tests showed that activation to control sentences did not differ significantly between groups in any

of these regions (t_{41} range = 0.09–1.60, all $P > 0.1$). The activation to target sentences was greater in the dissonance group than in the control group, in all of the regions that also showed a significant sentence type by experimental group interaction ($t_{41} = 2.35–4.20$, all $P < 0.05$); these effects did not reach significance in the right angular gyrus ($t_{41} = 1.72$, $P = 0.093$) and the left vPMC ($t_{41} = 1.02$, $P = 0.31$).

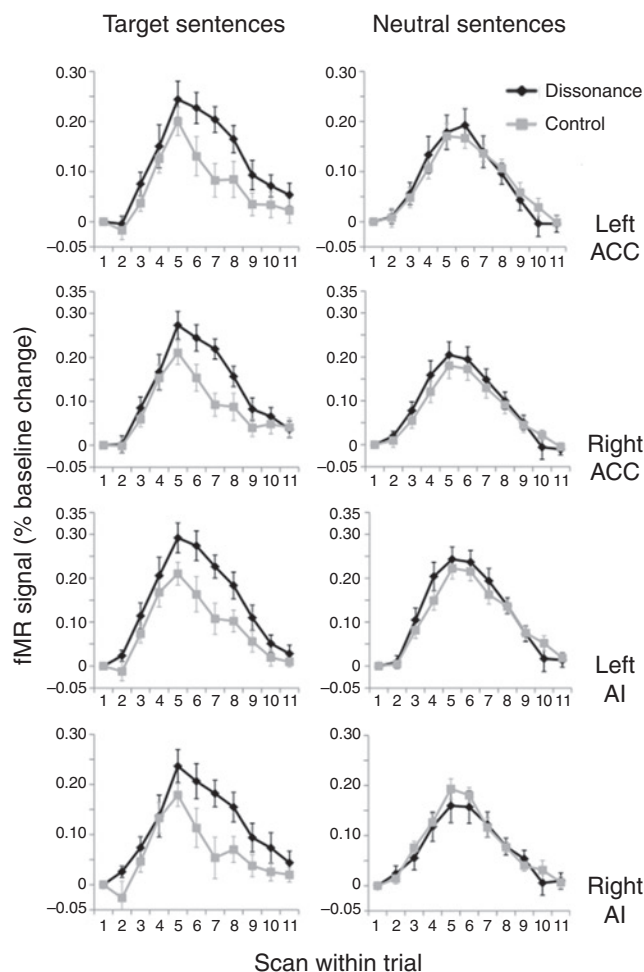


Figure 3 Time courses of activation of the left and right dACC and anterior insula (AI). Left, target sentences. Right, neutral sentences. Black indicates the dissonance group and gray indicates the control group. Error bars represent \pm s.e.m.

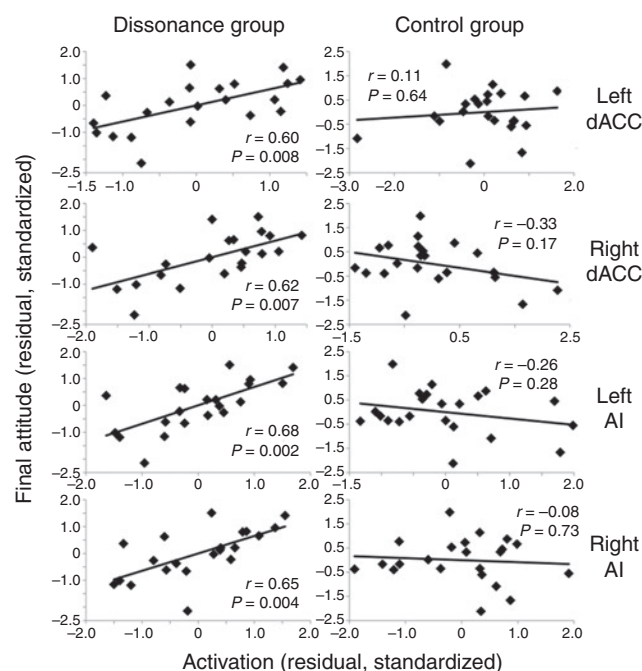


Figure 4 Partial correlation plots between activation and final attitude score, controlling for mean reaction time, pretest presence and pretest score. Left, dissonance group. Right, control group. Bilateral dACC and anterior insula activation predicted the final attitude score in the dissonance group, but not in the control group. Note that residual scores have been standardized within group.

To test whether pretest presence affected neural activity, we performed *post hoc* tests that included pretest presence as an additional between-group factor on the signals of each of these regions (averaged across voxels). These tests revealed no significant main or interaction effect of pretest presence for any of these regions in the three-way (experimental group by sentence type by pretest presence) interaction ($F_{1,39}$ range = 0.01–2.05, all $P > 0.1$) or when the two-way sentence type by pretest presence interaction was tested separately for the dissonance group ($F_{1,39}$ range = 0.02–1.56, all $P > 0.1$) and control group ($F_{1,39}$ range = 0.0004–2.96, all $P > 0.1$). The lack of any significant effects of pretest presence on any of these findings again suggested that the pretest did not influence the pattern of neural activation²¹.

We next investigated the relationship of brain activation to attitude change by correlating the activation to the target sentences with final attitude score for each region, and correcting for pretest score, pretest presence and mean reaction time to the target sentences. In the dissonance group, these partial correlations were significant for the bilateral dACC and bilateral anterior insula regions (partial r range = 0.60–0.68, all $P < 0.01$; **Fig. 4**); for those regions, these correlations were not significant in the control group (partial r range = –0.33–0.11, all $P > 0.1$). ANCOVA analyses verified that for the bilateral dACC and left anterior insula, these correlations were greater in the dissonance group than in the control group ($F_{1,35}$ range = 4.10–9.43, all $P < 0.05$). For the right anterior insula, the ANCOVA was marginally significant ($F_{1,35} = 3.186$, $P = 0.083$). We did not find any strong correlations for any of the remaining regions (**Supplementary Results**).

Instead of using pretest scores as a covariant of non-interest, it could be argued that taking the difference between pretest and final attitude scores might constitute a more direct, within-participant index of attitude change. For this reason, we calculated z scores of pretest and final attitude scores and took the difference between these two scores for each

participant. These attitude change scores were greater for the dissonance group than the control group ($t_{17} = 2.40$, $P = 0.028$). Correcting for reaction time, activation to the target sentences significantly predicted attitude change, as indexed by this difference measure, in the dissonance group in bilateral dACC and left anterior insula (partial r range = 0.72–0.75, all $P < 0.05$), but not in the control group (partial r range = –0.40–0.48, all $P > 0.2$). ANCOVA analyses verified that this correlation was significantly greater in the dissonance than in the control group in the right dACC ($F_{1,14} = 7.82$, $P = 0.014$) and left anterior insula ($F_{1,14} = 5.74$, $P = 0.031$) and was marginally significant in left dACC ($F_{1,14} = 3.69$, $P = 0.075$). These effects were not significant in right anterior insula or in any of the other identified regions (all $P > 0.1$).

We next tested correlations between activation and pretest scores, leaving aside final attitude scores. Activation correlated negatively with pretest scores in the dissonance group in the left dACC ($r = -0.79$, $P = 0.007$), the left PMd ($r = -0.70$, $P = 0.025$) and the left MFG ($r = -0.71$, $P = 0.022$); a trend toward significance was found in the pre-SMA ($r = -0.55$, $P = 0.099$). These correlations were not significant in the control group in these regions (r range = –0.11–0.17, all $P > 0.7$); however, the ANCOVA was only significant for the left dACC ($F_{1,15} = 6.36$, $P = 0.024$) and left MFG ($F_{1,15} = 5.88$, $P = 0.028$). With the exception of the right MFG, which showed a nonsignificant positive correlation ($r = 0.07$, $P = 0.84$), all of the other regions showed nonsignificant negative correlations in the dissonance group between activation and pretest score (r range = –0.47 to –0.11, all $P > 0.1$). None of the identified regions showed a significant correlation in the control group (r range = –0.30–0.26, all $P > 0.4$). Thus, the more that participants in the dissonance group ‘lied’, the greater was their activation of their left dACC, left PMd and left MFG.

DISCUSSION

These findings are consistent with a number of prior observations. Both cognitive dissonance and dACC and anterior insula activation have been associated with negative affect and autonomic arousal^{23–28}. These regions might therefore be responsible for representing or triggering the negative affect and related autonomic arousal associated with the dissonance. Another set of related observations is dACC activation during lying in guilty knowledge tasks^{29,30}. Our data expand on those findings, indicating that dACC activity during the counter-attitudinal argument, which is similar to lying, predicts subsequent attitude change, but only when counter-attitudinal behavior conflicts with other cognitions. In addition, one earlier study that could be interpreted as cognitive dissonance concerned US partisans reasoning about statements that either matched or conflicted with their pre-existing beliefs³¹. Although that study did not measure attitude change or correlations between brain activation and attitude change, it identified a dACC region similar to the ones engaged here. In short, our results are consistent with theories of cognitive dissonance that emphasize conflict between different cognitions, such as the original theory^{1,2}. In particular, our results are consistent with the action-based model of cognitive dissonance, which posits that conflict between cognitions evokes an aversive state because it potentially interferes with unconflicted, effective, goal-driven action^{6,32}. This theory has explicitly predicted involvement of the dACC in cognitive dissonance^{6,32}. Consistently, although we did not measure distress in a systematic way, the majority of participants in the dissonance groups indicated during debriefing that they felt uncomfortable or distressed about lying to the ‘patient’; none of the participants in the control group indicated feeling negatively toward responding as though their experience in the scanner was a pleasant one.

Our results do not appear to support other theories of attitude change in the induced compliance procedure that are not based on the

notion of conflict between cognitions. First, an important alternative to cognitive dissonance theory is self-perception theory³³. This theory poses that, instead of attitude change driven by conflict between cognitions, participants reflect on their own behavior when giving their final attitude measure and, except when they can attribute their behavior to a reward, simply infer their attitudes from their behavior. In other words, this theory assumes that the pattern of findings of the induced compliance procedure is only related to processes that occur during the phase in which participants give their final attitude measure and not to processes taking place during the actual counter-attitudinal argument. Our data, which relate neural activity during the counter-attitudinal argument to attitude change, is therefore not consistent with self-perception theory. A fervent supporter of this theory might argue that the relationship between neural activity and attitude change that we observed is related to encoding; those who encoded their behavior 'better' might have simply been better at retrieving their behavior when they were giving their final attitude. We do not believe that this line of reasoning is valid. If this were the case, we would expect to see activity in brain regions typically related to encoding into episodic memory, such as the medial temporal lobe^{34–36}, rather than in the dACC and anterior insula. Our data are therefore consistent with the reported finding that individuals with anterograde amnesia show similar amounts of attitude change as healthy participants³⁷.

A second competing theory is the new look theory¹². This theory holds that individuals change their attitude to reduce the negative consequences of their behavior that they feel personally responsible for, rather than being driven by the motivation to reduce conflict between two inconsistent cognitions. In our study, the negative consequence might be that the hypothetical patient will undergo an uncomfortable scanning session. Attitude change takes place to reduce these negative consequences; if the participant starts to believe that the scanner isn't such an uncomfortable environment after all, it's probably not that uncomfortable for the patient either, and the consequences of the participant's counter-attitudinal behavior were not that negative after all. The region most often associated with the anticipation of negative outcomes is the orbitofrontal cortex rather than the dACC^{38–40}. For instance, regret and its anticipation and avoidance have been tightly linked to orbitofrontal activation, rather than dACC⁴¹. We did not observe differential activity in orbitofrontal regions. However, it is not completely clear what the neural correlates are of the anticipation of negative consequences of actions that one feels personally responsible for. We therefore tentatively suggest that our data are not consistent with this theory either; however, more research is needed before we can be definitive about this issue. In sum, the involvement of the dACC and anterior insula, and their association with both conflict monitoring, including conflict with the self⁴², and negative affect and physiological arousal, provides insights into the importance of conflict or dissonance *per se* in attitude change in this procedure, and thus argues against the view that dissonance-induced attitude change is a result of inferences drawn from self-observation³³ or the anticipation of negative consequences¹² (see also **Supplementary Discussion and Supplementary Fig. 1**).

A possible source of confound in our study is the fact that the dissonant condition was more social than the control condition. This is an intrinsic confound in our design and of many other studies using the induced compliance procedure; this is because, as explained, the phenomenon tends to be greater when there are real-life consequences to the counter-attitudinal behavior¹². For example, the participant might have felt empathy toward the patient, and perhaps the dACC and anterior insula activation reflected this empathy. Indeed, both dACC and anterior insula have been associated with empathetic concern^{43–45} and it could therefore be argued that engagement of these regions reflects

empathy, rather than conflict or dissonance. However, this does not appear to be consistent with our finding that scanner comfort, as measured by the pretest, was negatively correlated with dACC activation in the dissonance group. Comfort and empathy or prosocial behavior tend to be positively correlated with each other^{46,47}; thus, if the dACC activation reflected empathy, it would most likely have correlated positively rather than negatively with scanner comfort. Furthermore, empathy in and of itself cannot explain the correlation between dACC and anterior insula activation and attitude change.

A recent fMRI study⁴⁸ used a different cognitive dissonance procedure. This study examined the neural activity during the final attitude score measurement, rather than during the dissonance-arousing behavior. Participants in that study performed a version of the 'free choice' or 'spreading of alternatives' procedure of cognitive dissonance; in this procedure, participants chose between two alternatives that they had previously rated either similarly or differently. It is commonly found in this procedure that participants align their attitudes to be more consistent with the chosen item and less with the rejected item following a difficult choice between two similarly rated items than following an easy choice between two differently rated items. Thus, attitude change is greater following a difficult choice than following an easy choice. Changes in the desirability of the chosen and rejected items were reflected in changes in activation of the dorsal striatum, commonly associated with reward assessment⁴⁹; thus, the conflict during the difficult choice resulted in changes in the hedonic value of the chosen and rejected items. These results are complementary to our own, as we expect that, had we somehow managed to measure neural activity during the final attitude score measurement, we would have observed changes in the dorsal striatum in the dissonance group reflecting their attitude change about the scanner environment. An important topic for future research and theory might focus on the actual mechanism of dissonance-induced attitude change, in addition to the causes of attitude change (conflict represented in the dACC) and its consequences (altered hedonic value of the object of attitude, reflected by modulated activation of the dorsal striatum).

Traditionally, researchers have assumed that the processes underlying dissonance and subsequent attitude change are the same across the different procedures used to study these processes, such as the induced compliance procedure that we used and the above-mentioned free choice procedure. We would predict that dissonance in other dissonance procedures would also be accompanied by dACC and anterior insula activation, and that the magnitude of these activations predicts attitude change in those procedures, as well. These hypotheses can be readily tested in future fMRI studies using different cognitive dissonance procedures.

Cognitive dissonance is an important concept in psychology and is used to explain and predict attitude change in a wide range of human behavior. Our results shed light on the cognitive and neurobiological basis of dissonance, and indicate that the magnitude of conflict-related dACC and anterior insula activation predicts the subsequent attitude change. This result supports the core assumption of dissonance theory, that attitude change in cognitive dissonance is driven by conflict¹. It is the basis for a neural prediction of attitude change. Our findings have considerable implications for understanding attitude change in a wide range of contexts in which cognitive dissonance theory has found broad practical application, including politics, marketing, management and the evaluation of psychotherapeutic interventions.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/natureneuroscience/>.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We are grateful to M. Greenberg, G. Hein and K. Ochsner for discussions, and to S. Ravizza, D. Tamir, S. Ursu and A. Watrous for help with data collection. This research was supported by grants from the US National Institute of Mental Health and the Burroughs-Wellcome Foundation to C.S.C.

AUTHOR CONTRIBUTIONS

V.v.V., J.W.S. and C.S.C. designed the study. M.K.K. gathered fMRI data and administered questionnaires. V.v.V. developed stimuli, gathered pilot data and conducted analyses. V.v.V. wrote the paper with contributions from M.K.K., J.W.S. and C.S.C.

Published online at <http://www.nature.com/natureneuroscience/>.

Reprints and permissions information is available online at <http://www.nature.com/reprintsandpermissions/>.

1. Festinger, L. *A Theory of Cognitive Dissonance* (Stanford University Press, Stanford, California, 1957).
2. Festinger, L. & Carlsmith, J.M. Cognitive consequences of forced compliance. *J. Abnorm. Soc. Psychol.* **58**, 203–210 (1959).
3. Harmon-Jones, E. & Mills, J. *Cognitive Dissonance: Progress on a Pivotal Theory in Social Psychology* (American Psychological Association, Washington, DC, 1999).
4. Jones, E.E. Major developments in social psychology during the past five decades. in *Handbook of Social Psychology: Vol. 1. Theory and Methods* (eds Lindzey, G. & Aronson, E.) 47–107 (Random House, New York, 1985).
5. Losch, D.E. & Cacioppo, J.T. Cognitive dissonance may enhance sympathetic tonus, but attitudes are changed to reduce negative affect rather than arousal. *J. Exp. Soc. Psychol.* **26**, 289–304 (1990).
6. Harmon-Jones, E. Contributions from research on anger and cognitive dissonance to understanding the motivational functions of asymmetrical frontal brain activity. *Biol. Psychol.* **67**, 51–76 (2004).
7. Elliot, A.J. & Devine, P.G. On the motivational nature of cognitive dissonance: dissonance as psychological discomfort. *J. Pers. Soc. Psychol.* **67**, 382–394 (1994).
8. Croyle, R. & Cooper, J. Dissonance arousal: physiological evidence. *J. Pers. Soc. Psychol.* **45**, 782–791 (1983).
9. Linder, D.E., Cooper, J. & Jones, E.E. Decision freedom as a determinant of the role of incentive magnitude in attitude change. *J. Pers. Soc. Psychol.* **6**, 245–254 (1967).
10. Hoyt, M.F., Henley, M.D. & Collins, B.E. Studies in forced compliance: confluence of choice and consequence on attitude change. *J. Pers. Soc. Psychol.* **23**, 205–210 (1972).
11. Harmon-Jones, E., Gerdjikov, T. & Harmon-Jones, C. 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 (2008).
12. Cooper, J. & Fazio, R.H. A new look at dissonance theory. in *Advances in Experimental Social Psychology* (ed. Berkowitz, L.) 229–264 (Academic Press, Orlando, Florida, 1984).
13. Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S. & Cohen, J.D. Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624–652 (2001).
14. van Veen, V. & Carter, C.S. Conflict and cognitive control in the brain. *Curr. Dir. Psychol. Sci.* **15**, 237–240 (2006).
15. Botvinick, M.M., Cohen, J.D. & Carter, C.S. Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn. Sci.* **8**, 539–546 (2004).
16. Kerns, J.G. *et al.* Anterior cingulate conflict monitoring and adjustments in control. *Science* **303**, 1023–1026 (2004).
17. van Veen, V. & Carter, C.S. Separating semantic conflict and response conflict in the Stroop task: a functional MRI study. *Neuroimage* **27**, 497–504 (2005).
18. Jones, A.D., Cho, R.Y., Nystrom, L.E., Cohen, J.D. & Braver, T.S. A computational model of anterior cingulate function in speeded response tasks: effects of frequency, sequence and conflict. *Cogn. Affect. Behav. Neurosci.* **2**, 300–317 (2003).
19. Van Overwalle, F. & Jordens, K. An adaptive connectionist model of cognitive dissonance. *Pers. Soc. Psychol. Rev.* **6**, 204–231 (2002).
20. Shultz, T.R. & Lepper, M.R. Cognitive dissonance reduction as constraint satisfaction. *Psychol. Rev.* **103**, 219–240 (1996).
21. Braver, M.C.W. & Braver, S.L. Statistical treatment of the Solomon four-group approach: a meta-analytic approach. *Psychol. Bull.* **104**, 150–154 (1988).
22. Solomon, R.L. An extension of control group designs. *Psychol. Bull.* **46**, 137–150 (1949).
23. Critchley, H.D. Neural mechanisms of autonomic, affective and cognitive integration. *J. Comp. Neurol.* **493**, 154–166 (2005).
24. Critchley, H.D. *et al.* Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. *Brain* **126**, 2139–2152 (2003).
25. Hajcak, G., McDonald, N. & Simons, R.F. Error-related psychophysiology and negative affect. *Brain Cogn.* **56**, 189–197 (2004).
26. Eisenberger, N.I., Lieberman, M.D. & Williams, K.D. Does rejection hurt? An fMRI study of social exclusion. *Science* **302**, 290–292 (2003).
27. Luu, P., Collins, P. & Tucker, D.M. Mood, personality and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error processing. *J. Exp. Psychol. Gen.* **129**, 43–60 (2000).
28. Craig, A.D., Reiman, E.M., Evans, A.C. & Bushnell, M.C. Functional imaging of an illusion of pain. *Nature* **384**, 258–260 (1996).
29. Nuñez, J.M., Casey, B.J., Egner, T., Hare, T. & Hirsch, J. Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage* **25**, 267–277 (2005).
30. Langleben, D.D. *et al.* Brain activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage* **15**, 727–732 (2002).
31. Westen, D., Blagov, P.S., Harenski, K., Kilts, C. & Hamann, S. Neural bases of motivated reasoning: an fMRI study of emotional constraints on partisan political judgment in the 2004 U.S. presidential election. *J. Cogn. Neurosci.* **18**, 1947–1958 (2006).
32. Harmon-Jones, E. & Harmon-Jones, C. Action-based model of dissonance: a review of behavioral, anterior cingulate, and prefrontal cortical mechanisms. *Soc. Pers. Psychol. Compass* **2**, 1518–1538 (2008).
33. Bem, D.J. Self-perception: an alternative interpretation of cognitive dissonance phenomena. *Psychol. Rev.* **74**, 183–200 (1967).
34. Kirchoff, B.A., Wagner, A.D., Maril, A. & Stern, C.E. Prefrontal circuitry for episodic encoding and subsequent memory. *J. Neurosci.* **20**, 6173–6180 (2000).
35. Wagner, A.D. *et al.* Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* **281**, 1188–1191 (1998).
36. Brewer, J.B., Zhao, Z., Desmond, J.E., Glover, G.H. & Gabrieli, J.D. Making memories: brain activity that predicts how well visual experience will be remembered. *Science* **281**, 1185–1187 (1998).
37. Lieberman, M.D., Ochsner, K.N., Gilbert, D.T. & Schacter, D.L. Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychol. Sci.* **12**, 135–140 (2001).
38. Ursu, S. & Carter, C.S. Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: implications for neuroimaging studies of decision-making. *Brain Res. Cogn. Brain Res.* **23**, 51–60 (2005).
39. Amodio, D.M. & Frith, C.D. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* **7**, 268–277 (2006).
40. Ursu, S., Clark, K.A., Stenger, V.A. & Carter, C.S. Distinguishing expected negative outcomes from preparatory control in the human orbitofrontal cortex. *Brain Res.* **1227**, 110–119 (2008).
41. Coricelli, G. *et al.* Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. Neurosci.* **8**, 1255–1262 (2005).
42. Takahashi, H. *et al.* When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. *Science* **323**, 937–939 (2009).
43. Singer, T. *et al.* Empathy for pain involves the affective, but not sensory components of pain. *Science* **303**, 1157–1162 (2004).
44. Immordino-Yang, M.H., McColl, A., Damasio, H. & Damasio, A. Neural correlates of admiration and compassion. *Proc. Natl. Acad. Sci. USA* **106**, 8021–8026 (2009).
45. Hein, G. & Singer, T. I feel how you feel but not always: the empathic brain and its modulation. *Curr. Opin. Neurobiol.* **18**, 153–158 (2008).
46. Lamm, C., Batson, C.D. & Decety, J. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *J. Cogn. Neurosci.* **19**, 42–58 (2007).
47. Batson, C.D. *et al.* "... As you would have them do unto you": does imagining yourself in the other's place stimulate moral action? *Pers. Soc. Psychol. Bull.* **29**, 1190–1201 (2003).
48. Sharot, T., De Martino, B. & Dolan, R.J. How choice reveals and shapes expected hedonic outcome. *J. Neurosci.* **29**, 3760–3765 (2009).
49. Delgado, M.R., Locke, H.M., Stenger, V.A. & Fiez, J.A. Dorsal striatum responses to reward and punishment: effects of valence and magnitude manipulations. *Cogn. Affect. Behav. Neurosci.* **3**, 27–38 (2003).

ONLINE METHODS

Participants. In total, 53 people participated in the experiment after having provided written informed consent in accordance with the Institutional Review Board of the University of California at Davis. Participants who indicated that they had been in an fMRI scanner before were declined participation in the study. We included 13 participants in the without-pretest control group (five females, eight males; mean age = 25 years old, s.d. = 2.7) and nine participants in the with-pretest control group (six females, three males; mean age = 23 years old, s.d. = 3.4). Initially, we included 20 participants in the without-pretest dissonance group. During debriefing, eight of these participants admitted to having doubts about or not believing the cover story and were excluded from the analyses. We also excluded one additional participant who opted not to comply with the instructions from the analyses, leaving a total of 11 participants in the without-pretest dissonance group (six females, five males; mean age = 24 years old, s.d. = 4.7). Initially, we included 11 participants in the with-pretest dissonance group (seven females, three males; mean age = 24 years old, s.d. = 4.3). None of these participants admitted to having doubts about the cover story. We excluded one participant for not complying with the instructions, leaving ten participants in the with-pretest dissonance group.

Materials. Stimuli were presented using E-Prime software (Psychological Software Tools). Throughout the task, the outline of six squares was visible at the bottom of the screen, with a + sign to the left, indicating agreement, and a – sign to the right, indicating disagreement. The session started with a 12,000-ms fixation screen.

Each trial started with a sentence being presented in the center of the screen for 8,000 ms, after which the sentence was replaced by a 7,000-ms fixation. Following a button press, the square corresponding at the bottom of the screen lit up; the square would dimmed again 1,500 ms before the onset of the next trial.

During the experiment, neutral and target sentences were presented in random order. Target sentences consisted of 25 sentences describing the participant's subjective experience in the scanner. Of these, 12 were phrased in negative terms, which we expected the participants to disagree with, and 13 were phrased in positive terms, which we expected the participants to agree with (Supplementary Table 1). In addition, we used 50 sentences as neutral stimuli, 25 of which concerned general attitudes and 25 of which were factual statements about the participant, the task and the scanner (Supplementary Table 2). Neutral and target sentences were matched for length.

Procedure. After providing consent, participants were shown the magnetic resonance facility and control room. They were informed that what they would see on screen while in the scanner was also visible on a screen in the control room. Participants in all groups first performed a forced-choice response task in a standard cognitive neuroimaging experiment, which lasted about 45 min (see Supplementary Methods). Following this experiment, a high-density structural magnetization-prepared rapid acquisition gradient echo (MPRAGE) image was acquired, which lasted about 9 min.

While participants were performing the first task, we randomly assigned them to either the control or dissonance group. Following this part of the scanning session, participants were taken partway out of the scanner and received instructions. Participants responded using a left- and a right-hand button glove. Participants were instructed to respond to each sentence with a button press, rating each sentence from completely agree to completely disagree (left ring finger = completely agree, left middle finger = mostly agree, left index finger = somewhat agree, right index finger = somewhat disagree, right middle finger = mostly disagree, right middle finger = completely disagree). Further instructions depended on experimental group.

Participants in with-pretest groups were first instructed to rate sentences honestly, without group-specific instruction, and the experiment was started. After five sentences, we interrupted the scanner, and we now instructed participants to respond counter-attitudinally in the same way as participants in without-pretest groups (that is, in the dissonance group, we gave participants the patient cover story, and in the control group, we told participants that they would be paid \$1 extra for each sentence they responded to as though they were enjoying the scanner). Critically, the five sentences presented before the experi-

mental manipulation included four randomly chosen neutral sentences and the sentence "Overall, I like my experience in the scanner". We took each participant's response to this item as a pretest measure of scanner enjoyment.

Following the scanning session, participants were led to a private waiting room (when participants in the dissonance groups exited the scanner and were led through the scanner control room to the waiting room, it was made apparent that several people were present in the scanner control room, including a 'patient', so as to not arouse suspicion). A questionnaire was then administered that contained the same target sentences; participants were now asked to respond to them honestly. There was an 8.9-cm horizontal line next to each sentence on the questionnaire; participants indicated by tick mark how much they agreed with the sentence. Instructions for filling in the questionnaire were given verbally and in writing in the top header of each sheet of the questionnaire (left = completely disagree, right = completely agree). After participants filled in the questionnaires, we carefully debriefed them. We excluded participants who admitted to having doubts about or did not believe the validity of the patient cover story from the analyses ($n = 8$ in the without-pretest dissonance group and $n = 0$ in the with-pretest dissonance group).

Attitude scoring. Final attitude questionnaire items were scored by measuring, for negatively phrased items, the distance (in cm) to the tick mark from the right, and for positively phrased items, the distance from the left. Thus, higher scores reflected greater enjoyment for all items. We then computed composite final attitude scores for each participant by averaging across all items of the questionnaire.

fMRI data acquisition and analysis. Functional images were acquired with a Siemens 3 T whole-body MRI system, using T2*-weighted gradient-recalled echo (field of view = 220 mm; matrix = 64×64 ; 28 oblique axial slices; slice thickness = 4.0 mm; repetition time = 1,500 ms; error time = 25 ms; flip angle = 90°). Data were preprocessed and analyzed using BrainVoyager QX software (Brain Innovation). The first eight images, corresponding to the initial 12,000-ms fixation screen, were discarded from the analysis. The remaining 825 images were preprocessed using interscan slice time correction, three-dimensional motion correction, three-dimensional Gaussian spatial filtering (full-width half maximum = 8 mm) and temporal high-pass filtering using a low cutoff frequency of three cycles per run. For all participants, high-density three-dimensional images (MPRAGEs) of the brain were acquired in between the 'boring' task and the attitude task; for each participant, the functional data were aligned to these images, transformed into Talairach space and interpolated into a $3 \times 3 \times 3$ -mm³ voxel size.

We constructed four regressors (target and neutral sentences, stimulus and response-locked) for each participant and convolved with a hemodynamic response function. These were entered into a first-level multiple regression. We then conducted a second-level, whole-brain experimental group (dissonance, control) by sentence type (target, neutral) ANOVA, once on the obtained betas associated with the stimulus-locked regressors and once on the obtained betas associated with the response-locked regressors, using participant as random factor (thresholded at $\alpha = 0.001$, uncorrected). For all analyses reported, temporal autocorrelation in the fMRI time series was corrected using an autoregressive function. As no significant results were obtained for the stimulus-locked regressors, only the results using the response-locked regressors are discussed. Monte Carlo simulations of the random process of image generation, followed by the injection of spatial correlations between neighboring voxels and voxel intensity thresholding, as implemented in BrainVoyager QX⁵⁰, were used to select a cluster size threshold that corrected the obtained statistical map for multiple comparisons at $\alpha = 0.01$.

For the activated regions, we obtained time courses of activation to the two sentence types for each group, as percentage change from baseline. We then performed partial correlations between activation and final composite attitude score by taking the average scan for each participant that represented the greatest between-group difference in response to the target sentences, controlling for the per-participant mean reaction time to the target sentences, pretest presence and pretest score.

50. Goebel, R., Esposito, F. & Formisano, E. Analysis of functional image analysis contest (FIAC) data with BrainVoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* **27**, 392–401 (2006).