Social Cognitive and Affective Neuroscience, 2021, 117–128

doi:10.1093/scan/nsaa115
Advance Access Publication Date: 7 October 2020
Original Manuscript

Interpersonal Synchrony Special Issue

Neural reference groups: a synchrony-based classification approach for predicting attitudes using fNIRS

Macrina C. Dieffenbach,[1] Grace S. R. Gillespie,[1] Shannon M. Burns,[1] Ian A. McCulloh,[2] Daniel L. Ames,[1] Munqith M. Dagher,[3] Emily B. Falk,[4] and Matthew D. Lieberman[1]

[1]Annenberg School of Communication, University of Pennsylvania, Philadelphia, Philadelphia, PA 19104, USA, [2]Accenture Federal Services, 800 N Glebe Rd, Arlington, VA 22203, [3]Independent Institute & Administration Civil Society Studies (IIACSS) Research Group, Al Hussam Center 2 270 Arar Mustafa Wahbii Al Tal, Amman, Jordan and [4]Annenberg School of Communication, University of Pennsylvania, Philadelphia, PA 19104, USA, Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA, Wharton Marketing Department, University of Pennsylvania, Philadelphia, PA 19104, USA, University of Pennsylvania

Correspondence should be addressed to Macrina C. Dieffenbach, Department of Psychology, University of California – Los Angeles, 4611 Pritzker Hall, Los Angeles, CA 90095, USA. E-mail: macrina.dieffenbach@gmail.com

Abstract

Social neuroscience research has demonstrated that those who are like-minded are also 'like-brained.' Studies have shown that people who share similar viewpoints have greater neural synchrony with one another, and less synchrony with people who 'see things differently.' Although these effects have been demonstrated at the 'group level,' little work has been done to predict the viewpoints of specific 'individuals' using neural synchrony measures. Furthermore, the studies that have made predictions using synchrony-based classification at the individual level used expensive and immobile neuroimaging equipment (e.g. functional magnetic resonance imaging) in highly controlled laboratory settings, which may not generalize to real-world contexts. Thus, this study uses a simple synchrony-based classification method, which we refer to as the 'neural reference groups' approach, to predict individuals' dispositional attitudes from data collected in a mobile 'pop-up neuroscience' lab. Using functional near-infrared spectroscopy data, we predicted individuals' partisan stances on a sociopolitical issue by comparing their neural timecourses to data from two partisan neural reference groups. We found that partisan stance could be identified at above-chance levels using data from dorsomedial prefrontal cortex. These results indicate that the neural reference groups approach can be used to investigate naturally occurring, dispositional differences anywhere in the world.

Key words: neural reference groups; neural synchrony; intersubject correlation; fNIRS; dmPFC

When people share similar ideas and opinions, they are often referred to as being 'like-minded.' In support of this metaphor, recent research demonstrates that people show greater neural synchrony (i.e. correlated neural fluctuations over time) with others who hold similar psychological perspectives and less neural synchrony with those who 'see' things differently. Thus,

Received: 3 February 2020; **Revised:** 19 June 2020; **Accepted:** 29 September 2020

studies have also identified distinguishable neural signatures between people who hold different perspectives at the group level (Nummenmaa *et al.*, 2018). Taking this idea one step further, recent studies have also shown that it is possible to predict the perspective that particular 'individuals' hold by comparing the amount of synchrony they show with groups of people who hold one perspective versus another, and then classifying them into whichever group they more closely resemble (Lahnakoski *et al.*, 2014; Yeshurun *et al.*, 2017). These studies applied synchrony-based classification approaches to predict differing mindsets that were experimentally induced. However, no published research has yet attempted to use a synchrony-based approach to predict naturally occurring, dispositional differences (i.e. longstanding psychological characteristics).

Furthermore, the synchrony-based classification studies conducted to date used functional magnetic resonance imaging (fMRI), which is expensive and immobile. Given that MRI machines are located in limited areas of the world (e.g. urban and mostly western locations), this imaging modality can only reach certain populations, which limits its generalizability and potential to study particular populations of interest. Thus, more work is needed to determine whether the same classification-based approaches used in the fMRI literature can be applied to data collected from portable neuroimaging devices, which are able to reach a broader population (Burns clustering to identify small group, 2019).

Therefore, in this study, we used a simple synchrony-based classification method, which we refer to as the 'neural reference groups' approach, to predict dispositional attitudes at the individual level. Furthermore, we applied this method to neural time series data collected using functional near infrared spectroscopy (fNIRS), a portable neuroimaging device. This research was conducted in the Middle East to demonstrate the possibility of conducting simple, naturalistic viewing studies anywhere in the world and also the feasibility of analyzing their data using a computationally accessible classification method.

The neural reference groups approach involves comparing an individual's brain data to data from groups of people with pre-identified distinct mindsets, and then 'matching' the individual into the group with which they have greater neural synchrony. Neural synchrony analyses were first developed to localize universal cognitive processes that occur during the processing of naturalistic stimuli. For instance, intersubject correlation is a neural synchrony approach that is commonly used for understanding which regions and networks of the brain are active across individuals during narrative comprehension (Hasson *et al.*, 2004; Nastase *et al.*, 2019). Such work has demonstrated strong synchronization in both low-level sensory regions and high-level association cortices, suggesting that individuals show similarities in their processing of both low- and high-level information features (Hasson *et al.*, 2004, 2010). Furthermore, regardless of the modality in which a narrative is presented, comprehension of its content tends to be associated with activation in the brain's default mode network (DMN; Wilson *et al.*, 2007; Jääskeläinen *et al.*, 2008; Honey *et al.*, 2012; Regev *et al.*, 2013).

Research using the intersubject correlation approach has also examined how neural responses differ across individuals who 'see things differently,' or are interpreting the same stimuli according to different frameworks. For instance, when individuals are told to attend to different aspects of a scene (e.g. scenery versus plot) while watching a movie, they show distinguishable differences in regions associated with attention and the processing of objects and scenes (parahippocampal gyrus, posterior parietal cortex and lateral occipital cortex; Lahnakoski *et al.*, 2014), such that people sharing a perspective show greater synchrony than those asked to see things differently. Further, individuals who are given alternative frames for interpreting an ambiguous narrative show differential neural responding in the brain's mentalizing network, language areas and subsets of the mirror neuron system (Yeshurun *et al.*, 2017).

Building on the group differences that they identified, these studies also applied classification-based machine learning and could reliably distinguish between individuals who interpreted the same information through two different frameworks. Lahnakoski *et al.* (2014) use a k-nearest neighbors machine learning approach, classifying participants based on the group membership of the participants with whom they show the greatest synchrony. In contrast, Yeshurun *et al.* (2017) use a k-nearest centroid approach, in which participants are classified based on showing greater synchrony with the average of one group of participants versus another. In this article, we refer to the approach used by Yeshurun *et al.* (2017) as the neural reference groups approach. This approach is simple to implement computationally and requires making few analytic choices, thus limiting 'researcher degrees of freedom' (Botvinik-Nezer *et al.*, 2020). In addition, it involves comparing new participants' data to group average timecourses, which are less noisy references for classification than neighboring individuals' timecourses.

Whereas these studies looked at experimentally manipulated differences in perspective, other research has examined how naturally occurring, dispositional differences influence neural synchrony (Finn *et al.*, 2020). For instance, researchers found that individuals with similar levels of trait paranoia (high or low) showed more similar neural responding in regions of the DMN (Finn *et al.*, 2018). Other researchers have found that individuals with similar sexual desire and self-control preferences have similar neural fluctuations in several brain networks, including the DMN (Chen *et al.*, 2020). Furthermore, individuals with the same cognitive style (analytical or holistic thinking) show synchrony in several cortical regions, including prefrontal cortex (Bacha-Trams *et al.*, 2018). Finally, other studies of have also found a strong relationship between similarities in self-reported experiences of narratives and neural responses (Jääskeläinen *et al.*, 2008; Nummenmaa *et al.*, 2012; Nguyen *et al.*, 2019; Saalasti *et al.*, 2019; Tei *et al.*, 2019).

Although this nascent body of research has examined the neural correlates of individual differences, no researchers have used a classification approach to make predictions about the dispositions of specific individuals using neural synchrony measures. From a basic science perspective, classification-based analyses have the advantage of being driven by reverse-inference rather than forward-inference, drawing a stronger link between brain activity and particular psychological functionality (Poldrack, 2011). From an applied science perspective, classification-based research can move beyond simply explaining differences in dispositional experience (i.e. what traditional, forward-inference studies do) to actually make real-world predictions about individuals whose dispositional characteristics are not known in advance.

To be clear, there is also significant literature on how differences in dispositional tendencies are associated with different neural responses to short, repeatable events (in contrast to more naturalistic timecourse data). For instance, many studies have shown that liberals and conservatives show differential neural responding in a number of regions, including the DMN, dorsolateral prefrontal cortex (dlPFC), anterior cingulate, amygdala and insula (Knutson *et al.*, 2006; Westen *et al.*, 2006; Kaplan *et al.*, 2007; Kanai *et al.*, 2011; Jost and Amodio, 2012; Ahn *et al.*, 2014; Van Bavel and Pereira, 2018). Other studies have

applied machine learning to univariate data to make predictions about other real-world characteristics, including physical and psychological well-being (Memarian et al., 2017) and political orientation (Ahn et al., 2014). Although these studies have been useful in illuminating naturally occurring differences in brain functioning, their use of event-based paradigms limits the ecological validity of their findings. In contrast, measuring brain fluctuations during unstructured experiences, such as watching a video or having a conversation, yields findings that are more likely to be generalizable to real-world experience. Furthermore, these naturalistic paradigms are simple to design and conduct, which is useful in terms of being able to use them to study a wide range of dispositional differences in a variety of contexts.

In summary, previous synchrony-based studies have taken a forward-inference approach, showing that individuals who share similar traits also show similar neural responses. Two synchrony studies to date have taken a reverse-inference approach to predict participants' temporary mindsets, which were experimentally induced, based on their neural fluctuations. The only studies that have made predictions about naturally occurring, 'dispositional' differences have been event-based, which can be limited in terms of their generalizability. Thus, there have been no classification-based synchrony studies that attempt to use naturalistic timecourse data to predict individuals' dispositional tendencies to process or experience the world differently. Furthermore, most research using a classification-based approach to predict dispositional tendencies has been conducted in highly controlled laboratory settings using fMRI, which is costly and limited in terms of the populations it can reach. Although fMRI research has been important in advancing classification-based methods, further work is needed to demonstrate the efficacy of conducting classification analyses on data acquired in more naturalistic, real-world settings. Therefore, we set out to examine whether it was possible to use a synchrony-based classification approach on neural timecourse data acquired in a non-standard lab setting using fNIRS, which is a less expensive and more portable neuroimaging modality than fMRI. Furthermore, we attempted to do so in a 'non-WEIRD population' in the Middle East, an area of the world in which neuroscience studies are rarely conducted outside of Israel (Burns et al., 2019).

The present study

In this study, our goal was to predict individuals' dispositional attitudes on a sociopolitical topic in a pop-up lab that was set up in an office space in Amman, Jordan. Given that attitudes can serve as interpretive frames that affect attention, mentalizing, counterarguing and other cognitive processes, we predicted that individuals with different attitudes should show differential neural responding in regions associated with these processes (i.e. lateral prefrontal cortex [lPFC] and medial prefrontal cortex [mPFC]). If this is the case, then it is possible to create neural reference group data by averaging across neural timecourses from the same brain region in participants who share similar attitudes or other hidden psychological characteristics. When two or more neural reference groups are obtained, new individuals whose attitudes or characteristics are not already known can be classified into one of the groups by comparing whether they show greater synchrony with one group versus another. In other words, two groups of people who have different attitudes about, for example, abortion, are likely to have different neural responses when listening to an anti-abortion message. A new individual listening to the same message will reveal greater similarity to one group (e.g. the pro-choice group) than to the other (e.g. pro-life group), indicating whether the new individual is likely to be pro-choice or pro-life. In tests of such classification strategies, the true dispositional attitude of the 'new individual' is actually known, but the classification process is blind to this information and only compared to this criterion in the final step to determine the accuracy of the classification method.

Only one other known study has used this neural reference groups method, predicting the experimentally manipulated perspective from which participants were understanding a narrative (Yeshurun et al., 2017). The present study was a first test of this method on dispositional attitudinal differences. Participants in the Middle East who held opposing views on a sociopolitical issue came to a pop-up neuroscience lab and viewed two videos in which other individuals expressed their opinions about the issue. While watching the videos, participants were scanned using fNIRS. Data were collected from channels positioned in lPFC and mPFC regions. Lateral prefrontal regions were selected due to previous associations of dlPFC with counterarguing behavior (O'Donnell et al., 2018; Liu et al., 2020). As part of the DMN, mPFC was selected due its association with social cognitive processes: A large body of evidence suggests that ventromedial cortex is associated with affective processing, anteromedial prefrontal cortex with self-referential thinking and dorsomedial cortex (dmPFC) with mentalizing and judgments about others, (Lieberman et al., 2019). Furthermore, prior work has shown that dmPFC synchrony can detect when individuals have more similar spontaneous interpretations of a narrative (Finn et al., 2018; Nguyen et al., 2019). Finally, collecting data from mPFC and lPFC regions minimized the chance of signal drop-out, as they are conveniently located beneath areas of the scalp that have less hair (i.e. the forehead).

We conducted analyses in two stages to determine whether members of the opposing ideological groups showed differentiable neural responses to the videos. First, we examined whether there were group-level differences. On a channel-by-channel basis, we averaged across the neural timecourses of all members within each ideological group, which created two group average timecourses per channel. We then conducted Euclidean distance analyses between these average timecourses to detect group-level differences. We hypothesized that we would find group differences between the timecourses of the two neural reference groups.

Second, we used the neural reference groups approach to make predictions about ideological stance at the individual level. The neural reference groups approach utilizes a leave-two-out procedure: the timecourses from pairs of participants are 'left out' from the dataset and are then compared to the timecourses of each neural reference group formed from the remaining data. Participants were classified as holding one ideological stance or the other based on which neural reference group their neural responses more closely resembled (i.e. which group they showed greater synchrony with). This process was repeated, holding out a different pair of participants in each iteration, until all participants have received predictions. In order to assess the accuracy of the neural reference groups approach, participants' true attitudes were compared to the model's predictions. Given that individuals who hold different ideological stances are likely to process sociopolitical content differentially, we hypothesized that we would be able to accurately predict participants' stances at the individual level.

Method

Participants

Participants ($N = 72$) were adult males who were recruited in Amman Jordan, for a video marketing study, from which the

authors obtained the data for analysis. All participants were screened over the phone in Arabic and were asked for their consent to participate. Total sample size was determined by how many participants could be scanned with the resources and time allotted to collecting data in a 10-day timespan. Participants were recruited such that half of the sample would hold one political stance and half would hold the opposite stance ($n = 36$ for each group). During pre-screening, participants used a 7-point scale (1 = 'strongly disagree,' 7 = 'strongly agree') to indicate their agreement with the following statement: 'Women who are raped should be allowed to have abortions.' This item was developed by the research team to assess attitudes on a facet of the abortion debate that was salient to the population being studied. In this article, we will refer to those in support of this sub-issue of abortion as being pro-choice and those who oppose it as pro-life, though the reader should consider that these terms are simplifications of a complex issue and, importantly, do not correspond directly to pro-choice and pro-life views as they are often defined in Western countries. Individuals who answered between 1 and 3 on the scale above were classified as being pro-life, and individuals who answered between 5 and 7 were classified as being pro-choice. Individuals who answered a '4' were not admitted into the study. For the final sample of participants who completed the study, the average opinion for pro-choice group members was a 6.47 ($SD = 0.71$) on the scale, whereas the opinion for pro-life group members was 1.67 ($SD = 0.80$).

Procedure

Participants came into an office space at the Independent Institute & Administration Civil Society Studies Research Group polling firm, where a pop-up fNIRS laboratory had been set up. After providing consent, participants' heads were measured and then fitted with an appropriately sized stretchy cap, which held the fNIRS optodes against the skull. The fNIRS equipment was then calibrated to ensure good signal quality between sources and detectors. During the fitting and calibration process, participants completed a questionnaire to assess their attitudes toward the abortion issue. This questionnaire included the original pre-screening item (i.e. whether women who are raped should be allowed to have abortions), which was used to confirm the participant's ideological stance on the day of the scan. The questionnaire also included a question that assessed whether participants thought abortion should be allowed in a series of different circumstances ('Do you agree or disagree with each of the following reasons for having an abortion?') For this question, participants rated a series of items, answering 'Agree,' 'Disagree' or 'No Opinion.' This question was included as a nuanced attitude measure for the purposes of tracking attitude change over time, although it was not analyzed in this study.

Next, participants completed the scanning portion of the study. During scanning, participants watched two 4- to 5-minute YouTube-style videos of Arabic speakers discussing their stance on the abortion issue in 2 separate functional runs. The order of the videos was counterbalanced across participants. The speaker in one video expressed a pro-choice stance, and the other expressed a pro-life stance. Scripts for the videos were written by the research team, translated into Arabic and then recorded by actors. After watching each video, participants completed a questionnaire in which they evaluated the quality of the speaker's arguments using a subset of items that were adapted from a validated scale of perceived argument strength (Zhao *et al.*, 2011). Participants used a Likert scale to indicate the extent to which they agreed with the following (translated) questions (1 = 'strongly disagree,' 3 = 'neither agree nor disagree,' 5 = 'strongly agree'): 'The person in the video gives convincing reasons for [increasing access to/preventing] abortion for women who are raped' and 'The reasons provided in the video are strong for [increasing access to/preventing abortion] for women who are raped.' Following the video portion of the scan, participants completed two functional localizers, which were translated into Arabic: the 'Why-How task,' a well-validated localizer of the brain's mentalizing system (Spunt and Adolphs, 2014), and a 'counter-arguing task' developed by our team (O'Donnell, in prep). The data from these localizer tasks were not used in the present analyses.

Data analysis

fNIRS acquisition and pre-processing

Acquisition. Participants were scanned using two NIRSport fNIRS units (NIRx, Los Angeles, CA), with a layout of 20 channels, composed of 8 light sources and 7 detectors (Figure 1). The NIRSport systems were selected due to their portability and compact size, as the machines were transported in carry-on luggage from the United States to Jordan and back. The layout was standardized using the 10-10 UI external positioning system. Channels were placed in medial and lateral prefrontal areas, which are associated with mentalizing (mPFC) and counterarguing (dlPFC) processes (Denny *et al.*, 2012; O'Donnell *et al.*, in prep). Data were collected at a sampling rate of 7.81 Hz at wavelengths of 760 and 850 nm. Given this high sampling rate, the timecourses for each video consisted a large number of timepoints (2195 for the pro-choice video, and 2531 for the pro-life video).

Pre-processing. Prior to data pre-processing, participants were excluded from all analyses if their answers on the primary attitudinal pre-screening question, indicated they had a neutral political stance when it was re-administered on the day of the scanning session (i.e. 4 on the 7-point scale; $n = 2$ participants recruited as pro-life). Participants were also excluded if their stance on the day of the scanning session conflicted with the stance they had been assigned during pre-screening ($n = 2$ recruited pro-life, $n = 1$ recruited as pro-choice). Participants were also excluded from analyses on a video-by-video basis if technical issues occurred during acquisition for that video ($n = 3$ pro-choice watching the pro-choice video; $n = 2$ pro-life watching pro-choice; $n = 2$ pro-life watching pro-life; $n = 2$ pro-choice watching pro-life). Following these exclusions, the following sample sizes remained for each political group watching each video type: $n = 32$ pro-choice Ps watching pro-choice videos, $n = 30$ pro-life Ps watching pro-choice videos, $n = 33$ pro-choice Ps watching pro-life videos and $n = 30$ pro-life Ps watching pro-life videos.

The remaining data were pre-processed using a customized fNIRS pre-processing pipeline that utilizes the HOMER2 analysis package (Huppert *et al.*, 2009). For each scan, data channels were marked as having usable signal if detector saturation did not occur for longer than 2 seconds at a time, and if the variation of the signal's power spectrum did not exceed a quartile coefficient of dispersion of 0.1 over the course of the scan. Then, the raw NIRS data were filtered using a bandpass filter of 0.005–0.5 Hz and corrected for motion artifacts using a PCA algorithm, converted into hemoglobin concentrations using the Modified Beer–Lambert Law, and then z-scored. Timecourses were truncated prior to the analyses,

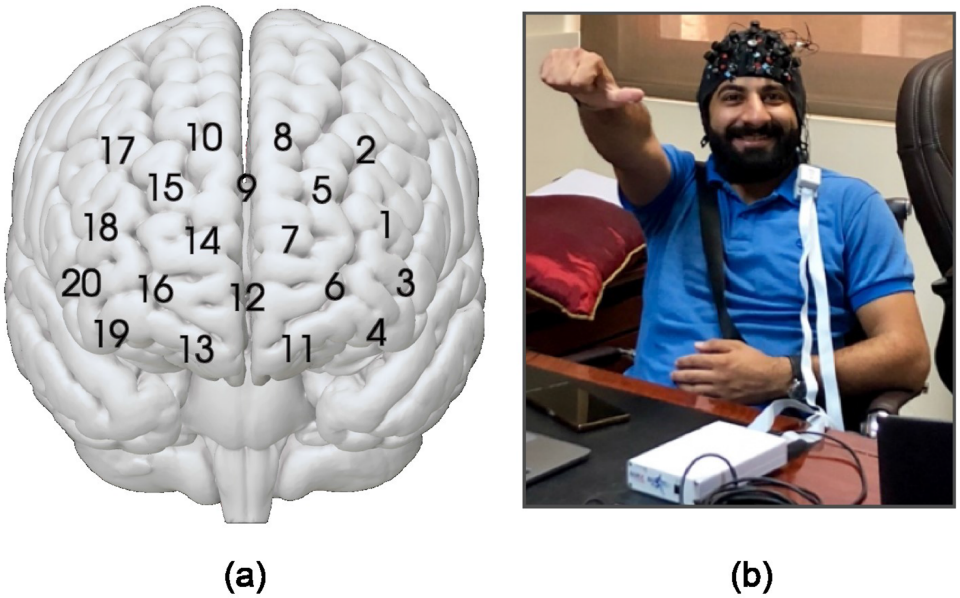

Fig. 1. (a) Locations of 20 NIRS channels, which are formed between adjacent sources and detectors. (b) Experimental setup, showing participant fitted with fNIRS cap in the mobile laboratory, which was established in a market research company's office space.

which included trimming off any scan time that occurred before or after the stimuli were displayed and removing the first 12 seconds of scan time during the video to account for delay in the hemodynamic response function. Analyses were conducted on oxygenated hemoglobin in accordance with our lab's prior work (Burns *et al.*, 2018, 2019). Research has shown that oxygenated hemoglobin has a stronger signal-to-noise ratio compared to deoxygenated hemoglobin (Strangman *et al.*, 2002). Furthermore, the oxygenated hemoglobin signal is more closely correlated with the fMRI BOLD signal (Cui *et al.*, 2011), which was relevant given that this study was replicating a method conducted on fMRI data (Yeshurun *et al.*, 2017).

In order to localize the data within a common brain space such that the present results could be compared with results from fMRI studies, approximate MNI coordinates were identified for each 10-10 channel position using a probabilistic registration method (Singh *et al.*, 2005). For visualization purposes, NIRS data were converted to *.img files using xjView (http://www.alivelearn.net/xjview/), and then overlaid on a 3D cortical surface using the software Surf Ice.

Measuring group-level neural differences

As a first analysis step, we examined whether participants in the pro-life and pro-choice groups showed distinguishable differences in their neural responses to the videos. We conducted this analysis on a channel-by-channel basis and for each video separately. First, we created average timecourses for each attitudinal group by calculating the mean across participants within a group at each timepoint (t). Then, to test for differences between the groups, we computed the Euclidean distance between the group average timecourses using the following formula:

$$D = \sqrt{\Sigma_t \left(\text{choice}\,(t) - \text{life}\,(t)\right)^2}$$

We determined whether the Euclidean distances obtained for each channel were significantly different from chance through a permutation testing procedure (see Yeshurun *et al.*, 2017). Participants' group membership was shuffled, while ensuring that the sample sizes of the shuffled groups were matched to the original groups. Then, Euclidean distances were computed between the shuffled groups. This procedure was repeated 10 000 times, such that the observed Euclidean distance values could be compared to a null distribution of 10 000 shuffled Euclidean distance values. For each channel and video, P values were calculating by dividing the number of shuffled values that exceeded the observed Euclidean distance by the number of repetitions (number exceeding the observed values + 1/10 000).

Synchrony-based classification analyses using 'neural reference groups'

Subsequent to the Euclidean distance group analyses, we used a classification-based machine learning approach to investigate whether participants' partisan stance (pro-life or pro-choice) could be predicted at the individual level. These classification analyses, which were conducted on individual channels, involved comparing a participant's neural timecourse to average timecourses from the two partisan neural reference groups (Figure 2). In other words, the reference group averages, which excluded the participant's own data, served as benchmarks to which the participants' neural data could be compared. Participants were classified as belonging to a group based on showing greater similarity to (as in greater synchrony with) one reference group over the other. For this analysis, Euclidean distance was used as a measure of neural synchrony.

Following Yeshurun *et al.* (2017), classification analyses were conducted on a channel-by-channel basis in regions of interest (ROIs) selected based on the results of the Euclidean distance analysis. Classifications were conducted on fNIRS timecourses for each video separately. For each channel's analysis, the sample size for each partisan group ranged from $n = 18$ to $n = 29$,

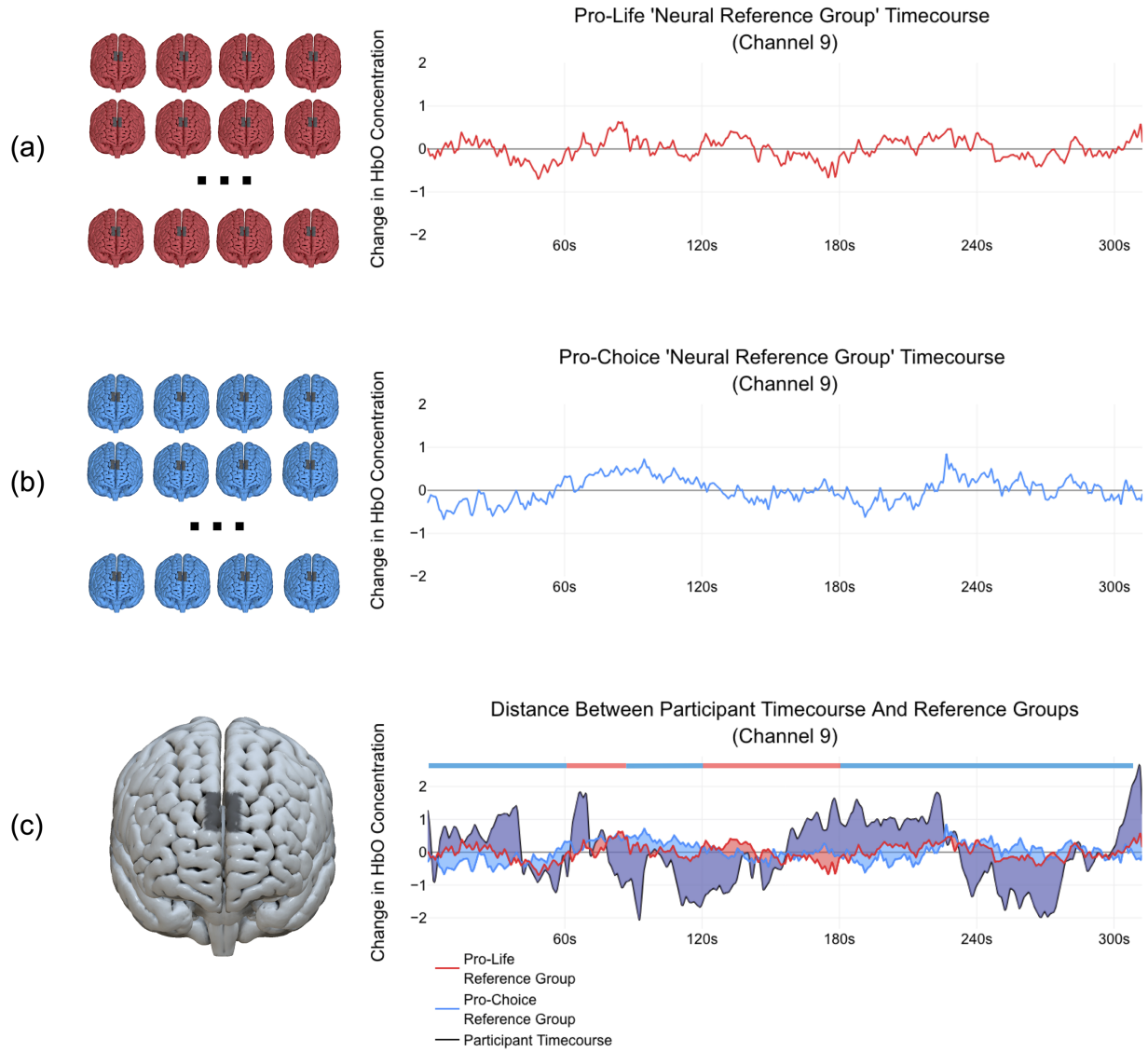

Fig. 2. Depiction of the neural reference group classification approach. (a) Neural timecourses from channel 9 for participants holding a pro-life stance are averaged together to form a pro-life neural reference group timecourse. (b) Timecourses for participants holding a pro-choice stance are averaged together to form a pro-choice neural reference group timecourse. (c) A participant's timecourse, whose data were not included in the reference group timecourses, is compared to the timecourses of the two neural reference groups. The participant is then categorized as belonging to one group or the other by demonstrating greater similarity with one group over the other, as measured by a distance metric (Euclidean distance in this case, though Pearson correlation might also be used). Areas that are shaded in purple demonstrate overlap where the participant's timecourse differed from both reference groups. Areas shaded blue or red correspond to where the participant's timecourse diverged more from one of the reference groups (blue=diverging further from pro-choice, red=diverging further from pro-life). These red and blue areas are key to determining which reference group the participant differs from most in order to match the participant as being likely to belong to one group or the other. Blue and red bars shown above the graph indicate sections of the timecourse where the participant differed more than (i.e. had a greater Euclidean distance from) one group or the other. For the participant shown here, a larger blue area than red area across all timepoints indicates that the participant differed more from the pro-choice group, and thus this participant was classified as being pro-life. In future studies, it may be valuable to examine regions of the timecourse when most participants tend to show similarity to one group over the other and identify moments in the video to which those timepoints correspond.

depending on how many participants had usable data within the channel. This sample size was deemed to be adequate based on the constraints of the study and previous classification-based neuroimaging work using similar sample sizes (Yeshurun *et al.*, 2017). For channels that had imbalanced data, such that there were different numbers of participants within each partisan group (or in machine learning terms, different numbers of 'samples' within each 'class'), we used a prototype generation algorithm to reduce the number of participants in the majority partisan group. This downsampling procedure, which was implemented using the imbalanced-learn Python package, utilizes k-means clustering to identify small groups of individual timecourses that cluster together within the majority partisan group (Lemaître *et al.*, 2017). It computes the average timecourse across participants within the identified clusters, and then replaces the original participant data with that newly

generated average. This process yielded an equal number of participants within each partisan group for each classification analysis.

To conduct the classification, a nearest centroid classifier was selected due to the study's small sample size, because it does not require the cross-validation procedure that is necessary for tuning hyper-parameters (see Yeshurun *et al.*, 2017). The classification procedure was implemented in Python using scikit-learn (Pedregosa *et al.*, 2011), where the accuracy of the classifier was tested using a leave-two-out process (i.e. leaving out one sample from each reference group to maintain equal numbers of samples within the two reference groups), with each sample being left out once. The model was tested on the left-out samples, having been trained on the remaining data.

We selected Euclidean distance to serve as the classifier's similarity index (i.e. the model's synchrony measure) and we selected the mean to represent the centroid, in accordance with standard defaults for the nearest centroid classifier and its use in previous work (Yeshurun *et al.*, 2017). During the classification procedure, for each fold in the leave-two out procedure, the Euclidean distance was computed separately between the neural timecourses of each of the two left-out samples and the mean timecourses of the remaining samples for the two partisan groups. Participants were classified as being a member of one group or the other based on which Euclidean distance value was lower. In other words, a participant was categorized as being likely to belong to whichever group's neural timecourse was more similar to their own timecourse within a given channel. For instance, if a participant's timecourse within a given channel was closer in Euclidean space to the average pro-life timecourse, that participant would be classified as being pro-life. In contrast, if a participant's timecourse was closer to the pro-choice timecourse, the participant would be classified as being pro-choice.

To obtain a measure of classification accuracy, the classifier's predictions were compared participants' true partisan positions, as measured by self-report. While the partisan position of each participant was known to the experimenters, the classification algorithm was blinded to the partisan position of the participants left out in any particular iteration. Classification accuracy scores were computed by dividing the number of participants that were classified correctly by the total number of participants included (number of participants correctly classified/number of classifications made). To obtain stable accuracy values, since different combinations of participants could be left-out in the leave-two-out procedure, the classification procedure was performed 1000 times within each channel. Final classification accuracy scores were computed as the average accuracy score from all 1000 repetitions. Permutation tests, where group membership labels were shuffled, were then used to test the significance of these accuracy scores. Classification accuracy scores were obtained for data shuffled over 10 000 repetitions and compared to the accuracy scores for the real dataset (number of null values larger than the real value + 1/10 000), an approach used by Yeshurun *et al.* (2017)

Results

Group-level behavioral differences

Prior to investigating for neural differences between the group, we first investigated whether there were differences in how members of the groups rated the videos. Specifically, we examined participants' perceptions about the argument strength of the videos. The two items used to assess the videos' perceived argument strength were highly correlated, and thus were combined into a composite variable for each video ($\alpha_{\text{pro-choice}} = 0.86$ [0.79, 0.93]); ($\alpha_{\text{pro-life}} = 0.89$ [0.83,0.94]).

We then conducted a repeated-measures ANOVA, with partisan group as a between-subjects factor and video type as a within-subjects factor. As predicted, there was a significant interaction in how participants from the two groups rated the perceived argument strength of the videos, $F(1,62) = 57.43$, $P < 0.001$, $\eta^2_p = 0.48$. Pro-choice participants rated the pro-choice argument as being of higher quality, ($M = 3.69$, $SD = 1.02$) than the pro-life participants ($M = 2.43$, $SD = 1.10$), $t(62) = -4.73$, $P < 0.001$. On the other hand, pro-life participants gave a higher rating to the pro-life argument ($M = 4.2$, $SD = 0.71$) than pro-choice participants ($M = 2.26$, $SD = 1.13$), $t(56.49) = -8.29$, $P < 0.001$. This indicated that the partisan groups were significantly different in terms of the extent to which they thought the videos contained strong, high-quality arguments.

Group-level neural differences

Given that behavioral differences were seen between the groups for the ratings of the arguments in the videos, we first examined whether there were also differences between the average neural timecourses of the two groups. For both videos, the greatest differences in neural responding between the pro-life and pro-choice groups were seen in channels located within the dmPFC, a region of the mentalizing network (Figure 3). In other words, participants in the two groups tended to respond more differently to the videos in this region. The largest Euclidean distance value, which was seen in channel 9 for the pro-life video, was marginally significant at $P < 0.06$. However, this effect was not significantly different from chance following False Discovery Rate (FDR) correction with a q criterion of 0.05 (Benjamini and Hochberg, 1995), which was used due to the large number of tests across videos and channels (2 videos $\times$ 20 channels = 40 tests). No other channels for either video showed significantly different Euclidean distances between the two groups.

Although these differences between the two partisan groups did not reach statistical significance at the group level, we also investigated whether it would be possible to make above-chance predictions about group membership at the individual level. Previous work has shown that in some instances, individual-level classification can achieve greater discriminatory power than group-level analyses due to inherent differences between the two methods (Arbabshirani *et al.*, 2017). Whereas the group-based difference analysis attempts to determine whether the partisan groups show different neural responses 'on average,' the individual-based classification analysis takes a slightly different approach. It investigates whether it is possible to categorize an individual as being likely to belong to one group or the other.

For the classification analyses, we began by implementing a simple ranked feature selection procedure to narrow down which channels would be used in order to reduce the number of statistical tests conducted. We selected the channels in which the group average timecourses were the farthest apart (>mean Euclidean distance value + 1*SD*) to serve as ROIs. The channels that passed this threshold were all located in dmPFC (channel 9 for the pro-choice video, channels 8, 9 and 10 for the pro-life

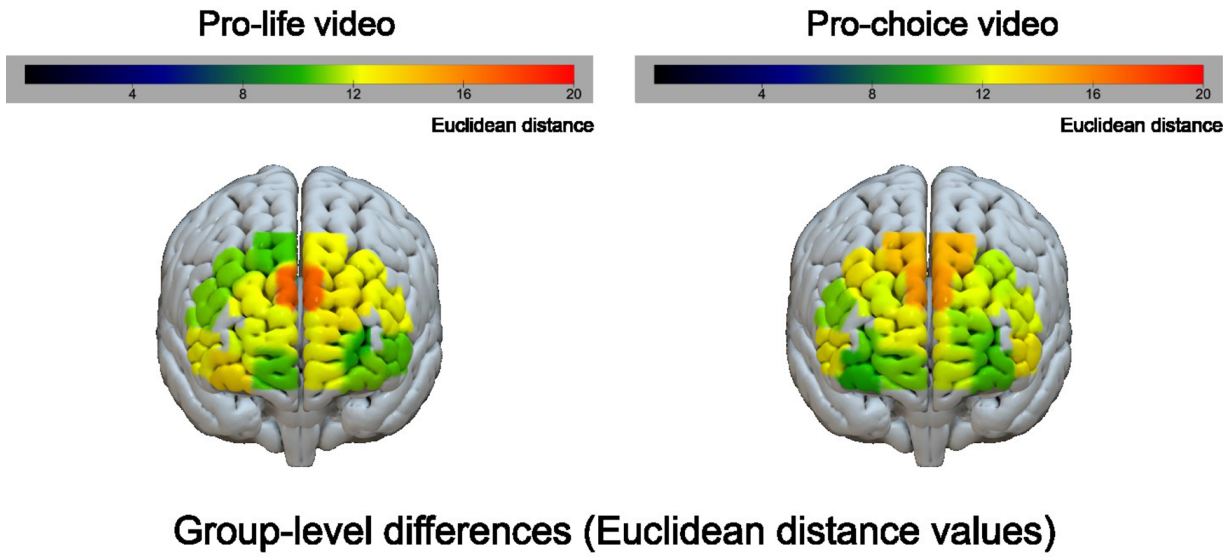

Fig. 3. For each video within each channel, group-level differences between pro-life and pro-choice participants were computed as the Euclidean distance between the mean timecourse for each group. These Euclidean distance values are shown projected onto a 3D cortical surface for each video: pro-life (left) and pro-choice (right). These maps were used to identify ROIs for conducting the classification-based analyses.

video). We modeled this ROI-based approach off of the procedure conducted by Yeshurun *et al.* (2017), which is analogous to the standard searchlight procedure developed by Kriegeskorte *et al.* (2006). In the majority of multivariate studies, a 'searchlight' is used to identify regions that show different levels of mean activity across conditions at the group level. Then, a classification analysis is applied on the same data at the individual level. This searchlight procedure was developed by the same research group that first raised methodological concerns about double dipping (Kriegeskorte *et al.*, 2009). According to Etzel *et al.* (2013), the searchlight procedure is not susceptible to the issues of double dipping given that the group- and individual-level analyses address fundamentally different questions: the group-level analyses examine mean differences, whereas the individual-level analyses examine individual differences. Furthermore, in our study, the ROIs were selected based on ranked distance values as opposed to using *P*-values generated through significance testing.

Synchrony-based classification results

To conduct the individual-level analyses, we trained a classifier in the selected ROIs within dmPFC for each video separately (Figure 4). For the pro-life video, only channel 9 passed the Euclidian distance threshold set, and hence we conducted the classification analysis within this channel only (dmPFC, [MNI: 2, 54, 38]). We found that participants' neural timecourses in channel 9 successfully predicted their attitudinal stance 66.52% of the time at above-chance levels ($P = 0.028$). Thus, it was possible to identify whether participants identified as being 'pro-choice' or 'pro-life' above chance based on how their dmPFC responded to an individual talking about his pro-life views. (Figure 4, left).

For the pro-choice video, we conducted analyses in channels 8, 9 and 10 of dmPFC, as all three surpassed the Euclidean distance threshold that we had set. We found that channel 8 (left dmPFC, [MNI: – 10, 44, 48]) predicted group membership 63.68% of the time, which was above chance ($P = 0.050$). Therefore, it was possible to identify participants' views based on how another region in dmPFC responded to an individual talking about his pro-choice views at better-than-chance rates (Figure 4, right). The classification analyses in channels 9 and 10 did not produce predictions at above-chance levels: the classification accuracy level was 62.31% ($P = 0.106$) for channel 9 and 50.33% ($P = 0.238$) for channel 10.

Therefore, we observed effects of dmPFC predicting partisan stance across both videos. Given this finding, we conducted an exploratory follow-up analysis to examine whether including data from both videos in a single analysis would improve the classifier's predictive ability. Channel 9 was selected as an ROI for this exploratory analysis given that its classification accuracy was greater than 60% for both videos. Participants were included in this analysis if they had usable data in channel 9 for at least one of the videos, which yielded a sample size of $N = 51$ ($n_{\text{pro-choice}} = 25$, $n_{\text{pro-life}} = 26$). For each video, a participant's time series data obtained in channel 9 was compared to the time series from the two reference groups. For participants who had quality data for both videos, this yielded four Euclidean distance values: (1) pro-life video time series (video) compared with pro-life reference group (ref); (2) pro-life video, pro-choice ref; (3) pro-choice video, pro-life ref and (4) pro-choice video, pro-choice ref. To calculate an average distance score relative to each reference group, we averaged the distance scores that were calculated relative to the same reference group across videos (i.e. 1 and 3, 2 and 4). Participants who had quality data for only one video had only 2 Euclidean distance value scores (one relative to each reference group for only one video), and thus, these were used to represent their average distance scores. Finally, a difference score between the average distances was used to classify participants as matching more closely with one reference group or the other. For instance, if a participant's average Euclidean distance from the pro-choice reference group was smaller than their distance from the pro-life group, they were classified as being pro-choice.

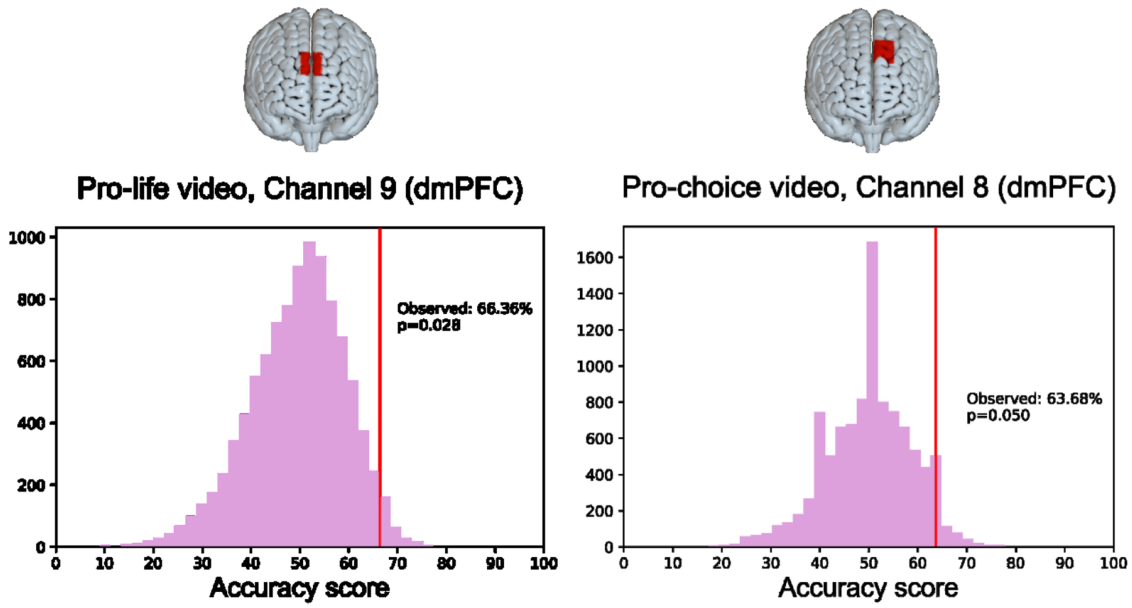

Fig. 4. Classification accuracy for channels (in dmPFC) that could distinguish between partisan groups at above-chance levels for the pro-life (left) and pro-choice (right) videos. For each video, the observed classification accuracy is shown relative to a null distribution of accuracy scores generated for shuffled data.

This approach did not yield a higher accuracy rate than what was achieved in channel 9 in the videos separately (accuracy = 54.90%, $P = 0.348$). However, an interesting finding emerged when we examined the extent to which there was consistency in classification across the pro-life and pro-choice videos. In other words, we investigated whether participants 'matched with' the same neural reference group for both videos. For instance, if a participant's timecourse for the pro-life video looked more similar to the pro-life reference group, and their timecourse for the pro-choice video also looked more similar to the pro-life reference group, they would be classified consistently as being pro-life. An inconsistently classified participant might show greater similarity to the pro-life reference group for one video, but greater similarity to the pro-choice reference group for the other, for instance. To be included in this analysis, participants were required to have usable data in channel 9 for both videos, which yielded a sample size of $N = 37$ ($n_{\text{pro-choice}} = 15$, $n_{\text{pro-life}} = 22$). Of the participants whose classification was consistent across videos ($n = 17$), 82.35% were classified accurately. In other words, if both classification tests yielded the same result, this result was highly diagnostic of the participant's true attitude. Permutation testing, which created a null distribution of accuracy scores obtained by comparing shuffled group assignments to the consistent participants' classified groups, indicated this was a significant result ($P = 0.001$); however, this analysis was conducted post hoc on a small sample and requires replication.

Discussion

Previous fMRI research has established that those who are 'like-minded' tend to show similarities based on how their brains respond to external stimuli (Parkinson *et al.*, 2018). Likewise, fMRI studies have shown that individuals who demonstrate differences in their internal states show differentiable neural responding (Lahnakoski *et al.*, 2014; Yeshurun *et al.*, 2017; Bacha-Trams *et al.*, 2018; Finn *et al.*, 2018 2020; Nguyen *et al.*, 2019; Chen *et al.*, 2020). Although studies have used neural synchrony measures to make predictions about experimentally induced psychological differences (Lahnakoski *et al.*, 2014; Yeshurun *et al.*, 2017), no synchrony-based studies to date have attempted to predict naturally occurring psychological characteristics, such as dispositional attitudes. Furthermore, no prior work has applied a classification-based approach to fNIRS data, which can be collected in more naturalistic environments as well as across culturally and demographically inclusive settings. Thus, the present study utilized fNIRS technology in a pop-up laboratory, measuring the neural responding of participants with two different partisan stances as they watched naturalistic video stimuli. The study's primary aim was to assess whether individuals' views could be predicted by applying a synchrony-based classification approach that compared individuals' neural data to data from neural reference groups.

Our results showed that we could predict participants' views on a specific abortion issue at above-chance levels. For two separate videos, classification could be achieved with significant accuracy using neural data acquired from dmPFC. In a subsequent exploratory analysis, participants who matched with the same neural reference group in dmPFC across both videos were classified at an even higher rate. This region is a part of the mentalizing network, a set of brain regions associated with thinking about mental states (Frith and Frith, 2006; Mitchell, 2009; Lieberman *et al.*, 2019). Prior fMRI and fNIRS studies have also demonstrated a positive association between dmPFC activity and perceptions of the effectiveness of persuasive messages (Klucharev *et al.*, 2008; Falk *et al.*, 2010 2013; Burns *et al.*, 2019). Thus, in the current study, participants in the two partisan groups were differentially responding in a region that has previously been associated with mentalizing and being persuaded by a message. Such a finding would track with differences observed in participants' self-report data, in which there were significant differences between the partisan groups in terms of how strong they found the video arguments to be.

For researchers who may be interested in conducting future research on synchrony-based classification using fNIRS data, it is worth noting that current fNIRS technology tends to have better signal in regions with thinner or no hair, and thus regions in prefrontal cortex, such as dmPFC, are optimal locations to measure. Whereas equipment constraints limited the number of regions that could be measured in the current study, future studies might also consider measuring signal in other DMN regions, such as the inferior parietal lobule and inferior parietal and temporoparietal junction. In addition, recent research has also demonstrated that friends, who tend to be similar to one another in terms of how they 'see' the world, show greater neural similarity in these regions (Parkinson *et al.*, 2018).

Although the same general brain region (dmPFC) yielded accurate classification across both videos in the current study, it is worth noting that the exact location of the channels that yielded the most accurate classifications for each video differed. For the pro-life video, significant classification was achieved using data from channel 9, but not from channel 8. The opposite was found for the pro-choice video (though here, channel 9 did show a trend towards significance). It is unclear why such a discrepancy may have occurred. It is possible that due to head movement, the fNIRS cap may have shifted such that the channels were in slightly different locations between the videos. However, we think this is unlikely given that the order of the videos was counterbalanced across participants. It is also possible that an inherent difference between the stimuli yielded differential activity in slightly different regions. We find it to be promising that similar effects were seen across the two videos, and yet we also would advocate for future research to attempt to obtain accurate classification in a consistent set of regions. Furthermore, we are encouraged by our finding that participants who were consistent in matching with the same reference group across stimuli within the same region were classified with a high degree of accuracy. This would suggest that future researchers who intend to use the neural reference groups approach in applied research might consider using 'neural synchrony consistency' across stimuli as a proxy for degree of confidence in predictions conducted at the individual level.

Despite it being possible to classify participants at the individual level in dmPFC channels, there were no significant differences in neural responses at the group level. Replications of this research may help explain why this occurred. One explanation for this could be that the partisan groups did not have truly dissociable neural data. We find this explanation to be unlikely due to a large body of evidence suggesting that individuals who hold different political beliefs show differential neural responding (Knutson *et al.*, 2006; Westen *et al.*, 2006; Kaplan *et al.*, 2007; Jost and Amodio, 2012; Ahn *et al.*, 2014; Van Bavel and Pereira, 2018).

An alternative explanation would be that the study was underpowered, such that the individual-based classification approach was more sensitive to neural differences than the group-level analyses. Previous research has shown that discrepancies can occur between these types of analyses due to differences in the research questions they attempt to address, and how they measure 'success' using different statistics (Arbabshirani *et al.*, 2017). It is possible that the fNIRS data collected in the pop-up lab in the Middle East were noisier than fNIRS or fMRI data from a traditional, controlled lab setting. Data collection was restricted to a 10-day timespan. Naturally, this meant that we did not have as large a sample as we would have liked. We are currently analyzing an analogous study run in our lab in the United States which has a larger sample size.

Nevertheless, accuracy rates of 66% and 63% in a binary classification are extremely typical for successful classification studies in neuroimaging. With high in-group variance resulting from a relatively small sample size and noisy data, it may be that we were underpowered to be able to detect statistically significant group-level differences. In contrast, the classification-based analysis focuses on comparing an individual's timecourse to the mean of each group and may be less sensitive to the amount of variance present. Even if both groups have high variance, accurate prediction may still occur if enough signal is present in the mean to facilitate the individual's matching with the correct group. Thus, further work examining the relationship between group-level and individual-level classification analyses on time series data may help explain why these discrepancies might happen in the context of this particular classification approach. In addition, future studies might consider collecting larger sample sizes, along with employing techniques to reduce statistical noise caused by participants and/or equipment.

Furthermore, it is possible that the low-budget quality of the stimuli used in the current experiment influenced statistical power. Given the study's time constraints, the actors in the videos used in the stimulus set alternated between making eye contact with the camera versus looking down at their scripts, which may have elicited muted emotional responses from participants. However, even if participants could recognize that the speakers in the videos were actors, the videos' political content was enough to elicit distinguishable neural responses between partisan groups. We believe that the limitations of our stimuli make the study's significant findings more impressive, and expect that richer stimuli might yield stronger effects. For instance, previous work has shown that highly engaging stimuli are more likely to evoke higher levels of neural synchrony (Cohen *et al.*, 2017). In terms of identifying distinguishable group differences, an ideal stimulus would be one that is highly engaging for individuals within a group and also polarizing between two or more groups. Future researchers who wish to apply the current classification approach should carefully consider the selection of their stimuli to optimize statistical power.

Finally, this study should be seen as a 'proof of concept,' demonstrating that it is possible to predict attitudes by conducting classification analyses on naturalistic timecourse data. More work is needed to demonstrate that models using the neural reference groups approach can make accurate out-of-sample predictions. It remains an open question whether this classification approach can generalize beyond a small sample of individuals who share similar demographics or if it becomes fine-tuned to the particularities of a specific population used in a particular study. For instance, this study used a small stimulus set focused on one sociopolitical issue, and it was conducted only among Arab males living in Jordan. Additional work will be required to replicate this work to ensure that the findings generalize to attitudes on other issues among other populations.

In summary, this study demonstrates that the neural reference groups approach can be used to make predictions about real-world differences using data collected in naturalistic settings around the world. Furthermore, such predictions can be made by using a synchrony-based classification approach that utilizes neural reference groups. The classification accuracy scores obtained in our study were greater than those that would be achieved by chance and are consistent with scores observed in a prior, analogous fMRI study (Yeshurun *et al.*, 2017). We find this result to be encouraging, given the challenges that were

posed by collecting data in a pop-up neuroscience lab with low-budget stimuli and time constraints. We are hopeful that it may be possible to obtain higher classification accuracies in fNIRS research as more advanced equipment and analysis techniques are developed, and as neuroimaging researchers learn how to optimize experimental design in naturalistic contexts.

Having the ability to take neuroimaging 'on the road,' and to make predictions about individuals based on their brain responses, is likely to open up new opportunities for field research in naturalistic settings with more diverse, non-WEIRD samples (Burns *et al.*, 2019). Recently, there has been growing interest in using portable neuroimaging, in combination with synchrony analyses, to understand social interactions in real-world settings (Dumas *et al.*, 2010; Dikker *et al.*, 2017). However, portable devices also afford the ability to conduct single-person analyses on any population, anywhere in the world and at low costs.

Using fNIRS or other neuroimaging modalities, it is at least plausible that the neural reference groups approach could predictively identify any hidden state or trait that influences how we process the world around us. For instance, one could determine whether individuals respond better to one teaching approach or another, resonate more or less with particular versions of public health messages or show neural responses more consistent with being open-minded or closed-minded in particular contexts. It is our hope that researchers will continue to build upon the neural reference groups approach to use neuroimaging in more applied and naturalistic settings.

Acknowledgements

We thank the staff at Independent Institute & Administration Civil Society Studies (IIACSS) for their hard work in translating study materials and providing the participant data.

Funding

This article is the result of funding from the U.S. Department of Defense's Minerva Initiative (13RSA281, PI: MDL) and National Defense Science & Engineering Graduate Fellowship (NDSEG) Program (Fellow: MCD).

Declaration of interest

The authors declare no conflict of interest.

References

Ahn, W.Y., Kishida, K.T., Gu, X., et al. (2014). Nonpolitical images evoke neural predictors of political ideology. *Current Biology*, *24*(22), 2693–9.

Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D. (2017). Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *NeuroImage*, *145*, 137–65.

Bacha-Trams, M., Alexandrov, Y.I., Broman, E., et al. (2018). A drama movie activates brains of holistic and analytical thinkers differentially. *Social Cognitive and Affective Neuroscience*, *13*(12), 1293–304.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F., et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582* 1–7.

Burns, S.M., Barnes, L.N., Katzman, P.L., Ames, D.L., Falk, E.B., Lieberman, M.D. (2018). A functional near infrared spectroscopy (fNIRS) replication of the sunscreen persuasion paradigm. *Social Cognitive and Affective Neuroscience*, *13*(6), 628–36.

Burns, S.M., Barnes, L.N., McCulloh, I.A., et al. (2019). Making social neuroscience less WEIRD: using fNIRS to measure neural signatures of persuasive influence in a Middle East participant sample. *Journal of Personality and Social Psychology*, *116*(3), e1-11.

Chen, P.H., Jolly, E., Cheong, J.H., Chang, L.J. (2020). Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. *NeuroImage*, *216* 116851.

Cohen, S.S., Henin, S., Parra, L.C. (2017). Engaging narratives evoke similar neural activity and lead to similar time perception. *Scientific Reports*, *7*(1), 1–10.

Cui, X., Bray, S., Bryant, D.M., Glover, G.H., Reiss, A.L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*, *54*(4), 2808–21.

Denny, B.T., Kober, H., Wager, T.D., Ochsner, K.N. (2012). A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *24*(8), 1742–52.

Dikker, S., Wan, L., Davidesco, I. et al. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, *27*(9), 1375–80.

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., Garnero, L., Lauwereyns, J. (2010). Inter-brain synchronization during social interaction. *PLoS One*, *5*(8).

Etzel, J.A., Zacks, J.M., Braver, T.S. (2013). Searchlight analysis: promise, pitfalls, and potential. *NeuroImage*, *78*, 261–69.

Falk, E.B., Berkman, E.T., Mann, T., Harrison, B., Lieberman, M.D. (2010). Predicting persuasion-induced behavior change from the brain. *The Journal of Neuroscience*, *30*, 8421–4.

Falk, E.B., Morelli, S.A., Welborn, B.L., Dambacher, K., Lieberman, M.D. (2013). Creating buzz: the neural correlates of effective message propagation. *Psychological Science*, *24*, 1234–42.

Finn, E.S., Corlett, P.R., Chen, G., Bandettini, P.A., Constable, R.T. (2018). Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nature Communications*, *9*(1), 2043.

Finn, E.S., Glerean, E., Khojandi, A.Y., et al. (2020). Idiosynchrony: from shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, *215* 116828.

Frith, C.D., Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–4.

Hasson, U., Malach, R., Heeger, D.J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–8.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–40.

Honey, C.J., Thompson, C.R., Lerner, Y., Hasson, U. (2012). Not lost in translation: neural responses shared across languages. *Journal of Neuroscience*, *32*(44), 15277–83.

Huppert, T.J., Diamond, S.G., Franceschini, M.A., Boas, D.A. (2009). Homer: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, *48*(10), D280–98.

Jääskeläinen, I.P., Koskentalo, K., Balk, M.H., et al. (2008). Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. *The Open Neuroimaging Journal*, 2, 14–9.

Jost, J.T., Amodio, D.M. (2012). Political ideology as motivated social cognition: behavioral and neuroscientific evidence. *Motivation and Emotion*, 36(1), 55–64.

Kanai, R., Feilden, T., Firth, C., Rees, G. (2011). Political orientations are correlated with brain structure in young adults. *Current Biology*, 21(8), 677–80.

Kaplan, J.T., Freedman, J., Iacoboni, M. (2007). Us versus them: political attitudes and party affiliation influence neural response to faces of presidential candidates. *Neuropsychologia*, 45(1), 55–64.

Klucharev, V., Smidts, A., Fernández, G. (2008). Brain mechanisms of persuasion: how 'expert power' modulates memory and attitudes. *Social Cognitive and Affective Neuroscience*, 3, 353–66.

Knutson, K.M., Wood, J.N., Spampinato, M.V., Grafman, J. (2006). Politics on the brain: an fMRI investigation. *Social Neuroscience*, 1(1), 25–40.

Kriegeskorte, N., Goebel, R., Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, 103(10), 3863–8.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P., Baker, C.I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, 12(5), 535.

Lahnakoski, J.M., Glerean, E., Jääskeläinen, I.P., et al. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, 100, 316–24.

Lemaître, G., Nogueira, F., Aridas, C.K. (2017). Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research*, 18(1), 559–63.

Lieberman, M.D., Straccia, M.A., Meyer, M.L., Du, M., Tan, K.M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): causal, multivariate, and reverse inference evidence. *Neuroscience and Biobehavioral Reviews*, 99, 311–28.

Liu, J., O'Donnell, M.B., Falk, E.B. (2020). Deliberation and valence as dissociable components of counterarguing among smokers: evidence from neuroimaging and quantitative linguistic analysis. *Health Communication*, 1–12.

Memarian, N., Torre, J.B., Haltom, K.E., Stanton, A.L., Lieberman, M.D. (2017). Neural activity during affect labeling predicts expressive writing effects on well-being: GLM and SVM approaches. *Social Cognitive and Affective Neuroscience*, 12(9), 1437–47.

Mitchell, J.P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1309–16.

Nastase, S.A., Gazzola, V., Hasson, U., Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, 14(6), 667–85.

Nguyen, M., Vanderwal, T., Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, 184, 161–70.

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I.P., Hari, R., Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*, 109(24), 9599–604.

Nummenmaa, L., Lahnakoski, J.M., Glerean, E. (2018). Sharing the social world via intersubject neural synchronisation. *Current Opinion in Psychology*, 24, 7–14.

O'Donnell, M.B., Coronel, J., Cascio, C.N., Lieberman, M.D., Falk, E.B. (2018, May). An fMRI localizer for deliberative counterarguing. Paper presented at The Social & Affective Neuroscience Society Annual Meeting, Brooklyn, NY.

Parkinson, C., Kleinbaum, A.M., Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, 9(1), 1–14.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–30.

Poldrack, R.A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron*, 72(5), 692–97.

Regev, M., Honey, C.J., Simony, E., Hasson, U. (2013). Selective and invariant neural responses to spoken and written narratives. *Journal of Neuroscience*, 33(40), 15978–88.

Saalasti, S., Alho, J., Bar, M., et al. (2019). Inferior parietal lobule and early visual areas support elicitation of individualized meanings during narrative listening. *Brain and Behavior*, 9(5), e01288.

Singh, A.K., Okamoto, M., Dan, H., Jurcak, V., Dan, I. (2005). Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. *NeuroImage*, 27(4), 842–51.

Spunt, R.P., Adolphs, R. (2014). Validating the why/how contrast for functional MRI studies of theory of mind. *NeuroImage*, 99, 301–11.

Strangman, G., Culver, J.P., Thompson, J.H., Boas, D.A. (2002). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *NeuroImage*, 17(2), 719–31.

Tei, S., Kauppi, J.-P., Fujino, J., et al. (2019). Inter-subject correlation of temporoparietal junction activity is associated with conflict patterns during flexible decision-making. *Neuroscience Research*, 144, 67–70.

Van Bavel, J.J., Pereira, A. (2018). The partisan brain: an identity-based model of political belief. *Trends in Cognitive Sciences*, 22(3), 213–24.

Westen, D., Blagov, P.S., Harenski, K., Kilts, C., Hamann, S. (2006). Neural bases of motivated reasoning: an fMRI Study of emotional constraints on partisan political judgment in the 2004 US presidential election. *Journal of Cognitive Neuroscience*, 18(11), 1947–58.

Wilson, S.M., Molnar-Szakacs, I., Iacoboni, M. (2007). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, 18(1), 230–42.

Yeshurun, Y., Swanson, S., Simony, E., et al. (2017). Same story, different story: the neural representation of interpretive frameworks. *Psychological Science*, 28(3), 307–19.

Zhao, X., Strasser, A., Cappella, J.N., Lerman, C., Fishbein, M. (2011). A measure of perceived argument strength: reliability and validity. *Communication Methods and Measures*, 5(1), 48–75.