

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Irani, Fatemeh; Maunula, Sini; Muotka, Joonas; Leppäniemi, Matti; Kukkonen, Maria; Monto, Simo; Parviainen, Tiina

**Title:** Brain dynamics of recommendation-based social influence on preference change : A magnetoencephalography study

**Year:** 2022

**Version:** Published version

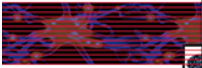
**Copyright:** © 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis

**Rights:** CC BY-NC-ND 4.0

**Rights url:** <https://creativecommons.org/licenses/by-nc-nd/4.0/>

**Please cite the original version:**

Irani, F., Maunula, S., Muotka, J., Leppäniemi, M., Kukkonen, M., Monto, S., & Parviainen, T. (2022). Brain dynamics of recommendation-based social influence on preference change : A magnetoencephalography study. *Social Neuroscience*, 17(5), 397-413.  
<https://doi.org/10.1080/17470919.2022.2126001>



# Brain dynamics of recommendation-based social influence on preference change: A magnetoencephalography study

Fatemeh Irani, Sini Maunula, Joonas Muotka, Matti Leppäniemi, Maria Kukkonen, Simo Monto & Tiina Parviainen

To cite this article: Fatemeh Irani, Sini Maunula, Joonas Muotka, Matti Leppäniemi, Maria Kukkonen, Simo Monto & Tiina Parviainen (2022): Brain dynamics of recommendation-based social influence on preference change: A magnetoencephalography study, *Social Neuroscience*, DOI: [10.1080/17470919.2022.2126001](https://doi.org/10.1080/17470919.2022.2126001)

To link to this article: <https://doi.org/10.1080/17470919.2022.2126001>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 26 Sep 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

## Brain dynamics of recommendation-based social influence on preference change: A magnetoencephalography study

Fatemeh Irani <sup>a,b</sup>, Sini Maunula<sup>a,b</sup>, Joonas Muotka <sup>a</sup>, Matti Leppäniemi<sup>c</sup>, Maria Kukkonen<sup>a</sup>, Simo Monto <sup>a,b</sup> and Tiina Parviainen <sup>a,b</sup>

<sup>a</sup>Department of Psychology, University of Jyväskylä, Jyväskylä, Finland; <sup>b</sup>Centre for Interdisciplinary Brain Research, University of Jyväskylä, Jyväskylä, Finland; <sup>c</sup>School of Business and Economics, University of Jyväskylä, Jyväskylä, Finland

### ABSTRACT

People change their preferences when exposed to others' opinions. We examine the neural basis of how peer feedback influences an individual's recommendation behavior. In addition, we investigate if the personality trait of 'agreeableness' modulates behavioral change and neural responses. In our experiment, participants with low and high agreeableness indicated their degree of recommendation of commercial brands, while subjected to peer group feedback. The associated neural responses were recorded with concurrent magnetoencephalography. After a delay, the participants were asked to reevaluate the brands. Recommendations changed consistently with conflicting feedback only when peer recommendation was lower than the initial recommendation. On the neural level, feedback evoked neural responses in the medial frontal and lateral parietal cortices, which were stronger for conflicting peer opinions. Conflict also increased neural oscillations in 4–10 Hz and decreased oscillations in 13–30 Hz in medial frontal and parietal cortices. The change in recommendation behavior was not different between the low and high agreeableness groups. However, the groups differed in neural oscillations in the alpha and beta bands, when recommendation matched with feedback. In addition to corroborating earlier findings on the role of conflict monitoring in feedback processing, our results suggest that agreeableness modulates neural processing of peer feedback.

### ARTICLE HISTORY

Received 7 December 2021  
Revised 15 August 2022  
Published online 25  
September 2022

### KEYWORDS





Social influence;  
magnetoencephalography;  
recommendation;  
agreeableness; neuronal  
oscillation

## Introduction

The question of whether others influence our behavior has been approached from a social psychological perspective broadly indicating that people's opinions are not guided only by their subjective values and personal experiences but greatly affected by other people (Cialdini & Goldstein, 2004). Two main motivations, seeking social approval and validating the correctness of opinion, underlie this influence. Correspondingly, it can lead to compliance, where people go along with the majority publicly while their internal opinions stay intact, or conformity, where the influence of others produces genuine attitude change (Cialdini et al., 1999). Positive and negative aspects of social influence have been studied in numerous domains. In fact, providing people with information about what others do in order to modify their behaviors often works better than increasing their factual knowledge about the matter (Miller & Prentice, 2016). Several studies have investigated this by using norms to increase pro-environmental behavior (Gugenishvili et al., 2021), health-related behavior (Templeton et al., 2016), and, related to the current

study, consumer behavior, where product ratings and reviews from others are shown to significantly affect preferences and behavior (Muchnik et al., 2013).

Neuroimaging studies have sought to provide insights into social influence by uncovering neural mechanisms that underlie changing one's behavior in line with other people's behavior or opinions in various forms of social influence. Exposure to the opinions of others has been shown to alter preferences for facial attractiveness (Klucharev et al., 2009; Zaki et al., 2011; Shestakova et al., 2013), trustworthiness (Zubarev et al., 2017), music (Berns et al., 2010; Campbell-Meiklejohn et al., 2010), and food choices (Nook & Zaki, 2015). Cascio et al. (2015) investigated the effect of social influence on an individual's preference change, where preference was assessed by making a recommendation to a peer. In this study, participants rated mobile game applications on a five-point scale in terms of their tendency to recommend them to a friend. Later, in a functional magnetic resonance imaging (fMRI) session, they were reminded about their initial rating followed by ratings of their peers, which could be higher, lower or

**CONTACT** Simo Monto  [simo.p.monto@jyu.fi](mailto:simo.p.monto@jyu.fi)  Department of Psychology, University of Jyväskylä, Jyväskylä; Tiina Parviainen  [tiina.m.parviainen@jyu.fi](mailto:tiina.m.parviainen@jyu.fi)  Department of Psychology, University of Jyväskylä, Jyväskylä

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

the same as the participant's rating, with the opportunity to update their initial rating. Results were similar to the effect of social influence on personal preference change: subjects tended to change their initial recommendation ratings in accordance with the group rating. Recently, Baek et al. (2021) used a similar paradigm, but instead of showing lower or higher group ratings, they provided a short recommendation text feedback from a peer, ranging in valence from negative to positive, similar to real-life online recommendations. Results showed that both positive and negative peer recommendations significantly affected participants' final ratings. In comparison to positive feedback, negatively framed peer recommendations had a larger effect on rating change, indicating that negativity may propagate more strongly in contexts related to consumer behavior and purchase decision.

A number of studies on neural mechanisms of social influence using fMRI have revealed that being exposed to a group opinion engages brain regions including medial frontal cortex (MFC), ventral striatum (VS), and anterior insula (AI). Activation in MFC and AI is positively correlated with social conflict and activation in VS is negatively correlated with social conflict (Campbell-Meiklejohn et al., 2010; Klucharev et al., 2009). These brain regions overlap with the network involved in reinforcement learning (RL), where involvement of VS, dorsal MFC, and AI reflect the differences between expectations and outcomes, that is, the general prediction error signals the need for behavioral adjustment (Zhang et al., 2020). Accordingly, Klucharev and colleagues suggested that personal opinion is adjusted by RL mechanisms toward social norms. In this account, the discrepancy between individual and group opinion is experienced as an error and requires correction of the deviance aligning one's opinion with normative opinions. In contrast, an agreement with group norms is experienced as a rewarding outcome, and no conformal adjustments follow.

Berns et al. (2010), in an fMRI study of the effect of group norm on adolescent's evaluations of music popularity, showed a positive correlation of AI and dorsal MFC activation, with one's tendency to change his/her evaluation: the higher the activation in these areas, the more likely the participant would change their rating toward a group norm. Because AI and dMFC activation is associated with aversive states and physiological arousal, researchers interpreted that conformity behavior in popularity ratings in adolescents works through the generated anxiety by the conflict between one's own preferences and those of the peer group. Supporting this idea, experiencing social exclusion showed neural activity in AI and dorsal anterior cingulate cortex (dACC)

as was observed in conflict with the group, indicating that activation in this area might reflect the threat of social rejection and call for behavior modification to keep individuals in harmony with the group norm (Wasylyshyn et al., 2018). Cascio et al. (2015) and Baek et al. (2021) suggested two major systems to be involved in the change of opinion based on group recommendation: First, the mentalizing system including medial prefrontal cortex (mPFC), temporoparietal junction (TPJ), and precuneus activates after observing conflict with group opinion. Second, the valuation system, which includes striatum and orbitofrontal cortex (OFC), activates when subjects change their opinion to match group recommendation.

Electroencephalographic (EEG) studies of social influence demonstrated that a mismatch between an individual and group opinion elicits a fronto-central voltage deviation in the event-related potential (ERP) component known as feedback-related negativity (FRN). FRN occurs between 200 and 400 ms post-onset of conflicting group feedback and localizes to MFC. The FRN has been suggested to reflect the neural response similar to punishment (negative) prediction errors (Chen et al., 2012; Kim et al., 2012; Shestakova et al., 2013; Yu & Sun, 2013). A mismatch between individual and group opinions also evoked a stronger N400 like component instead of FRN, which is predominantly involved in conceptual processing and violation of semantic expectations in language studies (Huang et al., 2014; Schnuerch et al., 2016). FRN is suggested to reflect reward prediction errors when an anticipated reward is omitted regardless of whether the violated expectancy is negative or positive (Chen et al., 2012). Most of the studies show stronger FRN for negative than positive reward prediction error (Sambrook & Goslin, 2015). However, in some studies, FRN amplitude is greater following neutral feedback than negative feedback (Walsh & Anderson, 2012). An MEG study of social conformity (Zubarev et al., 2017) indicated that electromagnetic brain responses to the disagreement between an individual and group opinion were similar to FRN component with larger negative deflection for conflict trials and originated in anterior and posterior cingulate cortex (PCC). This FRN response is accompanied by activity in TPJ and ventromedial prefrontal cortex (vmPFC). Together, these neural responses tracked the perceived discrepancy between an individual and group opinion.

Besides evoked responses, a number of electrophysiological studies have revealed an increase of oscillatory activity at 200–500 ms, specifically at beta band (13–30 Hz), after unexpected positive reward in gambling, as well as in probabilistic learning context when the outcome is better than expected (Hosseini & Holroyd, 2015;

Marco-Pallares et al., 2008, 2015). Also, in a social context, agreement with the group opinion was followed by an increase in the power of beta-band oscillations in vmPFC and ACC (Zubarev et al., 2017). On the other hand, mid-frontal theta-band activity was modulated by violation of expectations in a reinforcement learning task (Cavanagh et al., 2010; Cohen et al., 2007; van de Vijver et al., 2011) after errors in conflict tasks (Cavanagh et al., 2012), and after losses in gambling tasks (Marco-Pallares et al., 2008). Increases in theta-band oscillations after an incorrect response are suggested as a fundamental process associated with, or even underlying, the error-related negativity and FRN over MFC (van de Vijver et al., 2011). In the social context, conflict with group opinion was followed by an increase in theta power over the anterior and posterior medial cortices (Zubarev et al., 2017). Taken together, a brain network involving PFC, ACC, and TPJ regions manifested beta and theta-band oscillatory activity, as well as FRN-type of evoked activity, are thus likely to underlie the processing of socially meaningful peer feedback and its influence on one's behavior.

The way we interact with each other is highly individual. These individual differences in social interactions reflect behavioral tendencies, which likely build on neurodevelopmental characteristics. Out of temperamental and personality characteristics, agreeableness is most clearly linked with social influence and, together with conscientiousness, associated with social-emotional stability. Agreeableness also positively correlates with social conformity (DeYoung et al., 2002; Kosloff et al., 2017). High agreeableness is positively associated with helpfulness, friendliness, and compliance with the needs of others (DeYoung, 2010) and is negatively associated with aggression and interpersonal disputes (Jensen-Campbell & Graziano, 2001). It is also suggested that agreeableness can predict motivation for maintaining positive relationships, tendency to cooperate, and keeping social harmony in social relations (DeYoung, 2010). In addition to the possible link of agreeableness to conformity behavior, agreeableness is also involved in information sharing behavior. In fact, since people high in agreeableness are often helpful and cooperative with others, they could most probably participate in information sharing behavior in the form of collaboration and getting along with others within interpersonal relationships as employees and students (Matzler et al., 2008) and in the form of sharing ideas regarding products and services in communication among consumers. This knowledge-sharing propensity drives recommendation behavior and has a substantial impact on behavior in consumer context (Ali et al., 2022; Cascio et al., 2015). Contrary to the behavioral effect of agreeableness in

numerous domains, the possible neural correlates of agreeableness have not been addressed systematically. A recent study by Zhang et al. (2020) hypothesized that FRN signals can reflect individual differences in social feedback evaluation so that, *e.g.*, agreeableness would influence an individual's brain response in social influence tasks. However, they did not find evidence to support their hypotheses.

In this study, we explore behavioral effects and the brain basis of opinion discrepancy between the individual and the peer group during a recommendation-based social influence experiment. Previous studies have been mostly limited to the use of hemodynamic-based measures of fMRI, that do not provide direct access to the time-varying neural information processing. So, we measure the immediate neural markers of conflict between individual opinion and group recommendation by using magnetoencephalography. In addition to the event-related field (ERF), which reflects only the signal that is phase-locked to the stimulus and omits the majority of neural oscillatory activities, we addressed spectral differentiation of social conflict with time-frequency analysis. Oscillatory neural activity is considered essential in forming long-range functional networks, which might be crucial for conflict processing in social influence studies.

It has been more than a decade since researchers started moving their focus from merely identifying personality dimensions to the cognitive and neural underpinnings of personality traits. Here, we investigate the behavioral and neural association of agreeableness with social influence. We use a paradigm in which participants are subjected to peer group recommendation after their own initial recommendations of commercial brands. Unlike earlier studies, we tested conformity in the framework of recommendation behavior, because previous research on social influence has mainly examined how others' opinions influence personal opinions or evaluative judgments, that is, "people's attitudes or the overall degree to which they like or dislike any given object or concept" (Briñol et al., 2017). In general, evaluative judgments are not broadly subject to others' assessment, unless it is publicly declared. Instead, when people make recommendation decisions, they also evaluate how they are perceived by others (Barasch, 2020), making the task inherently social. Interestingly, the abundance of interaction in the new media environment, involving also various products and services, may have led to an increased need to be perceived positively. Indeed, there is a consensus that the tendency to self-enhance is a fundamental human motivation (Fiske, 2018), and sharing information with a wide range of real and imagined others through recommendation

can help to maintain reputation and bolster self-concept (Eisingerich et al., 2015).

Social influence has been a focus of extensive psychological research, and social psychology has predominantly studied the behavioral effects of this phenomenon. Hence, in line with former evidence, where aggregated group opinion was shown to influence an individual's decision and behavior, we hypothesized that recommendations from a peer group would change the value that individuals place on their recommendation intentions. This means that perceived positive and negative discrepancies with group recommendation will cause an individual's recommendation intention to change in a more positive or negative direction, respectively, thereby increasing the probability of future behavior change or so-called conformity toward peer group recommendation. Moreover, when there is no difference between the group and individual, the opinion will remain unchanged. Furthermore, based on the aforementioned grounds that the personality dimension of agreeableness may be associated with inter-individual differences in social influence and information sharing behavior, we expected agreeableness to modulate this behavioral adjustment as well as its neural correlates. More specifically, we expected high agreeable individuals to show stronger neural effects to conflicting opinions with the group and, subsequently, higher behavioral conformity compared to low agreeable ones. To test these hypotheses, we selected our subjects based on the agreeableness subscale of the five-factor model (FFM; Konstantel et al., 2012) to test whether high agreeable (HA) and low agreeable (LA) participants will conform to peer opinion in different ways and have distinct neural responses to peer feedback. Moreover, to detect neural processes predicting these behavioral effects and based on earlier neuroimaging findings, we further hypothesized that a discrepancy between an individual's preference and peer group opinions evokes responses similar to the FRN and an increase in the power of theta oscillations. Agreement with peer opinion was expected to induce an increase in beta-band oscillations.

## Materials and methods

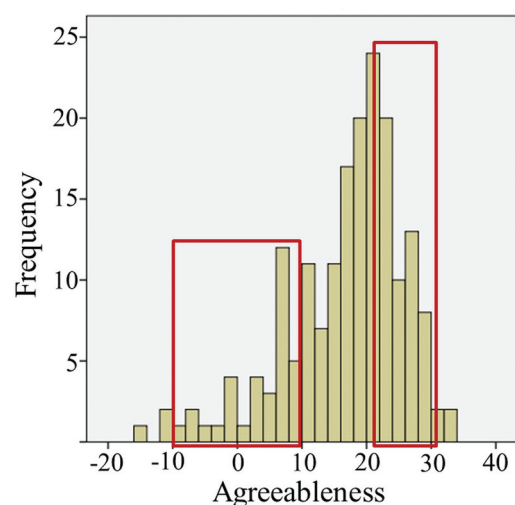
### Participants

One hundred and eighty-two individuals (141 females, age range 15–40 years old) filled the Short Five (S5) test, a short version of the five-factor model (FFM) personality inventory (Konstantel et al., 2012). We used the agreeableness subscale of S5 to select the participants for our MEG experiment. We recruited 15 participants with low

agreeableness (9 females, 6 males, mean age 24.80 years,  $SD = 3.629$ , 14 right-handed) and 15 with high agreeableness (11 females, 4 males, mean age 25.20 years,  $SD = 3.098$ , 14 right-handed). The low agreeableness group (LA) covers the S5 points ranging from  $-10$  to  $10$ , and the high agreeableness group (HA) from  $24$  to  $32$  (Figure 1). None of the participants reported a history of psychiatric or neurological illness, head trauma, or drug abuse. Twenty-five of the subjects were students and five employees. Subjects received a 20 euro grocery gift card as compensation for participation in the experiment. The ethical committee of the University of Jyväskylä approved the study, and all participants signed the informed consent form.

### Stimuli and task

The stimuli were national and global brands known to the Finnish population. These brands were categorized into five categories (food, clothing, home, technology, and personal care). A separate group of 20 individuals was recruited, to evaluate the familiarity of the commercial brands on a 5-point scale (1 = very unfamiliar; 5 = very familiar). This session was conducted in order to control the effect of brand familiarity on social influence and recommending behavior, and 210 brands in two categories of familiar and unfamiliar were used in our MEG experiment. The subjects in the MEG experiment evaluated the familiarity of the selected brands one day before the experiment to confirm that their evaluation



**Figure 1.** Distribution of the S5 trait “Agreeableness” from our online survey ( $n = 182$ ). Red rectangles refer to participants selected for the MEG study. The low agreeableness group (LA,  $n = 15$ ) had agreeableness values from  $-10$  to  $10$  and the high agreeableness group (HA,  $n = 15$ ) from  $24$  to  $32$ .

aligned with the familiarity evaluation of the independent test group.

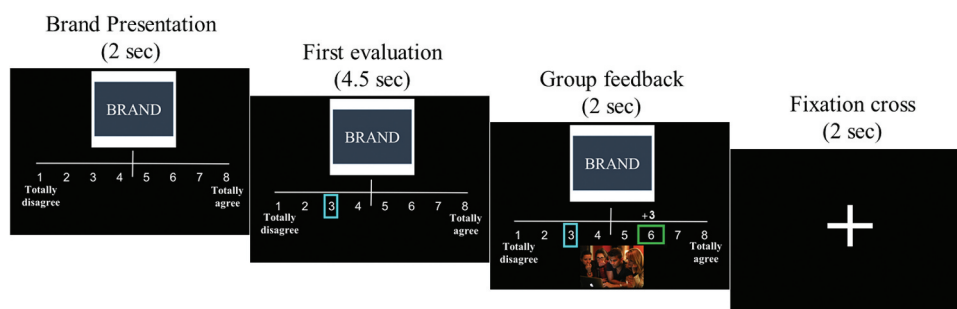
Before the start of the experiment, subjects were explained that they are participating in an experiment about how people make decisions between different brands and how personality affects these decisions and that they are informed about the average rating of 300 students from Jyväskylä university. Therefore, they were unaware of the real aim of the experiment. During the MEG measurement, subjects were first presented with the 210 brands (first rating session, session duration 45 min). Subjects were instructed to rate each brand on an 8-point scale, based on the claim ‘I would recommend this brand’ (1 = totally disagree; 8 = totally agree). A nonmagnetic, four-button device (Cambridge Research Systems, Ltd., UK) was used for selecting the rating by stepwise increasing or decreasing the value. At the beginning of the experiment, subjects rated eight practice trials with unrelated stimuli to get familiar with the use of the response buttons and the task. The structure and timing of the trials are illustrated in Figure 2. In each experimental trial, a brand was presented on the screen, together with a scale of 1 to 8. After 2 s, subjects were given 4.5 s to respond. In case the participant did not press a button within 4.5 s after the brand presentation, the trial ended, and the text “Too late” appeared on the screen. The subject’s rating was highlighted on the screen with a blue rectangular frame immediately after pressing the button. After a random 0.5–2 s delay, the participant was informed of the average recommendation of 300 students as peer group feedback. This was shown with a green rectangle for 2 s, together with a cue denoting the peer group (photograph of peers, Figure 2). The difference between the responses of the subject and the feedback group was shown above the green rectangle (0,  $\pm 2$ , or  $\pm 3$  points). In reality, the average feedback group

rating was randomized to either be the same (no-conflict condition) or 2–3 points below or above the subject’s rating (negative or positive *conflict* condition). This was done to manipulate the experience of social conflict between the subject and the peer group. The conditions were balanced individually among the 210 trials (1/3 conflict, 1/3 positive conflict, and 1/3 negative conflict), but this balance could have changed slightly depending on the subject’s initial ratings, although the subjects were instructed to use the whole scale (1 to 8). Stimulus presentation was controlled with the Presentation software (Neurobehavioral Systems, Inc., Albany, CA, USA).

Thirty minutes after the MEG measurement, the subjects were asked to rate the same items again, but without group feedback and in a new randomized order (second rating session, session duration 30 min). The second session was used to test whether subjects changed their initial recommendations after facing differing peer opinions, as predicted by social influence theory. Importantly, the subjects were not informed about this second session beforehand. At the end of the experiment, the subjects were interviewed and debriefed about the experiment, and the true nature of the experiment was revealed. Importantly, none of them reported guessing that the aim of the study was about social influence or conformity.

### MEG and MRI data acquisition

MEG data were collected with a whole-head 306-channel (102 magnetometer channels and 204 planar gradiometer channels) Elekta TRIUX MEG device (MEGIN Oy, Helsinki, Finland) located in a magnetically shielded room at the Center for Interdisciplinary Brain Research, Dept. Psychology, University of Jyväskylä, Finland, with a 1000 Hz sampling rate and 0.1–330 Hz



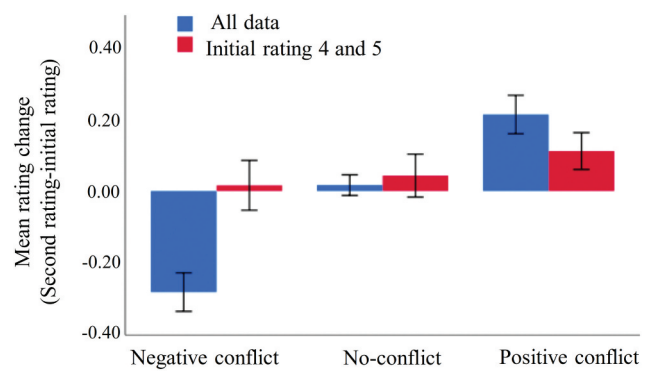
**Figure 2.** Schematic representation of one experimental trial of the first rating session in MEG, where subjects evaluated brands in terms of their intention to recommend them. The subjects were first presented with a brand logo (actual brands were presented instead of the word “brand”). After two seconds of initial display, they had 4.5 seconds to give their recommendation using response keys. Then, the peer evaluation, together with a cue image representing the peer group, was shown. The evaluation could be the same (no-conflict condition) or above or below the peer rating (conflict condition). Thirty minutes after the MEG session, subjects evaluated the same set of brands again in a behavioral session, where the trials were similar but without the group feedback phase.

band-pass acquisition filter. Prior to data acquisition, participants were checked for magnetic interference and instructed to keep their heads still as much as possible. Visual stimuli with the size of 19 cm × 25 cm were presented with a DLP projector on the center of the screen situated in front of the subject with a distance of 100 cm. Concurrently, an electrooculogram (EOG) was recorded using electrodes above the right and below the left eye for blink and saccade artifacts and electrocardiogram (ECG) with electrodes on the left and right clavicle to help detect cardiac artifacts. One ground electrode was attached to the right zygomatic bone. The head position was recorded with five head-position indicator (HPI) coils, three at the forehead and one behind each ear. HPI coil positions were determined with reference to three anatomical landmarks: nasion and left and right preauricular points, which define the head coordinate system. Additional scalp points were digitized for better co-registration of MEG with an individual's structural magnetic resonance images (MRI). All locations were digitized using a Polhemus Isotrak 3D tracker system (Polhemus, Colchester, VT, United States). The MR images were acquired from a private company (Synlab Oy, Jyväskylä). Two sets, a 3D SE T1 weighted and an FSE T2 weighted, were collected on a GE 1.5 T Signa HDxt MRI scanner using a standard head coil.

## Data analysis

### Behavioral data

To find out whether there is a change in participants' opinions after being exposed to the peer group feedback, we categorized behavioral responses based on the direction of the group's opinion compared to the subject's opinion in negative, positive, and no-conflict for each brand (Figure 3). We assessed the opinion change using linear two-level modeling with random slopes. The model's parameters were estimated with Bayesian estimation, as implemented in Mplus v8.2 (Muthén & Muthén, 2017). At the within-subject level, we assessed whether an individual's opinion was changed to the direction of the peer feedback between the first and the second rating (Figure 4(a)). Behavioral categories were transformed into dummy variables, which were used to estimate random slopes in the two-level model (Figure 4(b)). We contrasted positive conflict and negative conflict with the no-conflict condition (random slopes S1 and S2, respectively). The initial rating was added as a covariate (S3) to control for the effect of regression to mean (RTM) due to repeated measurements (Yu & Chen, 2014). At the between-subject level,



**Figure 3.** Mean behavioral change of the brand recommendation between the first and second session. Blue color shows mean change using all trials across all 30 subjects (total number of trials: 6278) and red color shows mean change when the effect of the regression to mean (RtM) is removed by using only initial ratings 4 and 5 (2300 trials). Bars indicate the standard error of the mean across subjects.

we inspected the effect of agreeableness on opinion change and the interaction between agreeableness and opinion change (Figure 4(b)). This analysis was carried out by adding agreeableness as a between-subjects level predictor in the two-level model.

### Processing of MRI data

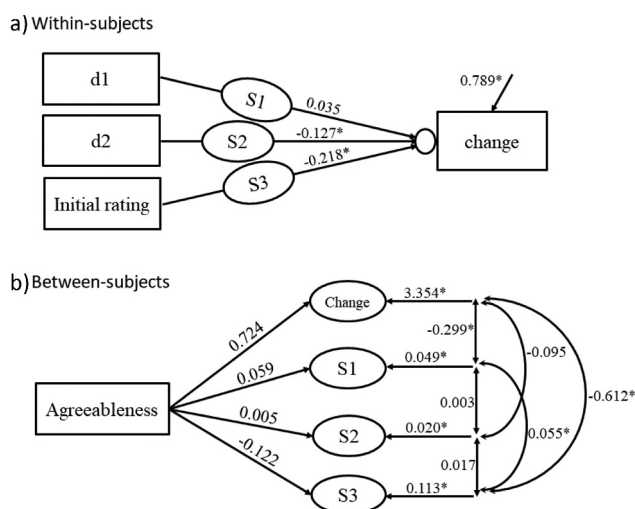
To localize source-level activity in the individual anatomy, we co-registered the anatomical MRIs with the MEG data. This was achieved by identifying in the MRI the anatomical landmarks that were used to determine the head coordinate frame during the MEG preparation (left/right preauricular point and nasion). Co-registration was confirmed and fine-tuned using points digitized on the scalp. The segmented MR images were used to reconstruct the cortical surface, skull, and skin (Dale et al., 1999). A 3-layer (inner skull, outer skull, and outer skin) boundary element model (BEM) was used in forward modeling, together with 5120 source dipoles distributed over the cortical surface.

The MR images were also used to find spherical morphing from individual anatomies to a single target anatomy, the "fsaverage" subject, using FreeSurfer (Fischl et al., 1999). This way we could compare the neural responses between subjects with high anatomical relevance.

### MEG data processing

First, a temporally extended signal space separation method (tSSS) from Maxfilter software (MEGIN, Helsinki, Finland) was used to remove external magnetic interference from MEG data and to correct for head movements, recorded with continuous HPI (Taulu & Hari, 2009). Head position was estimated in 200 ms time





**Figure 4.** Diagrams for behavioral analysis at within-subject level (A) and between-subject level (B). (A) d1 and d2 are dummy variables for the contrast between positive and negative conflict with the no-conflict condition, respectively. S1 and S2 indicate associated random slopes in the two-level model, and S3 is the random slope for modeling the effect of the initial rating. The values show mean differences in within-subject diagram for S1 and S2 and mean changes for S3. The asterisk indicates significant effects at 95% confidence level. (B) the between-subjects effect of agreeableness on rating change (second session rating minus first session rating), and the interaction between agreeableness and slopes S1-S3. The values on left show the regression coefficient of agreeableness on rating change and the interaction between agreeableness and slopes S1-S3. The values on the right indicate residual variances of change and slopes (horizontal arrows), and residual covariances between slopes and change.

windows with 10 ms step for movement compensation and transformed to the mean head position across the MEG session. Independent component analysis (Fast ICA; Hyvärinen & Oja, 2000) was applied to raw MEG data (low-pass filtered at 40 Hz and high pass filtered at 0.5 Hz) to manually identify and remove signal artifacts corresponding to horizontal saccades, blinks, and cardiac activity. In this analysis, the subsequent MEG data processing steps and forward modeling were performed in MNE-Python, v0.19 (Gramfort et al., 2013) using custom scripts. After ICA, the signals were downsampled by 3 (to 333.33 Hz) to reduce data size for further analysis. Since the experimental conditions defined by peer ratings were the main variable in our experiment, we segmented the continuous data to epochs time-locked to the presentation of the peer group feedback (Figure 2) to investigate event-related neural responses. We grouped the epochs into positive conflict, negative conflict, and no-conflict trials, depending on whether the group's ratings were higher, lower, or equal to the participant's rating, respectively. An epoch was rejected if any

magnetometer channel exceeded  $4e-12$  T or any gradiometer channel  $400e-12$  T/m. The trigger-to-stimulus delay, measured using a photosensitive resistor, was subtracted from each epoch.

### Sensor space event-related field (ERF) analysis

For the ERF analysis, the continuous MEG data was low-pass filtered at 40 Hz and high-pass filtered at 0.5 Hz using a zero-phase FIR filter. Epochs were selected from 200 ms before to 1000 ms after the onset of peer feedback presentation. Offsets were removed for each trial by subtracting the mean of the pre-stimulus interval from  $-200$  to  $0$  ms. Evoked fields were estimated by averaging across epochs within the conflict condition (negative and positive conflicts) and the no-conflict condition.

### Sensor-Space time-frequency response (TFR) analysis

For the frequency domain analysis, epochs from 1000 ms pre-stimulus to 2000 ms post-stimulus were first extracted from unfiltered raw data. Offsets were removed from each epoch by aligning the time series to the average of a 1000 ms pre-stimulus interval. The average evoked response was subtracted from each epoch to remove phase-locked activity. For time-frequency decomposition, we used Morlet wavelets, where the number of cycles was set to half of the center frequency. The frequency band of interest ranged from 4 Hz up to 60 Hz, with wavelet center frequencies in steps of 2 Hz. Each epoch was convoluted with the complex wavelet and then the absolute value was averaged across the epochs to obtain the amplitude. Then, the epochs were downsampled by 2 and trimmed by 500 ms at each end to avoid edge effects. Finally, the amplitude responses were converted to z-scores by subtracting the mean and dividing by the standard deviation, in order to reduce the impact of between-subjects variability in neural oscillation amplitudes.

### Statistical analysis

To determine whether there is a difference in neural responses between the conflict and no-conflict conditions and to identify the relevant time windows, we performed a non-parametric permutation test with a clustering method to correct for multiple comparisons (Maris & Oostenveld, 2007). This was done for the time window between 0 to 1000 ms for the ERF analysis and between 0 and 1500 ms for the TFR analysis. For each sample (time point for ERF or time-frequency point for

TFR), the difference between conflict and no-conflict was expressed as a dependent sample's  $t$ -statistic. Samples for which these  $t$ -statistics exceeded an uncorrected threshold of  $\alpha = 0.05$  were clustered based on spatial, temporal, and spectral adjacency. Cluster-level test statistics were calculated by summing the  $t$ -statistics of the samples belonging to the same cluster. The largest cluster-level  $t$ -statistic was used as a test statistic as suggested by (among others) Maris and Oostenveld (2007). Next, a permutation distribution of cluster-level statistics was calculated by randomly exchanging condition labels between epochs and calculating positive and negative cluster-level statistics for every permutation, for a total of 5000 permutations. The observed cluster-level statistic was tested against the surrogate distribution to find the permutation-based  $p$ -value.

### Source reconstruction

Source reconstruction was carried out to identify the brain areas underpinning the experimental effects detected at the sensor level, both for the event-related fields and the time-frequency data. Different source reconstruction approaches were implemented for each kind of data: linearly constrained minimum variance beamformer (LCMV) (Van Veen et al., 1997) for the ERFs, and a frequency-domain beamformer (DICS) (Gross et al., 2001) for the TFR data. For LCMV, we estimated the noise covariance matrix from a 500 ms pre-stimulus window using cross-validated Ledoit-Wolf estimator (Ledoit & Wolf, 2004) (as implemented with the "shrunk" method in MNE-Python). The data covariance was calculated across a time window of 100 ms to 600 ms relative to group feedback onset across both (conflict and no-conflict) conditions using the same method. We used a regularization parameter of 0.05 and depth weighting of 0.8. The rank of data was defined based on the degrees of freedom in the SSS transformation, subtracted from the number of removed ICA components. We computed the neural activity index (NAI; Hymers et al., 2010) using source orientation based on maximum power. Spatial filters were calculated using all epochs. Next, for each condition separately, the spatial filters were used to estimate source activity corresponding to the ERF, and the mean baseline activity was subtracted.

For the analysis of sources of the induced responses with DICS, the evoked response was first removed from each epoch and baseline correction was done based on a 1000 ms pre-stimulus window. The cross-spectral density (CSD) across epochs and between all pairs of MEG gradiometers was estimated using Morlet wavelets, like

for TFR, for computing the DICS filter weights. The active window was from 200 ms to 1200 ms, and the baseline window was  $-500$  ms to 0 ms. The CSDs for each wavelet center frequency were averaged within the frequency band of interest. The resulting CSDs were combined with the forward solution to calculate frequency-band specific spatial filters for each source location. For each epoch, the resulting DICS estimates were Hilbert transformed and the absolute values were averaged across epochs to provide condition-specific source-level oscillation amplitude responses.

## Results

### Behavioral results

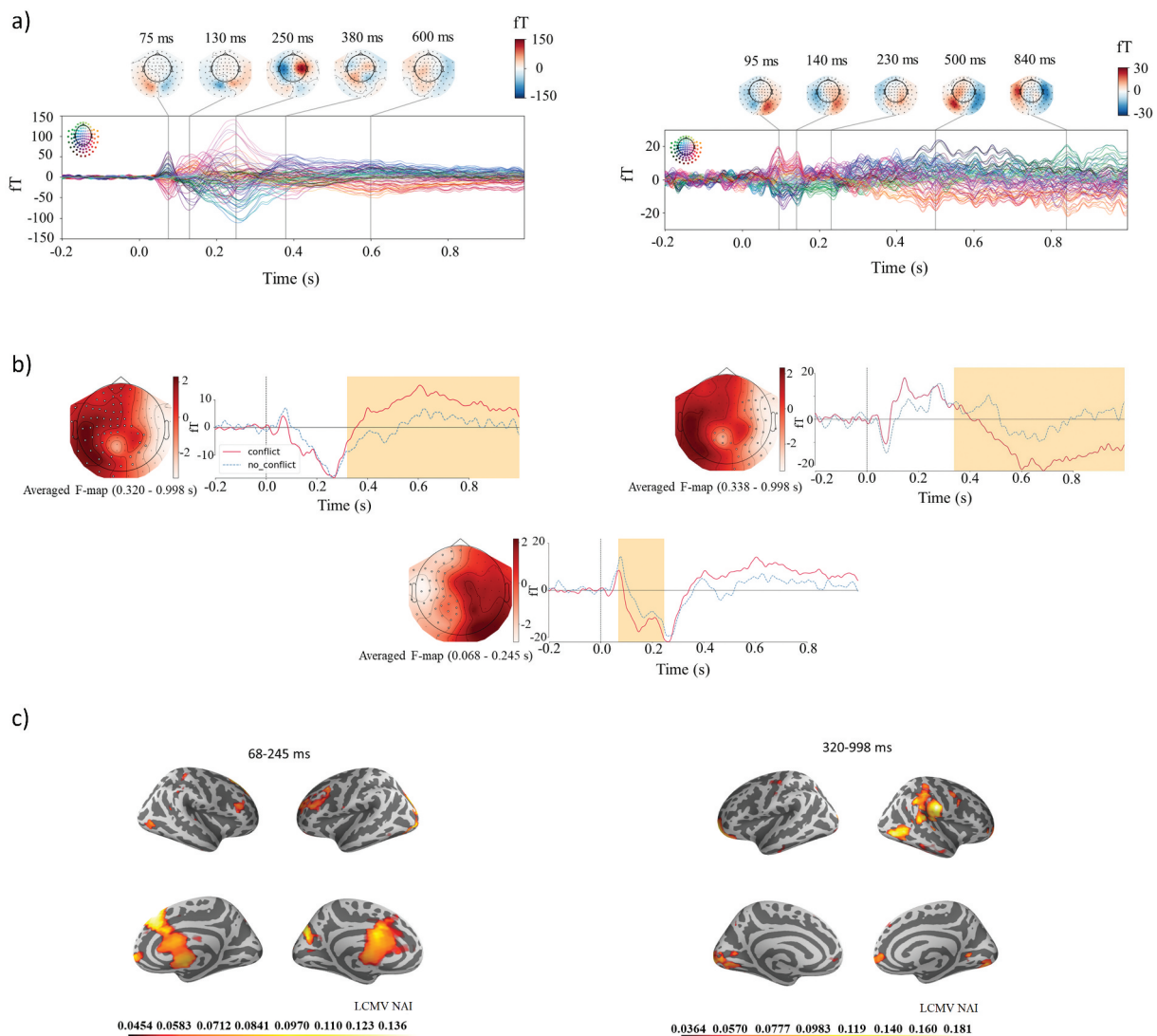
After categorizing trials based on the direction of the group's opinion (negative, positive, and no-conflict; see Figure 3 for behavioral effect), we tested whether participants' opinions changed toward the group opinion by comparing the 1st and 2nd ratings. In our multilevel model, the difference between negative and no-conflict feedback (the second random slope, S2; Figure 4(a)) was significant (mean difference =  $-0.127$ , confidence interval CI =  $[-0.204, -0.052]$ , observed post-hoc power for mean = 0.962), but no significant difference between positive and no-conflict feedback (random slope S1) was observed (mean difference = 0.035, CI =  $[-0.064, 0.134]$ , observed post-hoc power for mean = 0.136). To determine if not observing a significant effect for random slope S1 was due to insufficient power of our study, we performed sensitivity analysis, in addition to post-hoc power, to compute the required mean difference between positive and no-conflict feedback using Monte Carlo simulation. We found out that a mean difference of 0.13 would be required to get power of 0.8. The influence of the initial rating (first session) on subsequent rating change due to RtM was used as a covariate (S3). The initial rating was found to have an effect on rating change (mean =  $-0.218$ , CI =  $[-0.344, -0.093]$ , observed power for mean = 0.947), in line with previous similar studies (Levorsen et al., 2020; Nook & Zaki, 2015).

At the between-subject level (Figure 4(b)), agreeableness did not affect rating change (regression coefficient  $\beta = 0.724$ , CI =  $[-0.654, 2.108]$ ). The interaction between agreeableness and S1, S2, and S3 on rating change was not significant. Residual variances from the agreeableness main effect and interaction effects between agreeableness and slopes on opinion change and residual covariances between slopes and rating change are shown in Figure 4(b).

## Evoked responses

The butterfly plots of the evoked response, time-locked to the presentation of group ratings, in all magnetometer sensors are shown in Figure 5(a) for grand average (negative conflict plus positive conflict plus no-conflict; left panel) and condition contrast (conflict minus no-conflict; right panel). The spatial topography of the magnetic fields is given for the local maxima along the time course. A clear response peaking at 75 and 130 ms is visible, with the main contribution from the occipital areas, and later maxima can be seen in the frontal areas around 200 to 300 ms after

observing conflict and no-conflict feedback from the group. Statistical analysis of magnetometer data identified two spatiotemporal clusters, where the evoked activity in conflict trials differed from the activity in no-conflict trials (Figure 5(b)). A positive cluster in most of the left hemisphere sensors and a negative cluster broadly in the right hemisphere sensors appeared starting at 320 ms after the peer group rating onset, indicating a greater amplitude for conflict trials compared to no-conflict trials ( $p < 0.001$ ). Another cluster was found at 68–245 ms, indicating a greater amplitude during conflict trials compared to no-



**Figure 5.** Evoked response analysis. (A) Butterfly plots of magnetometer channels for evoked responses time-locked to the presentation of group feedback for grand average (left panel) and condition contrast (conflict minus no-conflict; right panel). Colored heads in the top-left corner show the individual waveforms' position in the sensor array and color scales in top-right indicate magnetic field strength. The time points for topographies are selected based on the peak activations (B) Cluster-based permutation test results. Time courses were obtained by averaging over magnetometers comprising of the two clusters identified by the permutation test. Orange boxes indicate the time windows in which statistically significant differences were observed. (C) Source reconstructions of the condition contrast (conflict minus no-conflict) in the time windows (68–245, 320–998 ms) suggested by sensor space cluster-based permutation test.

conflict trials ( $p < 0.029$ ; Figure 5(b)). Analysis of the gradiometer channels identified two spatiotemporal clusters in similar time windows to the clusters as in magnetometer analysis with greater amplitude for conflict trials.

Source reconstruction with LCMV beamforming was used to identify the brain regions underlying the effects identified in sensor data in the two time windows, (68–245 ms and 320–998 ms; Figure 5(c)). Source analysis of condition contrast showed differing activation between conflict and no-conflict trials mainly in medial and lateral frontal areas, including lateral prefrontal cortex (IPFC), mPFC, ACC, and cuneus in the early time window and in precentral, postcentral, and supramarginal cortices in the later time window.

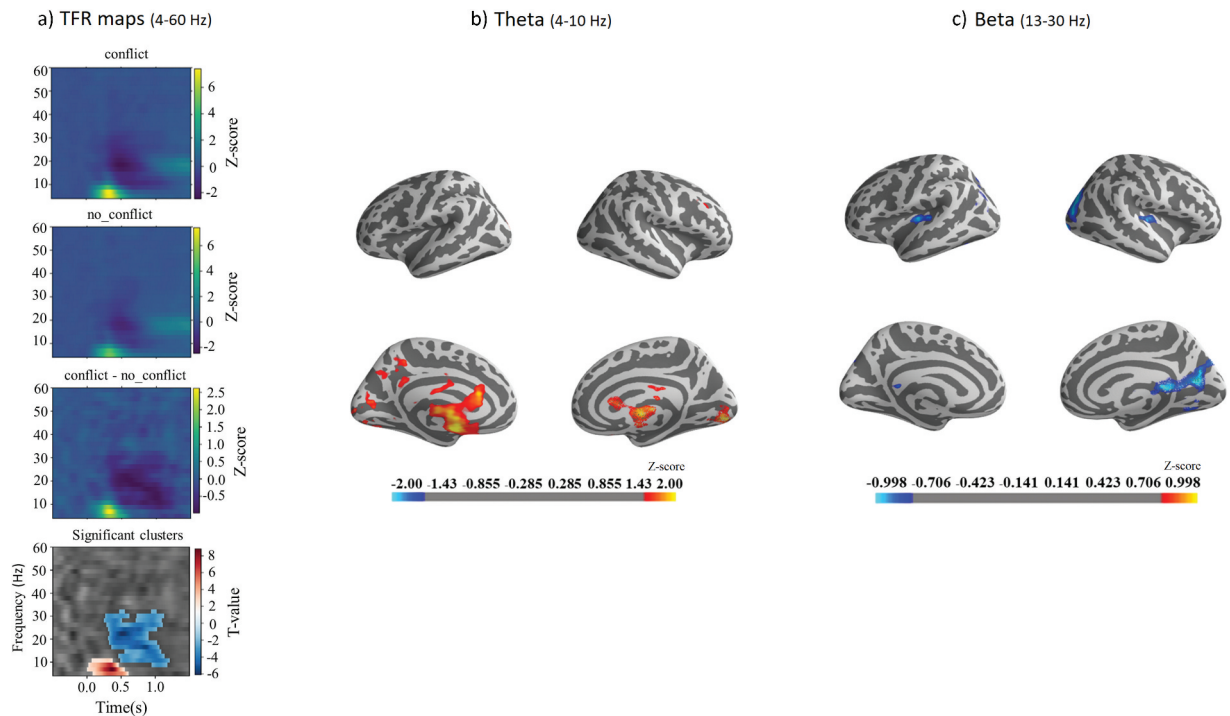
Cluster-based permutation tests on ERFs between the LA and HA groups did not reveal any significant group differences in sensor or source space, neither in the contrast condition (conflict vs. no-conflict) nor in any individual condition.

### Time-frequency analysis

In the TFR analysis, the focus was on induced oscillatory activity in the 306 channels (Figure 6(a)), with a time window –500 ms to 1500 ms and a frequency range

from 4 Hz to 60 Hz. Cluster-based permutation test for condition contrast (conflict minus no-conflict) revealed two significant clusters. The first cluster appeared between 10 and 590 ms time interval at 4 to 10 Hz frequency band, with a larger increase in induced amplitude for the conflict condition compared to the no-conflict condition (positive  $t$ -values,  $p < 0.001$ ). The second cluster appeared between 250 and 1200 ms time interval in 13–30 Hz frequency band, with a stronger decrease in induced amplitude for the conflict condition, compared to the no-conflict condition (negative  $t$ -values,  $p < 0.001$ ). Again, we constrained our source localization to the two time-frequency windows; 10–590 ms at 4 to 10 Hz (Figure 6(b)) and 250–1200 ms at 13–30 Hz. We employed DICS beamforming to estimate the sources of the rhythmic activities (Figure 6(c)). Sources of condition contrast (conflict minus no-conflict) in the theta frequency range mainly spread over the left and right medial surfaces, including parts of cuneus, precuneus, lingual and cingulate gyri. In the beta frequency range, sources were identified in the lateral occipital and parts of superior temporal and the lingual gyrus and cuneus.

The between-group differences (LA vs. HA) in induced oscillations analyzed for both conflict and no-conflict conditions (Figure 7(a)) revealed a significant difference



**Figure 6.** Analysis of induced oscillatory activity. (A) TFR maps for the frequency range 4–60 Hz and the time interval –500–1500 ms for conflict, no-conflict and the condition contrast, and the significant clusters resulting from the cluster-based permutation test. (B) Source reconstruction of the TFR effect (conflict minus no-conflict) in the significant time-frequency window of 10–590 ms at 4–10 Hz. (C) Source reconstruction of the TFR effect (conflict minus no-conflict) in the significant time-frequency window of 250–1200 ms at 13–30 Hz.

at sensor level in a late time window from 800 to 1500 ms in the beta frequency range (10–28 Hz) (positive  $t$ -value,  $p < 0.04$ ). To localize the cortical contributors of this difference, source reconstruction using DICS beamforming was applied in the time-frequency window limited to 800–1500 ms and 10–28 Hz (Figure 7(b)). Source plots of group contrast in the no-conflict condition in the beta range showed a difference in parts of lateral superior and inferior parietal and occipital cortices, along with precuneus and cuneus in medial cortices.

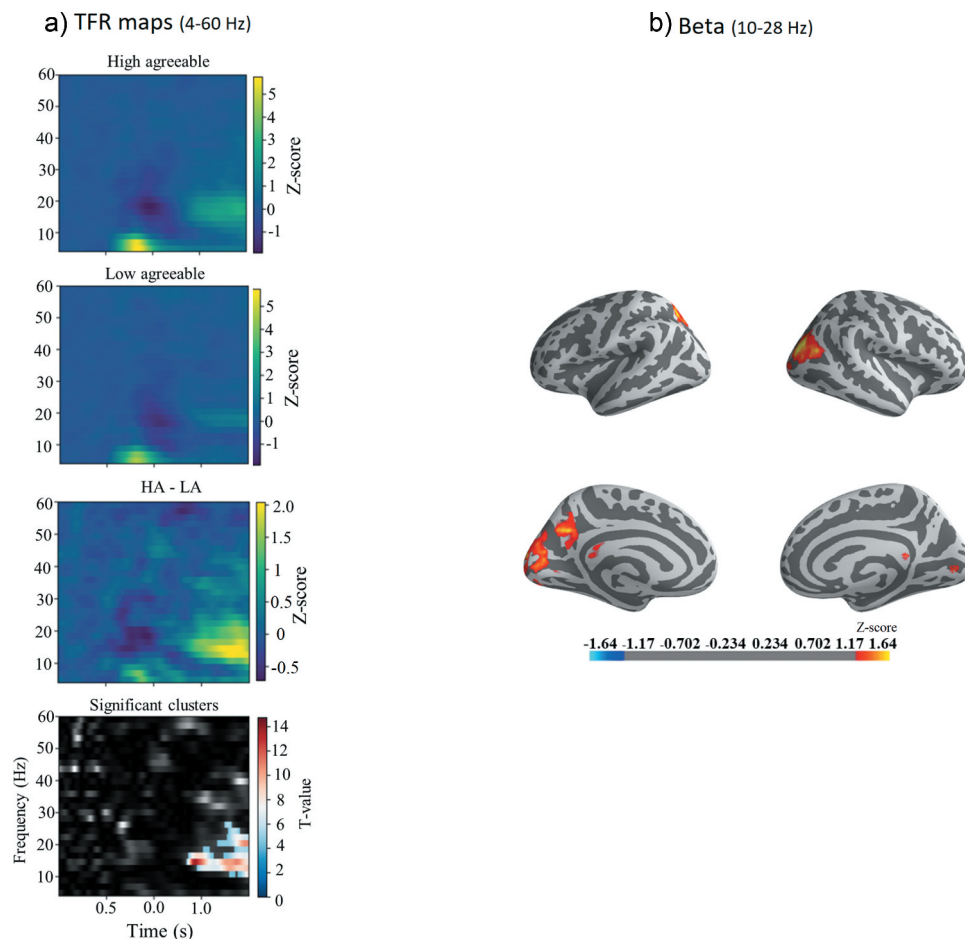
## Discussion

The objective of the present study was to examine the behavioral manifestation and the underlying neural dynamics involved in processing conflict between recommendation intentions of individuals and their peer group. Earlier studies on social influence have either focused on evaluative judgments (hence, not on recommendation) or used imaging modalities that do

not offer high enough temporal resolution to follow the fast neural dynamics of conflict processing. Here, we extended the field by using a temporally accurate imaging method to explore the immediate neural markers, that is, evoked and induced MEG responses, associated specifically with the recommendation-based task. Moreover, we also elucidated the behavioral and neural effects of individual differences in peer conflict processing by selecting our subjects based on agreeableness, as it is considered to be tightly linked with social influence and information sharing behavior.

### Behavioral conformity to group opinion

After correcting for the effect of regression to mean due to the initial rating on opinion change, the behavioral analysis revealed that only the conflict to the negative direction significantly modulated an individual's opinion. In other words, individuals decreased their preference after being exposed to lower ratings by the peer



**Figure 7.** Group-wise analysis of induced oscillatory activity (A) TFR maps for the low and high agreeable groups in the no-conflict condition; from top to bottom: high agreeable, low agreeable, high minus low and result from cluster-based permutation test. Frequency range was 4–60 Hz and the time interval was –500–1500 ms. (B) Source reconstructions of groups effect (high minus low) from significant time-frequency windows in 10–28 Hz at 800–1500 ms.

group, but they did not robustly increase their preference after being exposed to higher ratings. This result is in line with previous findings, where conflict in the negative direction had a stronger effect than that in the positive (Klucharev et al., 2009; Shestakova et al., 2013; Templeton et al., 2016; Zaki et al., 2011; Zubarev et al., 2017). Moreover, in previous recommendation-based studies of social influence (Baek et al., 2021; Cascio et al., 2015), behavioral evidence also showed that negative peer recommendations lead to greater conformity than positive recommendations. The specific conformity effect of our experimental paradigm, and using recommendation task instead of evaluative judgments, might highlight that especially from the perspective of recommending a brand, negatively framed peer opinion may be considered more informative or diagnostic than positively framed peer opinion. In other words, negativity may propagate more strongly in contexts related to consumer behavior when the financial outcome is more salient than, for example, rating facial attractiveness. Considering knowledge-sharing behavior, people share information in the desire to maintain a positive self-concept, and carefully selected recommendations can make them appear knowledgeable and trustworthy (Packard & Wooten, 2013). Finally, people tend to share positive product experiences of their own but negative product experiences of other people (Packard & Berger, 2017).

### **Brain evoked responses to conflicting group opinion**

At the brain level, perceived conflict with the group increased the evoked activation in early (68–245 ms) and late (320–998 ms) time windows. The medial surfaces including ACC, mPFC, cuneus, and precuneus as well as dorsolateral prefrontal cortex (dlPFC) were involved in processing the social feedback. Our results only partly complied with earlier M/EEG studies of social influence, where reinforcement learning was suggested as a mechanism of group feedback processing and subsequent behavior adjustment, evoking FRN component as characteristic of reinforcement learning in M/EEG studies (Chen et al., 2012; Kim et al., 2012; Shestakova et al., 2013; Zubarev et al., 2017). In our study, the first effect in the evoked response was somewhat earlier than what has been found in former studies. However, in line with the reported localization of the electromagnetic sources underlying FRN (Zubarev et al., 2017), our source analysis showed more activity in ACC (among other regions) in conflict trials compared to no-conflict trials.

In many social conformity studies, activity in ACC has been considered to reflect action monitoring and reinforcement learning (RL), but it is only one account for

conformity. In fact, the engagement of ACC alongside AI is aligned with earlier findings on social exclusion and rejection referred to as ‘social pain’. Some researchers have suggested that conformity to group opinions could be the product of negative affective states of social rejection threat, reflected in ACC and AI activity after a disagreement with the group (Berns et al., 2010; Wasylyshyn et al., 2018). Furthermore, the activity of ACC accompanied by dlPFC might also indicate negative emotion of pressure to change the initial preference in the face of new information, known as ‘cognitive dissonance’ (Izuma et al., 2010; Van Veen et al., 2009). Although in the context of social influence, others’ opinions provide useful information, there is a cost that an individual pays for altering their beliefs to gain social approval. This cost manifests itself in the form of psychological discomfort, specifically cognitive dissonance. In other words, inconsistencies in one’s knowledge or opinion about herself and her environment can create uncomfortable feelings as experienced during cognitive dissonance and thus strong motivation to retrieve an acceptable state (Festinger, 1957). Mainly, opinion updating is a tradeoff between conformity and cognitive dissonance. In other words, one must conform to the group to get an advantage, while keeping cognitive dissonance to a minimum by not distorting their expressed opinion too far from their real opinion (Seeme et al., 2019). Therefore, the activity of dACC and dlPFC in social conformity tasks may well reflect this negative emotion or psychological discomfort of two contrary beliefs when experiencing conflict with the group (Levorsen et al., 2020). The early time window of the effects observed in our study seems to be compatible with this interpretation, as cognitive dissonance due to difficult choices evoked a strong negative fronto-central response that peaked at 60 ms in a free-choice task (Colosio et al., 2017).

Social pain and cognitive dissonance in relation to social influence are particularly important in light of a recent study by Levorsen et al. (2020), who directly compared neural responses using fMRI between social conformity and reinforcement learning and did not find clear evidence for a common neural mechanism between these two processes. This finding undermines the reinforcement learning hypothesis of social conformity put forth by Klucharev et al. (2009), which suggests that conflict with social norms triggers a prediction error similar to the reinforcement learning signal, calling for adjustment of behavior. Levorsen and colleagues, however, argued that although normative conformity with subjective rating of stimuli does not show common neural resources with RL, informational conformity, where social conflict can serve as a teaching signal, might share neural mechanisms with RL.

According to Huang et al. (2014), N400-like modulations of neural activity in their study reflect a true conflict between an individual and group rather than reward prediction error as in FRN around 200 ms. Moreover, N400 encodes bi-directional conflict meaning that N400 is more negative for group's positive feedback than negative feedback but individual differences in conformity behavior were not associated with these differences in N400 response.

Previous findings (Baek et al., 2021; Cascio et al., 2015) in the recommendation-based conformity task showed involvement of valuation and mentalizing systems in processing conflicting messages from peers and subsequent opinion change. Moreover, Baek et al. (2021) found a marginally significant interaction between the sentiment of the recommendation and activity in the brain's mentalizing system to predict recommendation change. These findings suggest that the brain's mentalizing system may be recruited more strongly in situations where social consequences are the most salient, such as those that may signal negative outcomes. In fact, value-related brain signals seemed to track the value of the peer recommendation regardless of valence, while the activation in the mentalizing system was more responsive under some conditions than others, and negative reviews were shown to be more salient in this regard (Baek et al., 2021).

### **Neural oscillatory dynamics after conflict with the group**

Our analysis of induced oscillatory activity provided evidence toward the neural basis of social influence, in line with similar earlier EEG and MEG studies (Cavanagh et al., 2010; Cohen et al., 2007; Kim et al., 2012; Zubarev et al., 2017). The results from former studies imply that increases in theta oscillations over medial frontal cortices in processing negative feedback or unfavorable outcome in feedback-based response learning (van de Vijver et al., 2011). The increased theta in conflict trials compared to no-conflict trials in our study localized mainly in the medial surfaces, consistent with Zubarev et al. (2017) findings, where the difference between individual opinion and group's opinion in the judgment of trustworthiness evoked an increase in theta band oscillations in medial frontal regions. Theta-band oscillatory perturbation occurs with a similar topography and time range as the FRN response to punishment or negative feedback. In other words, time-frequency decomposition of the FRN reveals theta (4–7 Hz) activity in medial frontal electrodes and is thought to originate from the anterior cingulate cortex (ACC). So, theta-band oscillations have been suggested to underlie FRN (Cavanagh

et al., 2010). In our study, modulation of theta oscillation occurred in time windows as early as our evoked response with overlapping sources. However, it should be mentioned that precise determination of temporal, spatial, and frequency domain limits is not supported by the cluster-based permutation test (Sassenhagen & Draschkow, 2019) applied in the present study.

Following the increase in the low-frequency band, we observed a decrease in the induced amplitude of beta-band (13–30 Hz) oscillations as response to the conflict with the group preference, compared to an agreement with the group. This effect took place in a time window of 250–1200 ms and was localized widely to the medial and lateral surfaces. In previous studies, an increase of oscillatory activity in beta-band (12–30 Hz) was observed after the delivery of rewards in gambling and learning tasks (Marco-Pallarés et al., 2015; van de Vijver et al., 2011). Furthermore, in (Zubarev et al., 2017) study, when an individual's opinion was the same as the group opinion, an increase in oscillatory activity in the beta band, localized in the VMPFC, was interpreted to reflect reward processing related to subjective pleasure of being similar to the group. Engel and Fries (2010) related an increase in beta-band oscillations to maintenance of the current sensorimotor or cognitive state, consistent with reward signaling. In our study, the beta-band suppression in conflict trials can be interpreted as a call for a status change alongside theta increase, signifying error detection or displeasure.

### **The effect of agreeableness on processing group feedback**

In contrast to our hypotheses, agreeableness did not mediate behavior adjustment and social conformity in our study. Moreover, it did not influence conflict processing neither in evoked nor in induced brain responses. However, a difference between the low and high agreeable groups was observed in the time-frequency domain (800–1500 ms, 10–28 Hz) in the no-conflict condition, showing higher induced amplitude for the high agreeable group. The effect was observed mainly in lateral and medial occipital cortices. As mentioned earlier, an increase in beta-band oscillations has been associated with the agreement with group opinion in social conformity tasks (Zubarev et al., 2017), reward delivery, and winning in gambling tasks (Marco-Pallares et al., 2008; Marco-Pallarés et al., 2015). Interpreted in the context of present results, the match between an individual's and group's opinion (*i.e.* no-conflict trials) serves as a reward, which is reflected as a stronger increase in beta oscillations in individuals with higher agreeableness tendency. Matching this interpretation, Wang et al. (2019)

interestingly showed that highly agreeable participants were less affected by the negative feedback from their peers, can better tolerate someone opposing them, and can regulate their behavior in a more socially acceptable manner. Consequently, agreeable individuals seem likely to focus less on conflict-related signals or interpersonal disputes and more on cooperation and positive aspects (DeYoung, 2010). The lack of interaction between agreeableness and feedback processing in our evoked response data is also in line with earlier studies. In a recent study by Zhang et al. (2020), no significant effect of agreeableness was found on FRN or P300 response. Indeed, in line with our current results, the influence of the majority's opinion on individuals was not moderated by agreeableness, and individuals with higher agreeableness did not show greater FRN nor smaller P300 when their behaviors were inconsistent with the majority group.

Former psychological studies on agreeableness showed that agreeable individuals have social desirability to be in agreement with others, but this desire is not equal to being easily influenced by others (Jensen-Campbell et al., 2010). Moreover, Jensen-Campbell and Graziano (2001) showed that agreeable individuals are less likely to concentrate on conflicts and negative aspects and more likely to use conflict resolution approaches to replace struggles and challenges and negotiate outcomes that benefit a whole group. This might be because agreeable individuals have been found to automatically engage neural mechanisms associated with the self-control of emotions (i.e., right IPFC) to regulate negative affects associated with conflict-related signals (Haas et al., 2007).

It is important to mention that agreeableness is divided into two correlated subdimensions: Compassion — tendencies toward empathy, sympathy, and concern for others and Politeness — motivation to conform to social norms and avoid aggression and exploitation (Hou et al., 2017). These two subdimensions might be differentially related to aspects of social information processing and social conformity. Previous studies have shown that agreeableness and conscientiousness together can associate with susceptibility to social pressure, whereas extraversion and openness are negatively correlated with social conformity (DeYoung et al., 2002; Kosloff et al., 2017). Therefore, the use of only agreeableness in selecting our participants may have influenced our results. Besides, the distribution of agreeableness scores in our subjects was shifted, varying between  $-10$  and  $32$  (on a  $-40$  to  $40$  scale), and the subjects thus did not represent far extremes of the agreeableness trait. Naturally,

correlation measures would be more optimal to reveal the association between conformity behavior and agreeableness, but this is not feasible with the current sample sizes. Our findings, however, provide a valuable basis for building hypotheses for larger-scale correlative studies, emphasizing the importance of carefully controlling all the factors influencing the measures of conformity behavior.

In this study, although we observed negative conformity to group opinion (subjects changed their ratings toward lower group recommendation), we could not examine the neural predictors of behavioral conformity (social conflict followed by opinion change versus conflict not followed by change) due to the small number of trials after controlling for confounding factors. In other words, after selecting trials with intermediate initial ratings (4 and 5) to eliminate the regression-to-mean effect, we subsequently narrowed trials to ones with negative group feedback with a change to negative direction and trials with negative group feedback with no rating change or change to the positive direction used as control trials. After these steps, we were left with too little data to compare conformity trials versus trials without opinion change.

Taken together, our imaging results suggest that cooperation of the brain mentalizing and conflict monitoring networks might support processing of perceived discrepancy with peer group recommendation. In this view, the conflict monitoring network would follow the discrepancy by creating aversive feelings and the need for opinion change, while the role of the mentalizing network would be to evaluate the consequences of the social conflict. Based on our behavioral findings, the mentalizing system would more effectively resolve situations where social consequences are the most salient, that is, upon negative conflict. Furthermore, in this study, agreeableness was not found to mediate conformity behavior. However, a difference between high and low agreeable individuals was observed in beta-band oscillations when an individual's and group's recommendations matched, consistent with social reward.

## Acknowledgment

The authors thank Pessi Lyyra for his comments and insights in planning data analysis and Amit Jaiswal for his contribution to writing the MEG analysis pipeline.

## Disclosure statement

No potential conflict of interest was reported by the author(s).



## Funding

This study was funded by TEKES (Project 40232/142953/31/201) and the Academy of Finland (Project 298456).

## ORCID

Fatemeh Irani  <http://orcid.org/0000-0002-2519-8944>  
 Joona Muotka  <http://orcid.org/0000-0002-7113-903X>  
 Simo Monto  <http://orcid.org/0000-0002-8191-6146>  
 Tiina Parviainen  <http://orcid.org/0000-0001-6992-5157>

## References

- Ali, A., Ul Haq, J., Hussain, S., Qadir, A., & Bukhari, S. A. H. (2022). OCEAN traits: Who shares more word of mouth? *Journal of Promotion Management* 28, 749–773. <https://doi.org/10.1080/10496491.2021.2015510>
- Baek, E. C., O'Donnell, M. B., Scholz, C., Pei, R., Garcia, J. O., Vettel, J. M., & Falk, E. B. (2021). Activity in the brain's valuation and mentalizing networks is associated with propagation of online recommendations. *Scientific Reports*, 11(1), 1–11. <https://doi.org/10.1038/s41598-021-90420-2>
- Barasch, A. (2020). The consequences of sharing. *Current Opinion in Psychology*, 31, 61–66. <https://doi.org/10.1016/j.copsyc.2019.06.027>
- Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *NeuroImage*, 49(3), 2687–2696. <https://doi.org/10.1016/j.neuroimage.2009.10.070>
- Briñol, P., Petty, R. E., Durso, G. R. O., & Rucker, D. D. (2017). Power and persuasion: Processes by which perceived power can influence evaluative judgments. *Review of General Psychology*, 21(3), 223–241. <https://doi.org/10.1037/gpr0000119>
- Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology*, 20(13), 1165–1170. <https://doi.org/10.1016/j.cub.2010.04.055>
- Cascio, C. N., O'Donnell, M. B., Bayer, J., Tinney, F. J., & Falk, E. B. (2015). Neural correlates of susceptibility to group opinions in online word-of-mouth recommendations. *Journal of Marketing Research*, 52(4), 559–575. <https://doi.org/10.1509/jmr.13.0611>
- Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, 49(4), 3198–3209. <https://doi.org/10.1016/j.neuroimage.2009.11.080>
- Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. B. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology*, 49(2), 220–238. <https://doi.org/10.1111/j.1469-8986.2011.01293.x>
- Chen, J., Wu, Y., Tong, G., Guan, X., & Zhou, X. (2012). ERP correlates of social conformity in a line judgment task. *BMC Neuroscience*, 13(1), 43. <https://doi.org/10.1186/1471-2202-13-43>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55(May), 591–621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>
- Cialdini, R. B., Wosinska, W., Barrett, D. W., Butner, J., & Gornik Durose, M. (1999). Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality & Social Psychology Bulletin*, 25(10), 1242–1253. <https://doi.org/10.1177/0146167299258006>
- Cohen, M. X., Elger, C. E., & Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *NeuroImage*, 35(2), 968–978. <https://doi.org/10.1016/j.neuroimage.2006.11.056>
- Colosio, M., Shestakova, X. A., Nikulin, V. V., Blagovechtchenski, E., & Klucharev, X. (2017). *Neural mechanisms of cognitive Dissonance (Revised): An EEG study*. <https://doi.org/10.1523/JNEUROSCI.3209-16.2017>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. *NeuroImage*, 9(2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>
- DeYoung, C. G. (2010). Personality neuroscience and the biology of traits. *Social and Personality Psychology Compass*, 4(12), 1165–1180. <https://doi.org/10.1111/j.1751-9004.2010.00327.x>
- DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-Order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, 33(4), 533–552. [https://doi.org/10.1016/S0191-8869\(01\)00171-4](https://doi.org/10.1016/S0191-8869(01)00171-4)
- Eisingerich, A. B., Chun, H. H., Liu, Y., Jia, H. (., & Bell, S. J. (2015). Why recommend a brand face-to-face but not on Facebook? How word-of-mouth on online social sites differs from traditional word-of-mouth. *Journal of Consumer Psychology*, 25(1), 120–128. <https://doi.org/10.1016/j.jcps.2014.05.004>
- Engel, A. K., & Fries, P. (2010). Beta-Band oscillations-signalling the status quo? *Current Opinion in Neurobiology*, 20(2), 156–165. <https://doi.org/10.1016/j.conb.2010.02.015>
- Festinger, L. (1957). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.
- Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-Resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4), 272–284. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:4<272:AID-HBM10>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<272:AID-HBM10>3.0.CO;2-4)
- Fiske, S. T. (2018). *Social beings: Core motives in social psychology*. John Wiley & Sons.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R. Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python *Frontiers in Neuroscience*, 267. <https://doi.org/10.3389/fnins.2013.00267>
- Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 694–699. <https://doi.org/10.1073/pnas.98.2.694>
- Gugenishvili, I., Francu, R. E., & Koporcic, N. (2021). I give a dime if you do, too! The influence of descriptive norms on perceived impact, personal involvement, and monetary donation intentions. *Journal of Consumer Behaviour*, (June 2021), 167–179. <https://doi.org/10.1002/cb.1980>
- Haas, B. W., Omura, K., Constable, R. T., & Canli, T. (2007). Emotional conflict and neuroticism: Personality-dependent activation in the Amygdala and Subgenual anterior cingulate. *Behavioral Neuroscience*, 121(2), 249–256. <https://doi.org/10.1037/0735-7044.121.2.249>

- Hosseini, A. H., & Holroyd, C. B. (2015). Reward feedback stimuli elicit high-beta EEG oscillations in human dorsolateral prefrontal cortex. *Scientific Reports*, 5(July), 1–8. <https://doi.org/10.1038/srep13021>
- Hou, X., Allen, T. A., Wei, D., Huang, H., Wang, K., DeYoung, C. G., & Qiu, J. (2017). Trait compassion is associated with the neural substrate of empathy. *Cognitive, Affective & Behavioral Neuroscience*, 17(5), 1018–1027. <https://doi.org/10.3758/s13415-017-0529-5>
- Huang, Y., Kendrick, K. M., & Yu, R. (2014). Social conflicts elicit an N400-like component. *Neuropsychologia*, 65, 211–220. <https://doi.org/10.1016/j.neuropsychologia.2014.10.032>
- Hymers, M., Prendergast, G., Johnson, S. R., & Green, G. G. R. (2010). Source stability index: A novel beamforming based localisation metric. *NeuroImage*, 49(2), 1385–1397. <https://doi.org/10.1016/j.neuroimage.2009.09.055>
- Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13(3), 411–430. [https://doi.org/10.1016/s0893-6080\(00\)00026-5](https://doi.org/10.1016/s0893-6080(00)00026-5)
- Izuma, K., Matsumoto, M., Murayama, K., Samejima, K., Sadato, N., & Matsumoto, K. (2010). Neural correlates of cognitive dissonance and choice-induced preference change. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51), 22014–22019. <https://doi.org/10.1073/pnas.1011879108>
- Jensen-Campbell, L. A., & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality*, 69(2), 323–362. <https://doi.org/10.1111/1467-6494.00148>
- Jensen-Campbell, L. A., Knack, J. M., & Gomez, H. L. (2010). The psychology of nice people. *Social and Personality Psychology Compass*, 4(11), 1042–1056. <https://doi.org/10.1111/j.1751-9004.2010.00307.x>
- Kim, B.-R., Liss, A., Rao, M., Singer, Z., & Compton, R. J. (2012). Social deviance activates the brain's error-monitoring system. *Cognitive, Affective, & Behavioral Neuroscience*, 12(1), 65–73. <https://doi.org/10.3758/s13415-011-0067-5>
- Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, 61(1), 140–151. <https://doi.org/10.1016/j.neuron.2008.11.027>
- Konstabel, K., Lönnqvist, J. E., Walkowitz, G., Konstabel, K., & Verkasalo, M. (2012). The “Short Five” (S5): Measuring personality traits using comprehensive single items. *European Journal of Personality*, 26(1), 13–29. <https://doi.org/10.1002/per.813>
- Kosloff, S., Irish, S., Perreault, L., Anderson, G., & Nottbohm, A. (2017). Assessing relationships between conformity and meta-traits in an Asch-like paradigm. *Social Influence*, 12(2–3), 90–100. <https://doi.org/10.1080/15534510.2017.1371639>
- Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365–411. [https://doi.org/10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4)
- Levorsen, M., Ito, A., Suzuki, S., & Izuma, K. (2020). Testing the reinforcement learning hypothesis of social conformity. *Human Brain Mapping*, (November), 1–15. <https://doi.org/10.1002/hbm.25296>
- Marco-Pallarés, J., Cürcürell, D., Cunillera, T., García, R., Andrés-Pueyo, A., Münte, T. F., & Rodríguez-Fornells, A. (2008). Human oscillatory activity associated to reward processing in a gambling task. *Neuropsychologia*, 46(1), 241–248. <https://doi.org/10.1016/j.neuropsychologia.2007.07.016>
- Marco-Pallarés, J., Münte, T. F., & Rodríguez-Fornells, A. (2015). The role of high-frequency oscillatory activity in reward processing and learning. *Neuroscience and Biobehavioral Reviews*, 49, 1–7. <https://doi.org/10.1016/j.neubiorev.2014.11.014>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Matzler, K., Renzl, B., Müller, J., Herting, S., & Mooradian, T. A. (2008). Personality traits and knowledge sharing. *Journal of Economic Psychology*, 29(3), 301–313. <https://doi.org/10.1016/j.joep.2007.06.004>
- Miller, D. T., & Prentice, D. A. (2016). Changing norms to change behavior. *Annual Review of Psychology*, 339–363. <https://doi.org/10.1146/annurev-psych-010814-015013>
- Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, 341(6146), 647–651. <https://doi.org/10.1126/science.1240466>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus Version 8 User's Guide*. Muthén & Muthén 2017.
- Nook, E. C., & Zaki, J. (2015). Social norms shift behavioral and neural responses to foods. *Journal of Cognitive Neuroscience*, 27(7), 1412–1426.
- Packard, G., & Berger, J. (2017). How language shapes word of mouth's impact. *Journal of Marketing Research*, 54(4), 572–588. <https://doi.org/10.1509/jmr.15.0248>
- Packard, G., & Wooten, D. B. (2013). Compensatory knowledge signaling in consumer word-of-mouth. *Journal of Consumer Psychology*, 23(4), 434–450.
- Sambrook, T. D., & Goslin, J. (2015). A Neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, 141(1), 213–235. <https://doi.org/10.1037/bul0000006>
- Sassenhagen, J., & Draschkow, D. (2019). Cluster-Based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, 56(6), 1–8. <https://doi.org/10.1111/psyp.13335>
- Schnuerch, R., Richter, J., Koppehele-Gossel, J., & Gibbons, H. (2016). Multiple neural signatures of social proof and deviance during the observation of other people's preferences. *Psychophysiology*, 53(6), 823–836. <https://doi.org/10.1111/psyp.12636>
- Seeme, F., Green, D., & Kopp, C. (2019). Pluralistic ignorance: A trade-off between group-conformity and cognitive dissonance BT. In T. Gedeon, K. W. Wong, & M. Lee (Eds.), *Neural information processing* (pp. 695–706). Springer International Publishing.
- Shestakova, A., Rieskamp, J., Tugin, S., Ossadtchi, A., Krutitskaya, J., & Klucharev, V. (2013). Electrophysiological precursors of social conformity. *Social Cognitive and Affective Neuroscience*, 8(7), 756–763. <https://doi.org/10.1093/scan/nss064>
- Taulu, S., & Hari, R. (2009). Removal of magnetoencephalographic artifacts with temporal signal-space separation: Demonstration with single-trial auditory-evoked responses. *Human Brain Mapping*, 30(5), 1524–1534. <https://doi.org/10.1002/hbm.20627>
- Templeton, E. M., Stanton, M. V., & Zaki, J. (2016). Social norms shift preferences for healthy and unhealthy foods. *PLoS ONE*, 11(11), 1–16. <https://doi.org/10.1371/journal.pone.0166286>

- van de Vijver, I., Richard Ridderinkhof, K., & Cohen, M. X. (2011). Frontal oscillatory dynamics predict feedback learning and action adjustment. *Journal of Cognitive Neuroscience*, 23(12), 4106–4121. [https://doi.org/10.1162/jocn\\_a\\_00110](https://doi.org/10.1162/jocn_a_00110)
- Van Veen, V., Krug, M. K., Schooler, J. W., & Carter, C. S. (2009). Neural activity predicts attitude change in cognitive dissonance. *Nature Neuroscience*, 12(11), 1469–1474. <https://doi.org/10.1038/nn.2413>
- Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44(9), 867–880. <https://doi.org/10.1109/10.623056>
- Walsh, M. M., & Anderson, J. R. (2012). NIH public access. *Neuroscience & Biobehavioral Reviews*, 36(8), 1870–1884. <https://doi.org/10.1016/j.neubiorev.2012.05.008.Learning>
- Wang, F., Wang, X., Wang, F., Gao, L., Rao, H., & Pan, Y. (2019). Agreeableness modulates group member risky decision-making behavior and brain activity. *NeuroImage*, 202(June), 116100. <https://doi.org/10.1016/j.neuroimage.2019.116100>
- Wasylyshyn, N., Falk, B. H., Garcia, J. O., Cascio, C. N., O'Donnell, M. B., Bingham, C. R., Simons-Morton, B., Vettel, J. M., Falk, E. B. (2018). Global brain dynamics during social exclusion predict subsequent behavioral conformity. *Social Cognitive and Affective Neuroscience*, 13(2), 182–191. <https://doi.org/10.1093/scan/nsy007>
- Yu, R., & Chen, L. (2014). The need to control for regression to the mean in social psychology studies. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2014.01574>
- Yu, R., & Sun, S. (2013). To conform or not to conform: Spontaneous conformity diminishes the sensitivity to monetary outcomes. *PLoS ONE*, 8(5), 1–9. <https://doi.org/10.1371/journal.pone.0064530>
- Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science*, 22(7), 894–900. <https://doi.org/10.1177/0956797611411057>
- Zhang, L., Guo, D., Wen, X., & Li, Y. (2020). Effect of other visible reviews' votes and personality on review helpfulness evaluation: An event-related potentials study. *Electronic Commerce Research*, (0123456789). <https://doi.org/10.1007/s10660-020-09419-y>
- Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience*, (July). <https://doi.org/10.1093/scan/nsaa089>
- Zubarev, I., Klucharev, V., Ossadtchi, A., Moiseeva, V., & Shestakova, A. (2017). MEG signatures of a perceived match or mismatch between individual and group opinions. *Frontiers in Neuroscience*. <https://doi.org/10.3389/fnins.2017.00010>