Hautala, Annika; Kluge, Annika; Hameiri, Boaz; Zebarjadi, Niloufar; Levy, Jonathan

Examining implicit neural bias against vaccine hesitancy

Published in:
Social Neuroscience

DOI:
10.1080/17470919.2022.2162119

Published: 28/12/2022

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Hautala, A., Kluge, A., Hameiri, B., Zebarjadi, N., & Levy, J. (2022). Examining implicit neural bias against vaccine hesitancy. *Social Neuroscience*. https://doi.org/10.1080/17470919.2022.2162119

Social Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/psns20

Examining implicit neural bias against vaccine hesitancy

Annika Hautala, Annika Kluge, Boaz Hameiri, Niloufar Zebarjadi & Jonathan Levy

To cite this article: Annika Hautala, Annika Kluge, Boaz Hameiri, Niloufar Zebarjadi & Jonathan Levy (2022): Examining implicit neural bias against vaccine hesitancy, Social Neuroscience, DOI: 10.1080/17470919.2022.2162119

To link to this article: https://doi.org/10.1080/17470919.2022.2162119

Published online: 28 Dec 2022.

SOCIAL NEUROSCIENCE
https://doi.org/10.1080/17470919.2022.2162119

Routledge
Taylor & Francis Group

Examining implicit neural bias against vaccine hesitancy

Annika Hautala[a], Annika Kluge[a], Boaz Hameiri[b], Niloufar Zebarjadi[a] and Jonathan Levy[a,c]

[a]Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland; [b]The Evens Program in Conflict Resolution and Mediation, Tel Aviv University, Tel Aviv-Yafo, Israel; [c]Baruch Ivcher School of Psychology, Reichman University (Interdisciplinary Center Herzliya), Herzliya, Israel

ABSTRACT

COVID-19 pandemic has changed the world in many ways. At the societal level, disparities in attitudes toward the COVID-19 vaccines have led to polarization and intense animosity. In this study, we use a novel paradoxical thinking intervention that was found to be effective in difficult and violent intergroup contexts, and measure its effectiveness in a novel unobtrusive way in an important and timely context, namely prejudice against vaccine hesitancy. In the midst of a vaccination campaign, 36 young Finnish adults either went through the intervention or through a control condition. Magnetoencephalography then measured a neural response that is thought to reflect intergroup bias and possibly implicit prejudice. This neural response was reduced among the participants receiving the intervention, compared to the control group, thereby suggesting a potential mechanism of intergroup bias that is affected by a psychological intervention even during a campaign that castigates aggressively vaccine-hesitant individuals. The findings reported here contribute to the recent accumulating evidence of the potential of neuroimaging to reveal covert mental effects by psychological interventions. They may also have societal implications for moderating the polarized attitudes in a new era of pandemics.

ARTICLE HISTORY
Received 15 August 2022
Revised 30 November 2022
Published online 29 December 2022

KEYWORDS
Vaccine hesitancy; Implicit neural prejudice; intergroup bias; Implicit association test (IAT); alpha rhythm

Introduction

The COVID-19 pandemic has developed to be one of the largest public health crises in the world (Lades et al., 2020). Beyond the imminent implications on public health and economic power, the pandemic impacted group relations (Adler et al., 2021; Fuochi et al., 2021; Han et al., 2021), for instance by finding scapegoats to blame for the negative outcomes of the pandemic (Adler et al., 2021; Cohn, 2012). Around the onset of 2021, vaccines targeting COVID-19 were introduced and this event gradually led to heated polarization (Bajwa, 2021; Jiang et al., 2021). While most societies embraced the introduction of the vaccines, a minority of individuals were hesitant to uptake the vaccine and in some cases expressed strong opposition or conspiracy theories. Although vaccine hesitancy has long preceded the COVID-19 pandemic and was directed toward various types of vaccines for decades (Schuster et al., 2015), this societal division and the ensuing expressions of prejudice and animosity have been reinforced by the COVID-19 pandemic as never before (Tram et al., 2022).

One of the most prevalent approaches for evaluating implicit prejudice or intergroup bias is by implementing the Implicit Association Test (IAT) (Chang et al., 2016; Hewstone et al., 2002). IAT relies on the slower behavioral responses on incongruent (where the outgroup is good) versus congruent (where the outgroup is bad) pairs of stimuli. Despite the extensive use of IAT to study intergroup bias and implicit prejudice (Kurdi et al., 2018), IAT has been criticized to hardly reflect real-life intergroup behavior (Hofmann et al., 2005; Oswald et al., 2013; Wilson & Scior, 2014). However, although bias and prejudice may be captured by implicit tests such as the IAT in certain contexts, these processes often rely on complex, automatic interactive, and consecutive mental mechanisms that cannot be reflected simply by variations in response-time as IAT traditionally measures. Traditional models of prejudice (Crandall & Eshleman, 2005; Devine, 1989; Dovidio & Gaertner, 1986; Plant & Devine, 1998) have indeed suggested that prejudice relies on such automatic and covert mechanisms, while receiving empirical support from early psychophysiological explorations (Cooper & Siegel, 1956; Rankin & Campbell, 1955). However, it was not until less than two decades ago that the technological progress in neuroimaging has enabled delineating the different processes during prejudice and their neural underpinnings: detection of bias, control of prejudice and inhibition of prejudice control (Amodio & Cikara, 2021; Amodio & Devine, 2006; Amodio & Ratner, 2011). Furthermore, relying on neuroimaging not only uncovers important psychological processes but has been shown to predict real-life future behavioral

CONTACT Jonathan Levy jonathan.levy@aalto.fi Department of Neuroscience and Biomedical Engineering, Aalto University, Rakentajanaukio 2, 02150 Espoo, Finland

outcomes (Falk et al., 2015; Gabrieli et al., 2015; Poldrack et al., 2018). Recently, a new approach was implemented by using the IAT framed within the Israeli-Palestinian context during MEG recording of rhythmic neural activity, while individuals were performing the IAT task inside the MEG room (Levy et al., 2021). Similar to well-studied behavioral IAT index, to obtain a neural marker of intergroup bias, the analysis focused on contrasting the recorded data in the incongruent (i.e., ingroup bad, outgroup good) condition with the congruent (i.e., ingroup good, outgroup bad) condition. By examining neural rhythmic activity across a broad range from very low to very high-frequency bands, the only rhythm found to underlie the intergroup bias was posterior alpha rhythm. Importantly, after the MEG session, participants from both conflict groups (i.e., Palestinians and Israelis) in that study dialoged with each other under quasi-naturalistic settings, and in-depth interview evaluated their intergroup attitudes. These data were used to evaluate real-life intergroup behavior and attitudes. The results from that study not only revealed a neural marker (i.e., posterior alpha) of implicit prejudice and intergroup bias but also an association between this marker and those real-life representations of intergroup behavior and attitudes (Levy et al., 2021). In addition, another recent study found that the attenuation of this neural alpha rhythm by another psychological intervention can predict future support in peacemaking even years later (Levy et al., 2022). None of the self-reported or behavioral IAT measures in these studies could predict any real-life behavior or future change in peacemaking attitudes (Levy et al., 2021, 2022). Hence, examining this neural marker during the IAT is a promising emerging strategy for examining covert processes and mental mechanisms that may reflect real-life intergroup behavior and attitudes, but more empirical evidence from other social and experimental contexts is needed to further test the validity of this claim.

In parallel to studying prejudice and its mechanisms, in recent decades, social scientists moved forward to create evidence-based knowledge on how to devise and structure interventions that could be useful for practitioners and decision makers for reducing prejudice (Badea & Sherman, 2019; Paluck, 2016) or promoting intergroup reconciliation (Bruneau et al., 2020; Hameiri et al., 2018; Hasler et al., 2021; Hasson et al., 2019). One of the promising interventions for moderating heated intergroup attitudes is the recent paradoxical thinking intervention (Bar-Tal et al., 2021). This intervention has been found to be efficient in unfreezing of previously held conflict-supporting attitudes and open the mind of individuals to different information in various contexts, ranging from intractable and violent conflicts, to immigration and gender (Hameiri et al., 2020; Knab & Steffens, 2022; Swann et al., 1988). The intervention leans on the strategic use of information (often transmitted via auditory statements, visual banners or vignettes), which is consistent with individuals' beliefs, but is also extreme, exaggerated, or even absurd. This approach is thought to circumvent personality defense mechanisms; individual attitudes are not being directly confronted (Bar-Tal et al., 2021). As a consequence, individuals may paradoxically perceive their own social attitude beliefs as irrational, and thereby mentally "unfreeze", and this can potentially lead to attitude moderation (Hameiri et al., 2014). Although this phenomenon has been tackled in numerous studies, it has not been investigated using neuroimaging, which may uncover neural consequences of such "unfreezing" of mental states.

In the current study, we first examined implicit neural prejudice (Levy et al., 2021) toward COVID-19 vaccine hesitancy by implementing the Implicit Association Test (IAT) (Kurdi et al., 2018) during a magnetoencephalography (MEG) scan. In a previous study, the reduction of the implicit neural prejudice response (i.e., the occipital alpha rhythm during the IAT) was found to predict real-life intergroup dialog style and support for reconciliation (Levy et al., 2021). Second, we used a promising intervention for moderating heated intergroup attitudes, the paradoxical thinking intervention, which has been repeatedly evidenced to be effective in moderating well-entrenched negative intergroup attitudes (e.g., prejudice) (Bar-Tal et al., 2021). By exposing participants to multiple paradoxical auditory statements targeting COVID-19 Vaccination (e.g., "Due to the limited number of intensive care units and ventilators in Finland, their capacity should be maintained by not providing intensive care to the anti-vax individuals at all. The place in the intensive care unit should be earned by taking vaccination"), we tested whether the intervention would reduce implicit neural prejudice and the behavioral IAT bias. A recent study found that another intergroup intervention can attenuate this neural marker, not the behavioral IAT marker, and that this attenuation is predictive of future support in intergroup peacemaking (Levy et al., 2022). We finally aimed to explore whether these measures can be explained by explicit negative attitudes toward the targeted individuals (i.e., vaccine-hesitant individuals) and self-reported reactions toward the intervention stimuli.

Methods

Participants

A priori power analysis was conducted based on the previous study that examined the neural implicit prejudice response (Levy et al., 2021) with a Cohen's d effect size =.70. This effect was calculated at the sensor level by localizing the peak alpha-band sensor effect of the neural IAT effect (i.e., incongruent vs. congruent contrast), thereby matching the similar sensor effect that was hypothesized in the present study. An *a priori* power analysis for computing the required sample size was conducted using G*Power (Faul et al., 2007) for a single-tailed (i.e., alpha reduction) difference in means of matched pairs. This calculation indicated that a sample size of 17 per group would be sufficient to detect this targeted neural response (i.e., sensor-level alpha reduction) at 85% power. We therefore recruited 36 participants (23 females, mean age ± SD = 24.5 ± 7.27). Noteworthy, this sample calculation should have considered the differences in experimental contexts and hypotheses and thereby increase the power and sample aimed for; this is elaborated in the Discussion. All 36 subjects were right-handed, native Finns, over 18 years old and did not have any psychiatric or neurological disorders. MEG compatibility and history of neurological and psychiatric disorders were checked before the recruitments by asking subjects to fill a primary online survey. The questions in the online survey were about the history of disorders mentioned earlier, MEG compatibility and their demographics (such as age, education level, and gender). After recruiting the participants, the participants were pseudo-randomly divided into two groups: control group and intervention group. Acquired data were excluded from the analysis only under circumstances of lack of compliance with the specifications in the experimental task. In the current study, the acquired data of one participant was excluded from the analysis due to the failure to complete the IAT. All participants read an information sheet and a privacy notice paper. In addition, they signed the participation confirmation form. All were approved by the Aalto University Research Ethics Committee.

Experimental design

As illustrated in Figure 1, prior to attending the experiment, participant's attitudes toward anti-vax individuals were evaluated via a brief survey containing questions related to participant's education, political orientation, political activism, their negative perception of anti-vax individuals and perceived threat related to the anti-vax individuals. The information collected from this questionnaire was used to ensure a balance between the two studied groups (intervention and control): for each participant the numerical response of their answers was stored in the Excel and the average was calculated from these numbers. Negative perception of anti-vax individuals was assessed via a 7-item 7-point Likert-type scale ranging from totally disagree (1) to totally agree (7). The items were as follows: i) "I trust the vaccines ability to improve our health"; ii) "I feel anger toward people who refuse to get vaccinated against corona without a medical reason"; iii) "I feel that the people refusing to get the corona-vaccine are downplaying the seriousness of the disease"; iv) "I think the people refusing to get corona-vaccines without a medical reason are selfish"; v) "I feel positively toward people who have gotten vaccinated against the coronavirus"; vi) "I think the vaccination passport is fair"; vii) "We should make people get vaccinated against coronavirus using whatever means necessarily, if they do not have a medical reason to refuse". Moreover, the feeling of threat that the anti-vax individuals induced was assessed via a 4-item 7-point Likert-type scale ranging as well from totally disagree (1) to totally agree (7). This perceived threat scale assessed to what extent the participant feels that the anti-vax individuals endanger i) the security of Finland, ii) the welfare state of Finland, iii) the society of Finland, or iv) the participant and their family.

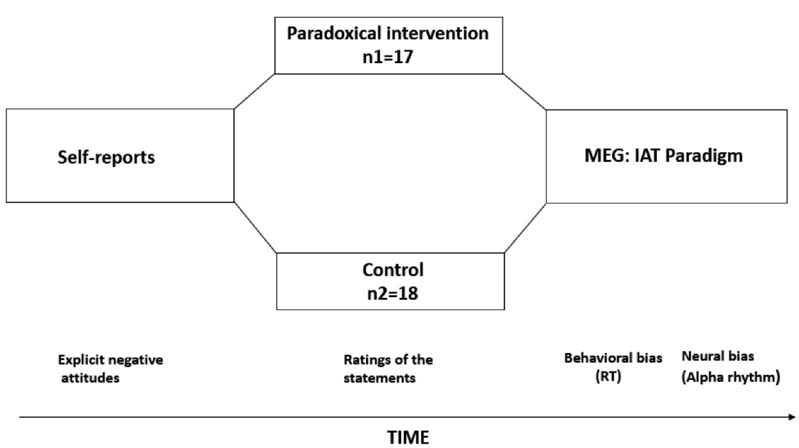

Figure 1. The experimental design of the study. The experiment consisted of three parts: explicit attitudes towards anti-vax individuals, intervention/control, and the IAT-MEG session.

Intervention/Control stimuli

During the intervention/control, participants listened to 22 auditory statements, and after each, a question prompting them to rate their level of agreement with the statement's message. The answer choices were as follows: Do not agree (1), I don't know (2), and Agree (3). After the statement ended, it took 1.5 s until this question appeared on the screen. After the participant answered the question, it took 5 s until the next statement appeared. Overall, depending on the participant's response time, this part of the measurement lasted from 10 min to 12 min. The audio stimuli were presented with a Panphonics SoundShower speaker and with the Presentation software (Neurobehavioral Systems Inc.; version 22.0, Berkeley, CA, USA). All participants were seated in the MEG room in front of a projector placed at a viewing distance of 120 cm. All participants in both experimental groups received auditory statements. Participants in the intervention group perceived statements following the paradoxical thinking principles (Hameiri et al., 2014, 2016; Knab et al., 2021), whereas participants in the control group perceived statements following the conventional inconsistency approach (Bartunek, 1993). The experimental rationale of intervention was that the paradoxical thinking intervention would induce threat to self-identity and thereby moderate attitudes toward anti-vax individuals, in contrast to the control which would not moderate the attitudes, and the participants instead would clearly disagree with those statements (Hameiri et al., 2018). The intervention statements positively related to individuals who have taken the corona vaccine and negatively related to those who did not take the vaccine, but these statements were extreme, exaggerated, or even absurd. By contrast, the control statements positively related to individuals who did not take the vaccine and negatively about those who took the vaccine. All statements were collected from social media, news, and other internet forums. The length of the paradoxical thinking stimuli in the experiment was 468 (11.87 ± 1.21) seconds and 521 (23.68 ± 2.75) words and the length of the conventional inconsistent stimuli was 460 (12.25 ± 1.12) seconds and 497 (22.59 ± 3.07) words. There was no significant difference between the statements in the two groups ($p = 0.21$).

MEG recordings

To evaluate neural intergroup bias, we acquired IAT data while MEG collected continuous oscillatory activity using a 306-channel MEG device (MEGIN Oy, Helsinki, Finland) with 102 radial magnetometers and 204 planar gradiometers in a 3-layered magnetically shielded room (Imedco AG, Hägendorf, Switzerland). Before the MEG measurements, multiple preparations were done to the participants. First, we placed electrooculogram (EOG) electrodes to the participant's forehead and over and under the eyes to obtain eye blinking data. In addition, we placed one electrode on the neck to ground the participants. Also, we attached continuous Head Position Indicator (HPI) coils to the heads of the participants, and lastly, we performed digitization of the participants anatomical features, for example nose and head, to correct head movements. In addition, we familiarized the participants with the scanning procedure and asked them to avoid any bodily movements during the scan.

Implicit association test (IAT) measures implicit negative associations toward preselected targets, for example toward outgroup members (Greenwald et al., 1998). In this study, the selected outgroup was "anti-vax individuals", and we assumed that there will be negative implicit association toward the members of this outgroup. To adapt the standard IAT procedure (Greenwald et al., 2003) to MEG settings, we based our design on a previous MEG IAT study (Levy et al., 2021). There were 10 blocks in total used in the IAT of this study. Blocks 1 to 4 and 6 to 9 were used as practice blocks, and blocks 5 and 10 were used in the analysis. Blocks 5 and 10 took up 78% of the whole IAT experiment and were the congruent and incongruent condition. These blocks used for the analysis had altogether 200 stimuli. Blocks were counterbalanced across participants so that half of the participants saw blocks 6 to 10 before blocks 1 to 5 and other half of the participants saw blocks 1 to 5 before the blocks 6 to 10. The order of stimuli inside a block was randomized, and there were instructions on how to sort the stimuli at the start of each block. The IAT stimuli were presented until a response from a participant was registered. These responses were delivered by a response pad by lifting either the middle or index finger of the right hand. Index finger corresponded to the left side of the screen and middle finger to the right side of the screen, thus targeting one of the four IAT categories. Participants were notified when an error was made – if the word or picture was assigned to the opposite category it belonged – and asked to correct their response.

In the first block, the task was to sort bad and good words into the right categories as quickly as possible. The discrimination was designed to be as easy as possible, thus the words used as a stimulus were clearly either unpleasant or pleasant. In the second block, the task was to sort the pictures of the people who support the COVID-19 vaccines and pictures of the people who do not support the COVID-19 vaccines into the current categories. The categories were “Anti-vax” and “Pro-vax”. In the third block, the participants sorted stimuli from all categories, i.e., words and pictures and all four categories, “Good”, “Bad”, “Pro-vax”, and “Anti-vax”, were displayed on the screen. In the fourth block, the task was the same as in the previous block, but the names of the categories were not displayed anymore in order to practice the task with fewer eye movements for the experimental block that was used for the analysis. As these blocks were the congruent condition, on the one side of the screen were the categories “Good” and “Pro-vax” and on the other side of the screen were the categories “Bad” and “Anti-vax”. The experimental fifth block had the same logic as the block 4 – sort the words and pictures without seeing the name of the categories anymore. This fifth block was also a congruent block. Blocks 6 to 10 followed the same logic as blocks 1 to 5, but in these blocks when the task was to sort stimuli from all four categories (blocks 8 to 10), the sides of the “Pro-vax” and “Anti-vax” category names were switched so that the categories “Pro-vax” and “Bad” were on the one side of the screen and the “Anti-vax” and “Good” on the other side of the screen. These blocks were presenting the incongruent condition.

IAT neuroimaging stimuli

The visual stimuli were presented with a PROPixx projector (1920 × 1080 at 120, max 1440 Hz) and with the Presentation software (Neurobehavioral Systems Inc.; version 22.0, Berkeley, CA, USA). Participants were seated in the MEG room in front of a projector placed at a viewing distance of 120 cm, except for one participant, who was placed at a viewing distance of 90 cm due to the strong myopia. The visual stimuli were presented in the center of the screen.

There were two types of visual stimuli used in the IAT part of the study. The first stimuli type was a series of pleasant words (hereafter labeled “Good”) and unpleasant words (hereafter labeled “Bad”). There were 10 words in each category. The second type of stimuli were pictures of anti-vax individuals and pictures of pro-vax individuals also 10 in each category. Fundamentally, IAT relies on contrasting trials with congruent (C) versus incongruent (IC) implicit associations. In other words, it relies on the slower behavioral responses on incongruent (where the outgroup is good) versus congruent (where the outgroup is bad) pairs of stimuli (Levy et al., 2021). In this study, in the congruent trial contained the following category name arrangements, where “Good” and “Pro-vax” were on the one side, and in turn, the categories “Bad” and “Anti-vax” on the other side. In an incongruent trial, the situation was reversed so that the categories “Pro-vax” and “Bad” were on the one side of the screen and the “Anti-vax” and “Good” on the other side of the screen.

Data preprocessing

The preprocessing of the collected data was performed by using Spyder software and MNE-Python toolbox (Python 3.9.4 with the MNE-python 0.21.1). This preprocessing step consisted of four different steps. At the very beginning of the first preprocessing step, we marked all bad channels manually. These marked channels, for example, included channels with a lot of noisiness or channels that were not functioning properly, such that they, for instance, did not show any signal at all. In this first step, the data were also lowpass filtered at 40 Hz using linear-phase FIR-filter with delay compensation. After these steps, we used Maxwell filtering to attenuate measurement artifacts and magnetic interference from inside and outside of the sensory array as well as transform data due to head movements. In the second step, we bandpass-filtered the data at 1–200 hz and performed an independent component analysis (ICA) to detect and remove heart and eye artifacts. Although there was automatic detection of eyeblinks and saccades by electrocardiography (ECG) and heartbeat by electro-oculogram (EOG), the artifacts were manually inspected and excluded as well. In the third step, we aligned the data with the onset of the stimulus and detected the trial events. As stated earlier (c. f. sub-section “IAT design”), we took only stimuli from blocks 5 and 10 into account during the analysis, because other blocks were used as test blocks in the IAT. In the last step, we performed epoching. We created time windows from −0.5 to 2 s relative to stimulus onset and extracted the response times related to the stimuli for each participant. The extracted response times for congruent and incongruent conditions were the average response times of the conditions. Also, in this step, we extracted the number of errors made in each condition.

Behavioral data analysis

We extracted the average response times related to the stimuli for each participant for congruent and incongruent conditions. We collected these extracted response times into Excel for further behavioral-level analysis. In order to test the first hypothesis at the behavioral level, that is, whether there is an IAT bias toward the anti-vax people in the control group, we calculated a response time difference between the incongruent and congruent conditions. To test the second hypothesis that the intervention would reduce this behavioral-level bias in the intervention group but not in the control group, we compared a response time difference between the two conditions in the two groups. We tested the significance of the response time differences in the congruent and incongruent conditions and between the control group and intervention group by using a two-tailed Student's t-test

Time-frequency analysis

For the neural-level analysis, we computed the time-frequency representation (TFR) by applying the Hanning tapers method on each trial after the four pre-processing steps. We also calculated the average power over epochs for congruent and incongruent conditions. We performed this time-frequency analysis on the data by using MATLAB 7.12.0 (MathWorks®, Natick, MA, USA), and MATLAB R2021 (MathWorks®, Natick, MA, USA) and the FieldTrip software toolbox. The oscillatory components often show power changes related to experimental events (Tallon-Baudry et al., 1999). Thus, by calculating the average power over epochs, the oscillations that are related to the repetition of the same condition can be seen. To evaluate Time-Frequency Representations (TFRs) of power for each trial and to compute Fast Fourier Transform (FFT) for short sliding time windows of 500 ms in the 1–40 Hz frequency range, we applied a Hanning taper to each epoch of the 306-sensor data (Zebarjadi et al., 2021). More specifically, the time window was sliding from −0.5 to 2 s in step of 50 ms. In addition, we subtracted evoked responses from the induced activity and eventually calculated TFRs for the statistically significant contrast between the two conditions (congruent and incongruent).

Statistical analysis

We performed the statistical tests described in this paragraph to examine the first hypothesis and the second hypothesis. We tested the significance of the response time differences in the congruent and incongruent conditions and between the control group and intervention group by using a two-tailed Student's t-test (c.f. subsection "Behavioral data analysis"). Moreover, statistical procedures on the collected MEG data relied on a two-tailed non-parametric approach based on the earlier study of Levy et al. (2018) and Levy et al. (2019) and these procedures did assess the significance of the power values (Levy et al., 2018, 2019). This two-tailed non-parametric approach took the cross-participant variance into consideration and was used for a correction for multiple comparisons. This procedure was cluster-based, and it allowed a correction for multiple comparisons at all sensor analyses. As stated in the study of Levy et al. (2019), this used approach is beneficial and valuable, because it does not make any assumptions related to the underlying distribution and, moreover, it is unaffected by partial dependence that occurs between neighboring time-frequency pixels (Levy et al., 2018). Hence, the same approach was used in the present study.

More specifically, what happened during this two-tailed non-parametric approach used on the MEG data was that we first computed t-values per participant, channel, time, and frequency. These computations represented the contrast between the incongruent and congruent conditions. Secondly, we pooled t-values over all participants in order to define the test statistic. During this second step, we searched the time-frequency clusters with effects that were significant. The effects were significant, if there was significance at the random effects level after correcting for multiple comparisons along the time dimensions and frequency dimensions.

To test the significance of the group-level statistic, we randomly multiplied the t-value of each individual by 1 or −1 and then summed them over all participants. This multiplication with 1 or −1 corresponded to the permutation of the original conditions in that participant. We repeated this randomization 1000 times in order to receive randomization distribution for the group-level statistic. Furthermore, for each of these 1000 randomizations, we kept only the maximal and minimal cluster-level test statistic across all clusters. Then, for each cluster from the data, we determined the fraction of the maximum or minimum cluster-level test statistic that was higher or smaller than the cluster-level test statistic and then the smaller of the two fractions was retained and divided it by 1000 and this gave the multiple comparisons corrected significance thresholds for a two-sided test. Finally, we defined the p-value by the proportion of values in the randomization distribution that exceed this test-statistic.

Correlation analysis

We used two different correlation approaches to explore whether the expected reduction in the observed bias can be explained by the level of agreement with the paradoxical thinking statements: the more the paradoxical thinking statements tend to induce disagreement the less bias can be observed and to test the fourth hypotheses, that is, whether the traditional IAT bias could be explained by self-reported negative attitudes toward anti-vax individuals and the feeling of threat that they may impose; We conducted these correlation analyses in Excel. To calculate the correlations between variables, we calculated the Pearson's correlation coefficient. We used Pearson's correlation coefficient because we measured the relationship between parametric values. Furthermore, we examined the potential outliers from both groups by calculating the average and standard deviation of the parameter: if the value was higher or lower than the group average ± 2*SD, then it was outlier. Accordingly, we examined whether the neural reduction would be explained by the agreement rating with each one of the paradoxical thinking stimuli. Pearson's correlation coefficient was calculated and revealed that (without an outlier exceeding 2 SD from the mean of the agreement measures) the correlation between the neural-level bias and the agreement with the paradoxical statements was significant ($r = -0.52$, $p = 0.037$). Noteworthy, with the outlier the correlation was not significant ($r = -0.42$, $p = 0.092$). The observed reduction in the neural bias can be explained by the level of agreement with the paradoxical statements: the more the paradoxical thinking statements induced disagreement, the less bias was observed. This can be interpreted that the more the stimuli induced identity threat, the more the neural prejudice response reduced.

Results

Participants on average tended to moderately agree with the explicit negative perception of vaccine-hesitant individuals equally across both intervention (scale 1–7 with 7 representing total agreement; mean ± SD 5.15 ± 1.09) and control (5.01 ± 1.15) groups ($p = .71$ for the between-group t-test). Furthermore, as a manipulation check, we found that the paradoxical thinking stimuli, unlike the control stimuli, did not trigger strong disagreement or resistance, in line with the paradoxical thinking principles (Bar-Tal et al., 2021). The group average for the control group was 1.11 (Rating 1 to 3, with 1 being disagree and 3 agree; mean ± SD 1.11 ± 0.11) and for the intervention group 2.04 (mean ± SD 2.04 ± 0.35).

During the MEG scan, as expected, participants responded significantly ($p = 0.002$) slower in the incongruent condition (1003.175 ms ±208.783) compared to the congruent condition (803.668 ms ±155.970) in the control group, thereby confirming the typical behavioral IAT effect – that an implicit bias was held toward vaccine-hesitant individuals. In the intervention group, this bias was also present, and there was no significant difference (p-value = 0.99, Cohen's $d = -0.002$) between the two groups on this implicit behavioral measure, thereby suggesting that the intervention did not yield any effect at the typical IAT behavioral level. At the neural level, we followed a previous MEG study that found an implicit neural IAT bias in the 8–10 Hz and 100–550 ms range (Levy et al., 2021). Our analysis replicated this perceptual intergroup bias by revealing a similar time-frequency (Figure 2A) and posterior (Figure 2B) pattern in the control group ($P_{cluster\text{-}cor} = 0.0009$) (Levy et al., 2021). Further, to examine whether the paradoxical thinking intervention reduced this neural bias, we examined the neural pattern in the intervention group (Figure 2A), and this analysis revealed a non-significant effect ($P_{cluster\text{-}cor} = 0.10$) thereby ruling out robust neural prejudice after the intervention. Importantly, this neural pattern was significantly more robust in the control compared to the intervention group (Figure 2A) when examining the all-sensor effect ($P_{cluster\text{-}cor} = 0.03$) and the posterior sensors ($P_{cluster\text{-}cor} = 0.002$). These findings suggest that participants in the control, but not in the intervention group, revealed a significant and robust neural prejudice toward the vaccine-hesitant individuals, and that the neural prejudice was significantly more robust in the control group. Hence, this result suggests that the paradoxical thinking attenuated neural implicit prejudice toward vaccine-hesitant individuals.

Finally, we examined whether the neural reduction would be explained by the agreement rating with each one of the paradoxical thinking stimuli. Pearson's correlation coefficient was calculated and revealed that (without the outlier) the correlation between the neural-level bias and the agreement with the paradoxical statements was significant ($r = -0.52$, $p = 0.037$). Noteworthy, with the outlier, the correlation was not significant ($r = -0.42$, $p = 0.092$). The observed reduction in the neural bias can be explained by the level of agreement with the paradoxical statements: the more the paradoxical thinking statements induced disagreement, the less bias was observed. This can be interpreted that the more the stimuli induced identity threat, the more the neural prejudice response is reduced. Furthermore, across all subjects, there was a significant positive correlation ($r = 0.37$, $p = 0.029$) between the behavioral level implicit bias (i.e., RT) and the self-reported negative attitudes toward the vaccine-hesitant individuals, thereby suggesting that the implicit negativity (at the behavioral level) can be explained by explicit negativity. Overall, the results reveal two separate mechanisms: (1) One by which participants explicitly relate negatively to vaccine-hesitant individuals; this mechanism is clearly expressed at the implicit behavioral level but is not affected by the paradoxical thinking intervention. By contrast, (2) the second mechanism is expressed as neural prejudice that is clearly reduced following the paradoxical thinking intervention, thereby suggesting a covert neural impact of the intervention.

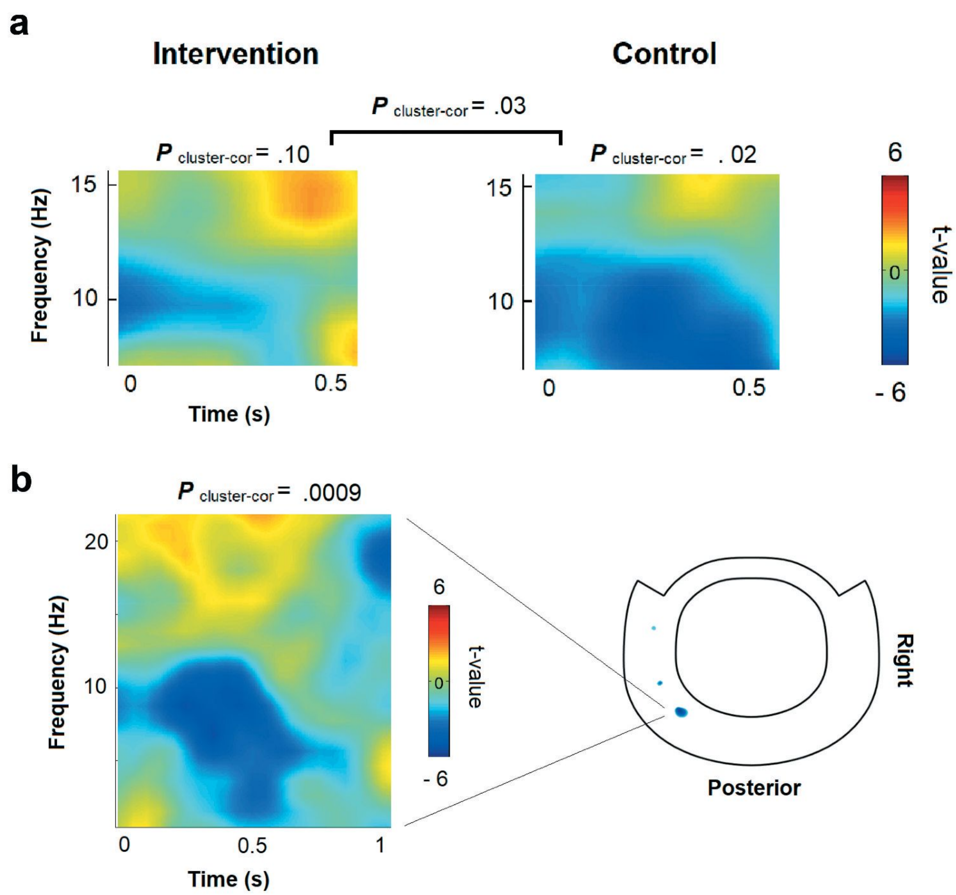

Figure 2. Neural representations implicit prejudice. (a) the alpha rhythm underpinning of implicit prejudice was found statistically significant in the control and statistically more robust than in the intervention group. (b) on the right is a topographical representation of the alpha representation in the control group, highlighting the implication of left posterior channels ($P_{cluster\text{-}cor} < 0.05$). On the left is the TFR representation of the induced activity emanating from those channels, thereby pointing out a clear alpha rhythmic activity ($P_{cluster\text{-}cor} = 0.00090$).

Discussion

The introduction of the COVID-19 vaccination has triggered skepticism, negativity, polarization, and societal division (Bajwa, 2021; Jiang et al., 2021) between the anti- and pro-vax individuals (Poddar et al., 2021). The present study aimed to address this timely societal issue: the negative intergroup attitudes that were recently triggered due to vaccination in the ongoing pandemic. The study represents two novel approaches in the study of prejudice and intergroup attitudes. First, building on a recent study that was conducted in the context of a well-entrenched intergroup conflict (Levy et al., 2021), we evaluate a neural response reflecting implicit prejudice. Second, we conducted an intervention to examine whether it can reduce this prejudice response; this is a very innovative strategy that has hardly been implemented at all in the study of intergroup interventions, and it can potentially uncover effects of the intervention that might otherwise get overlooked by the evaluation of conventional behavioral and self-reported measures.

However, before examining the neural implicit prejudice response, we first inspected the classical behavioral-level bias, based on response times. As expected, we found a strong behavioral bias toward the anti-vax individuals – 97.22% of participants revealed a statistically significant behavioral bias. Furthermore, we examined whether such bias could be explained by the self-reported negative attitudes toward anti-vax individuals and the feeling of threat that they may impose. A positive correlation between the behavioral-level bias and self-reported negative attitudes toward anti-vax individuals was revealed: the more biased the individual was at the behavioral level, the higher self-reported negativity was. Previous studies have examined how explicit and implicit racial attitudes of white people are related to impressions and behaviors in interactions with other-race people (Dovidio et al., 2002; Knutson et al., 2007). For example, the study by Dovidio et al. (2002), found that explicit racial attitudes that white people self-reported significantly predicted bias in verbal behavior toward other-race people, i.e., the outgroup (Dovidio et al., 2002).

Likewise, a study by Knutson et al. (2007) found that the explicit measure of racism correlated with the implicit IAT bias (Knutson et al., 2007). Hence, the findings reported here on the implicit bias and its association with explicit self-reports further consolidate the relationship between these constructs, at least under certain circumstances. Examining this in the context of the relations between anti- and pro-vax individuals is new and adds another empirical evidence to this literature, but this time in a new and timely context.

Then, we examined the neural implicit prejudice response based on a very recent study (Levy et al., 2021); our study reproduced this neural response. This is an important result that further support the validity of this novel neural response, and importantly, in a totally different context: while the previous study was conducted in the context of intractable intergroup conflict wherein attitudes and beliefs are socialized from an early age, the current study examined this neural response in a context in which new groups were formed ad hoc and only very recently. The ability to measure this neural response in such different intergroup contexts suggests that this response may reflect a basic mechanism of prejudice. Future studies that would examine this response in other contexts would enable to further evaluate the validity and implications of this interesting and new neural mechanism.

Importantly, to test the effect of the paradoxical intervention, we evaluated both the neural and the behavioral indices – in both experimental groups: the intervention and the control groups. The results of these analyses revealed a significant reduction of the neural response following the intervention. Interestingly, the behavioral response remained unchanged. As a corollary, an important question arises: why was the behavioral-level bias not reduced? First, it is noteworthy that using neural and behavioral measures to evaluate the impact of intergroup interventions is a very new direction of research and there is hardly any previous evidence that can be compared to for bringing up assumptions and forming interpretations of this differential effect. Nevertheless, one possible explanation for this difference between neural and behavioral effects may relate to the temporal dynamics of the two processes. Schiller and colleagues found that the latency delay reflecting the implicit behavioral effect was reflected in a neural delay right before button press, around 600 ms post-stimulus onset (Schiller et al., 2016). In other words, increased time delay in that late phase was associated with an increased implicit behavioral bias. By contrast, the neural response begins much earlier – already at 100 ms post-stimulus onset – and is not correlated with the behavioral bias – neither in our study nor in the previous one that measured this same response (Levy et al., 2021). It is therefore very plausible that the neural response, which begins very early in time, represents one process, whereas the behavioral response, which occurs much later in time – reflects a different process. Furthermore, it might not be a coincidence that the neural, but not the behavioral response was impacted by the intervention: the neural response correlated with the level of agreement with the paradoxical thinking items, therefore implying the attunement of the neural response with the interventional effect.

But what psychological processes do the two mechanisms reflect? Paradoxical thinking is thought to trigger unfreezing, which in turn, can lead to attitudes change, but not necessarily (Bar-Tal et al., 2021). People can reconsider their views and decide that they should stick with them. Hence, the differential effect reported in the present study might suggest that the intervention triggered unfreezing, as can be detected by the reduction of the neural response, but not the immediate downstream consequences, that is behavior change (possibly relating to attitudinal change). This interpretation further corroborates with previous behavioral studies of the paradoxical thinking intervention: In many cases, the effects were not assessments of intergroup attitudes but rather of unfreezing and openness (Hameiri et al., 2016, 2018; Knab et al., 2021). Moreover, in many of the cases in which there were effects on intergroup attitudes, these were not assessed immediately after the intervention, but only after some time had passed after the first exposure to it (Hameiri et al., 2014, 2016; Knab & Steffens, 2022). This new piece of empirical finding reported here is therefore valuable for the growing evidence pointing to the ability of neuroimaging data to quantitatively assess various cognitive and affective processes and to predict outcomes better than traditional behavioral and self-reported measures (Falk et al., 2015; Gabrieli et al., 2015). Future studies should further re-test this radically important assumption, and focus not only on immediate outcomes of interventions but also on long-lasting effects. This direction of research can transform the domain of intergroup interventions which exactly needs new approaches and long-term assessments (Paluck et al., 2021).

The current study has potential implications for both theory and practice on a timely societal matter. In particular, the study (i) further consolidates a new neural response reflecting implicit prejudice; (ii) it opens new vistas for examining both neural and behavioral impact of intergroup interventions and the mechanisms that each one of them may reflect; and (iii) it can have societal implications for moderating the polarized attitudes in a new era of pandemics. Nevertheless, it is important to note that the generalizability of the findings reported here requires further research. Although the sample of participants in this study relies on a priori sample size calculation, it is noteworthy that the hypotheses and experimental contexts in the current study (i.e., vaccine hesitancy in Finland) and in that reported by Levy and colleagues (i.e., violent ethnic conflict in Israel) (Levy et al., 2021) are very different. Hence, more studies are needed in the future with larger sample sizes, in other societies, and with additional controls such as neutral-outcome controls (e.g., no intervention). It is our hope that the current study will motivate such future work as this new direction of research is very relevant and timely with the prolonged COVID-19 pandemic and with its long-term consequences on many levels of the global society.

Disclosure statement

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

Funding

The work was supported by the Kulttuurin ja Yhteiskunnan Tutkimuksen Toimikunta [328674].

ORCID

Jonathan Levy http://orcid.org/0000-0003-1007-2393

References

Adler, E., Hebel, S., Lesehm, O. A., Levy, J., & Halperin, E. (2021). A social virus: Intergroup dehumanization and unwillingness to aid amidst COVID-19 – Who are the main targets?. *International Journal of Intercultural Relations*, *86*, 109–121. https://doi.org/10.1016/j.ijintrel.2021.11.006

Amodio, D. M., & Cikara, M. (2021). The social neuroscience of prejudice. *Annual Review of Psychology*, *72*(1), 439–469. https://doi.org/10.1146/annurev-psych-010419-050928

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*(4), 652–661. https://doi.org/10.1037/0022-3514.91.4.652

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*, *20*(3), 143–148. https://doi.org/10.1177/0963721411408562

Badea, C., & Sherman, D. K. (2019). Self-Affirmation and prejudice reduction: When and why?. *Current Directions in Psychological Science*, *28*(1), 40–46. https://doi.org/10.1177/0963721418807705

Bajwa, A. (2021). Information disorder, the Triumvirate, and COVID-19: How media outlets, foreign state intrusion, and the far-right diaspora drive the COVID-19 anti-vaccination movement. *The Journal of Intelligence, Conflict, and Warfare*, *4*(2), 2. https://doi.org/10.21810/jicw.v4i2.3067

Bar-Tal, D., Hameiri, B., & Halperin, E. (2021). Paradoxical thinking as a paradigm of attitude change in the context of intractable conflict. In B. Gawronski (Ed.), *Advances in experimental social psychology* (pp. 129–187). Elsevier Academic Press.

Bartunek, J. M. (1993). The multiple cognitions and conflicts associated with second order organizational change. In John Keith Murnighan (Ed.), *Social psychology in organizations: Advances in theory and research*. Prentice Hall.

Bruneau, E. G., Kteily, N. S., & Urbiola, A. (2020). A collective blame hypocrisy intervention enduringly reduces hostility towards Muslims. *Nature Human Behaviour*, *4*(1), 45–54. https://doi.org/10.1038/s41562-019-0747-7

Chang, L. W., Krosch, A. R., & Cikara, M. (2016). Effects of intergroup threat on mind, brain, and behavior. *Current Opinion in Psychology*, *11*, 69–73. https://doi.org/10.1016/j.copsyc.2016.06.004

Cohn, S. K. (2012). Pandemics: Waves of disease, waves of hate from the Plague of Athens to A.I.D.S. *Historical Research*, *85*(230), 535–555. https://doi.org/10.1111/j.1468-2281.2012.00603.x

Cooper, J. B., & Siegel, H. E. (1956). The galvanic skin response as a measure of emotion in prejudice. *The Journal of Psychology*, *42*(1), 149–155. https://doi.org/10.1080/00223980.1956.9713031

Crandall, C. S., & Eshleman, A. (2005). The justification-suppression model of prejudice: An approach to the history of prejudice research. In Crandall, C.S., Schaller, M. (Eds.), *Social psychology of prejudice: Historical and contemporary issues/edited by Christian S. Crandall, Mark Schaller*. *Lawrence, Kan.: Lewinian Press*, *c2005*, (pp. 237–268). Lewinian Press.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Dovidio, J. F., & Gaertner, S. L. (1986). *Prejudice, discrimination, and racism*. Academic Press.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68. https://doi.org/10.1037/0022-3514.82.1.62

Falk, E. B., O'Donnell, M. B., Cascio, C. N., Tinney, F., Kang, Y., Lieberman, M. D., Taylor, S. E., An, L., Resnicow, K., & Strecher, V. J. (2015). Self-affirmation alters the brain's response to health messages and subsequent behavior change. *Proceedings of the National Academy of Sciences*, *112*(7), 1977–1982. https://doi.org/10.1073/pnas.1500247112

Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fuochi, G., Boin, J., Voci, A., & Hewstone, M. (2021). COVID-19 threat and perceptions of common belonging with outgroups: The roles of prejudice-related individual differences and intergroup contact. *Personality and Individual Differences*, *175*, 110700. https://doi.org/10.1016/j.paid.2021.110700

Gabrieli, J. D. E., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, *85*(1), 11–26. https://doi.org/10.1016/j.neuron.2014.10.047

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Hameiri, B., Idan, O., Nabet, E., Bar-Tal, D., & Halperin, E. (2020). The paradoxical thinking 'Sweet Spot': The Role of recipients' latitude of rejection in the effectiveness of paradoxical thinking messages targeting anti-refugee attitudes in Israel. *Journal of Social and Political Psychology*, *8*(1), 1. https://doi.org/10.5964/jspp.v8i1.1158

Hameiri, B., Nabet, E., Bar-Tal, D., & Halperin, E. (2018). Paradoxical thinking as a conflict-resolution intervention: comparison to alternative interventions and examination of psychological mechanisms. *Personality & Social Psychology Bulletin*, *44*(1), 122–139. https://doi.org/10.1177/0146167217736048

Hameiri, B., Porat, R., Bar-Tal, D., Bieler, A., & Halperin, E. (2014). Paradoxical thinking as a new avenue of intervention to promote peace. *Proceedings of the National Academy of Sciences*, *111*(30), 10996–11001. https://doi.org/10.1073/pnas.1407055111

Hameiri, B., Porat, R., Bar-Tal, D., & Halperin, E. (2016). Moderating attitudes in times of violence through paradoxical thinking intervention. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(43), 12105–12110. https://doi.org/10.1073/pnas.1606182113

Han, B., Ho, B., Rhee, I., & Tabakis, C. (2021). *Fear and favoritism in the time of COVID-19* (SSRN Scholarly Paper ID 3914606). Social Science Research Network. https://doi.org/10.2139/ssrn.3914606

Hasler, B. S., Landau D, H., Hasson, Y., Schori-Eyal, N., Giron, J., Levy, J., Halperin, E., & Friedman, D. (2021). Virtual reality-based conflict resolution: The impact of immersive 360° video on changing view points and moral judgment in the context of violent intergroup conflict. *New Media & Society*, *23*(8), 2255–2278. https://doi.org/10.1177/1461444821993133

Hasson, Y., Schori-Eyal, N., Landau, D., Hasler, B. S., Levy, J., Friedman, D., Halperin, E., & Triberti, S. (2019). The enemy's gaze: Immersive virtual environments enhance peace promoting attitudes and emotions in violent intergroup conflicts. *PLoS One*, *14*(9), e0222342. https://doi.org/10.1371/journal.pone.0222342

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup Bias. *Annual Review of Psychology*, *53*(1), 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality & Social Psychology Bulletin*, *31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Jiang, X., Su, M. -H., Hwang, J., Lian, R., Brauer, M., Kim, S., & Shah, D. (2021). Polarization over vaccination: Ideological Differences in twitter expression about COVID-19 vaccine favorability and specific hesitancy concerns. *Social Media + Society*, *7*(3), 20563051211048412. https://doi.org/10.1177/20563051211048413

Knab, N., & Steffens, M. C. (2022). "Why do you think Christmas will never ever be celebrated again?" a paradoxical thinking intervention's potential to affect conflict-related concerns, willingness to compromise, and openness towards refugees. *European Journal of Social Psychology*, *52*(3), 2833. https://doi.org/10.1002/ejsp.2833

Knab, N., Winter, K., & Steffens, M. C. (2021). Flexing the extremes: Increasing Cognitive flexibility with a paradoxical leading questions intervention. *Social cognition*, *39*(2), 225–242. https://doi.org/10.1521/soco.2021.39.2.225

Knutson, K. M., Mah, L., Manly, C. F., & Grafman, J. (2007). Neural correlates of automatic beliefs about gender and race. *Human brain mapping*, *28*(10), 915–930. https://doi.org/10.1002/hbm.20320

Kurdi, B., Seitchik, A., Axt, J., Carroll, T., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A., & Banaji, M. (2018). Relationship between the implicit association test and intergroup behavior: A meta-analysis. American Psychologist, 74(5), 569–586. https://doi.org/10.1037/amp0000364

Lades, L. K., Laffan, K., Daly, M., & Delaney, L. (2020). Daily emotional well-being during the COVID-19 pandemic. *British Journal of Health Psychology*, *25*(4), 902–911. https://doi.org/10.1111/bjhp.12450

Levy, J., Goldstein, A., & Feldman, R. (2019). The neural development of empathy is sensitive to caregiving and early trauma. *Nature communications*, *10*(1), 1. https://doi.org/10.1038/s41467-019-09927-y

Levy, J., Goldstein, A., Influs, M., Masalha, S., & Feldman, R. (2021). Neural rhythmic underpinnings of intergroup bias: Implications for peace-building attitudes and dialogue. *Social Cognitive And Affective Neuroscience*, *17*(4), nsab106. https://doi.org/10.1093/scan/nsab106

Levy, J., Goldstein, A., Pratt, M., & Feldman, R. (2018). Maturation of pain empathy from child to adult shifts from single to multiple neural rhythms to support interoceptive representations. *Scientific Reports*, *8*(1), 1. https://doi.org/10.1038/s41598-018-19810-3

Levy, J., Influs, M., Masalha, S., Goldstein, A., Feldman, R., & Nelson, K. E. (2022). Dialogue intervention for youth amidst intractable conflict attenuates neural prejudice response and promotes adults' peacemaking. *PNAS Nexus*, *1*(5), pgac236. https://doi.org/10.1093/pnasnexus/pgac236

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. https://doi.org/10.1037/a0032734

Paluck, E. L. (2016). How to overcome prejudice. *Science*, *352*(6282), 147. https://doi.org/10.1126/science.aaf5207

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, *72*(1), 533–560. https://doi.org/10.1146/annurev-psych-071620-030619

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and*

Social Psychology, *75*(3), 811–832. https://doi.org/10.1037/0022-3514.75.3.811

Poddar, S., Mondal, M., Misra, J., Ganguly, N., & Ghosh, S. (2021). Winds of change: Impact of COVID-19 on Vaccine-related opinions of twitter users. *ArXiv:2111.10667 [Cs]*. http://arxiv.org/abs/2111.10667

Poldrack, R. A., Monahan, J., Imrey, P. B., Reyna, V., Raichle, M. E., Faigman, D., & Buckholtz, J. W. (2018). Predicting violent behavior: What can neuroscience add? *Trends in cognitive sciences*, *22*(2), 111–123. https://doi.org/10.1016/j.tics.2017.11.003

Rankin, R. E., & Campbell, D. T. (1955). Galvanic skin response to Negro and white experimenters. *Journal of Abnormal and Social Psychology*, *51*(1), 30–33. https://doi.org/10.1037/h0041539

Schiller, B., Gianotti, L. R. R., Baumgartner, T., Nash, K., Koenig, T., & Knoch, D. (2016). Clocking the social mind by identifying mental processes in the IAT with electrical neuroimaging. *Proceedings of the National Academy of Sciences*, *113*(10), 2786–2791. https://doi.org/10.1073/pnas.1515828113

Schuster, M., Eskola, J., & Duclos, P. (2015). Review of vaccine hesitancy: Rationale, remit and methods. *Vaccine*, *33*(34), 4157–4160. https://doi.org/10.1016/j.vaccine.2015.04.035

Swann, W. B., Pelham, B. W., & Chidester, T. R. (1988). Change through paradox: Using self-verification to alter beliefs. *Journal of Personality and Social Psychology*, *54*(2), 268–273. https://doi.org/10.1037/0022-3514.54.2.268

Tallon-Baudry, C., Bertrand, O., Tallon-Baudry, C., Bertrand, O., Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences*, *3*(4), 151–162. https://doi.org/10.1016/S1364-6613(99)01299-1

Tram, K. H., Saeed, S., Bradley, C., Fox, B., Eshun-Wilson, I., Mody, A., & Geng, E. (2022). Deliberation, dissent, and distrust: understanding distinct drivers of coronavirus disease 2019 vaccine hesitancy in the United States. *Clinical Infectious Diseases*, *74*(8), 1429–1441. https://doi.org/10.1093/cid/ciab633

Wilson, M. C., & Scior, K. (2014). Attitudes towards individuals with disabilities as measured by the implicit association test: A literature review. *Research in Developmental Disabilities*, *35*(2), 294–321. https://doi.org/10.1016/j.ridd.2013.11.003

Zebarjadi, N., Adler, E., Kluge, A., Jääskeläinen, I. P., Sams, M., & Levy, J. (2021). Rhythmic Neural patterns during empathy to vicarious pain: beyond the affective-cognitive empathy dichotomy. *Frontiers in human neuroscience*, *15*, 380. https://doi.org/10.3389/fnhum.2021.708107