UNIVERSITY OF CALIFORNIA

Los Angeles

Neural Representations of Attitude Polarization and Open-Mindedness

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Psychology

by

Macrina Cooper Dieffenbach

2021

ABSTRACT OF THE DISSERTATION

Neural Representations of Attitude Polarization and Open-Mindedness

By Macrina Cooper Dieffenbach

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2021

Professor Matthew Lieberman, Chair

When individuals *see* the world differently, the divide between their subjective construals manifests in differential neural responses. Research on neural polarization has found that individuals who share similar viewpoints tend to synchronize their brain responses and those with different viewpoints show distinguishable brain responses. In this dissertation, I attempt to build upon and extend this nascent literature on neural polarization by demonstrating two novel ways in which neural synchrony analyses can shed light on people's viewpoints. This research utilizes a cost-effective and portable neuroimaging tool called functional near infrared spectroscopy (fNIRS), which is optimized for measuring brain responses in the mentalizing network. In Chapter 2, I demonstrate that a classification technique called the 'neural reference groups' approach can be used to predict individuals' political viewpoints at an above-chance level from the prefrontal cortex. In Chapter 3, I explore how the neural reference groups approach can also be used to detect whether an open-mindedness intervention has impacted individuals' subjective construal

processes. In Chapter 4, I provide a comprehensive review of open-mindedness interventions that have been developed in order to provide a roadmap for researchers who may be interested in applying the neural reference groups approach to other interventions.

The dissertation of Macrina Cooper Dieffenbach is approved.

Dr. Carolyn M. Parkinson

Dr. Naomi I. Eisenberger

Dr. Rick Dale

Dr. Matthew D. Lieberman, Committee Chair

2020

This dissertation is dedicated to three people who have

supported, encouraged, and loved me endlessly:

my mother, Pamela

my father, Michael,

and my husband, James.

Without you, this would not be possible.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2 - Neural reference groups: a synchrony-based classification approach for predicting attitudes using fNIRS

Chapter 3 - Leveraging Differences to Bring People Together: Using Neural Synchrony to Detect Polarized Thinking and Evaluate Open-Mindedness Interventions

Chapter 4 - Open-Mindedness Interventions: An Integrative Review and Roadmap for Future Research

LIST OF TABLES

Chapter 3 - Leveraging Differences to Bring People Together: Using Neural Synchrony to Detect Polarized Thinking and Evaluate Open-Mindedness Interventions

Chapter 4 - Open-Mindedness Interventions: An Integrative Review and Roadmap for Future Research

ACKNOWLEDGEMENTS

First and foremost, I am incredibly grateful to my advisor, Dr. Matthew Lieberman. I always come out of our conversations reassured, inspired, and brimming with ideas, which must be the sign of a brilliant mentor. Thank you for always making the time when I needed it, especially in this final stretch!

Thank you to my committee members, Drs. Carolyn Parkinson, Naomi Eisenberger, and Rick Dale. Your work and thinking has truly shaped my own over the past years and I am very grateful for that. To my research assistants: in particular, Janet, Caleb, Kris. Thank you for making me a better mentor, collecting data for hours on end, and for sharing your ideas and feedback with me. I would also like to thank Dan, Shannon, and Ian for helping to design and create some of the stimuli used in these studies. Further, I would like to thank Devon, Ren, Brian, and Patrick for providing their creative talents to act in my video stimuli. I am also so appreciative for the staff at IIACSS in Amman, Jordan. Thank you for conducting marathon data collection days, translating study materials into Arabic, and enlightening me as to how real hummus should taste.

A big thank you goes to my funders for supporting this work: The National Defense Science and Engineering Graduate Fellowship, the United States Department of Defense MINERVA Initiative, and to the Norma & Seymour Feshbach Doctoral Dissertation Award. In addition, I would like to thank Oxford University Press, who has permitted unrestricted use, distribution, and reproduction of the following article, a version of which is reproduced in this dissertation in Chapter 2: Dieffenbach, M. C., Gillespie, G. S., Burns, S. M., McCulloh, I. A., Ames, D. L., Dagher, M. M., ... & Lieberman, M. D. (2021). Neural reference groups: a synchrony-based classification approach for predicting attitudes using fNIRS. *Social Cognitive and Affective Neuroscience, 16*(1-2), 117-128.

I would like to send out a big thank you to my colleagues at OpenMind. Thanks especially to Caroline for downloading the encyclopedia of your brain into mine over the past two years, and to Jon, for helping me to realize that when you are trying to do interventions, you have to think about

group dynamics. To my labmates - thank you for strengthening my ideas through your thoughtful ideas and feedback. To my cohortmates, Carrie and Jessica, I am very grateful to have gone through this wild ride together. A big thank you goes to my cohortmate, labmate, and friend, Shannon. You paved the way for fNIRS and I have been lucky to have the honor of riding on your coattails.

Thank you to Melo. I am grateful for the writing breaks you encouraged me to take and for your constant companionship. You were the silver lining for me as I spent this final year and a half at home.

To all of my friends, who have provided support and affirmation for many years. To Devon. To my Cooper and Dieffenbach family. Thank you for your cheerleading! To John and Anne - I am so thankful to have a second set of loving parents. Matt and Emily, thank you for sorely needed pizza and game nights in our quarantine bubble. To my parents. Without everything you have done for me, this would not be possible. Dad, you are the wind “above” my wings. Thank you for your cheerful phone calls and for helping me deal with “real life stuff” like my taxes amidst writing a dissertation. Mom, I still will not be able to live up to the number of degrees you hold, but at least I got close. Thank you for modeling a love of learning, for debating about Freud, and for answering all of my questions since I first started asking them. Thank you for always being there to tell me “You’ll get it done. You always do.” Thankfully, you haven’t been wrong yet!

Finally, to James. You are the best life partner, editor, coach, and best friend a person could hope for. Thank you for embarking on this cross-country adventure with me and for celebrating every step along this journey. This success belongs equally to you.

VITA

EDUCATION

2019	Candidate in Philosophy in Psychology; University of California, Los Angeles
2016	Master of Arts in Psychology, University of California, Los Angeles
2012	Bachelor of Arts in Cognitive Science, Yale University

FELLOWSHIPS AND HONORS

2020	UCLA Psychology Department's Norma & Seymour Feshbach Doctoral Dissertation Award
2018	Sao Paulo School of Advanced Science in Social and Affective Neuroscience Summer Traineeship
2018	Duke University Summer School in Social Neuroscience & Neuroeconomics Traineeship
2016-2019	U.S. Department of Defense National Defense Science and Engineering Graduate Fellowship
2016	fMRI Training Course Fellowship, University of Michigan
2016	National Science Foundation Graduate Fellowship, Honorable Mention
2015-2017	UCLA Graduate Dean's Scholar Award
2015	UCLA Distinguished University Fellowship

PUBLICATIONS

Welborn, B.L., **Dieffenbach, M.C.**, Lieberman, M.D. (under review). Default egocentrism: An MVPA approach to overlap in own and others' attitudes.

Dieffenbach, M.C., Gillespie, G.S., Burns, S.M., McCulloh, I., Ames, D., Dagher, M.M., Falk, E., & Lieberman, M. (2021). Neural reference groups: a synchrony-based classification approach for predicting attitudes using fNIRS. *Social Cognitive and Affective Neuroscience, 16,* 117 - 128.

Sahi, R.S., **Dieffenbach, M.C**., Gan, S., Lee, M., Hazlett, L.I., Burns, S.M., Lieberman, M., Shamay-Tsoory, S., & Eisenberger, N. (2021). The comfort in touch: Immediate and lasting effects of handholding on emotional pain. *PloS one*, 16 2, e0246753 .

PRESENTATIONS

**Formerly presented under Cooper-White (maiden name)*

Dieffenbach, M., Sawaoka, T., Votta, F., Haidt, J. (February 2021) *OpenMind: A scalable online intervention to depolarize campuses and communities.* Symposium chair at the 2021 annual meeting of the Society for Personality and Social Psychology, Virtual Convention.

Cooper-White, M., Gillespie, G. S. R., Ames, D. L., Burns, S., McCulloh, I. A., Dagher, M. M., & Lieberman, M. D. (May 2019) *Neural Partisanship In the Middle East: An fNIRS Study.* Poster presented at the annual meeting of the Social and Affective Neuroscience, Miami, FL, USA.

Cooper-White, M., Gillespie, G. S. R. Ames, D. L., Burns, S., & Lieberman, M. D. (Feb. 2019) *An Investigation of Self-Affirmation As A Tool For Reducing Defensiveness Against Opposing Political Views.* Poster accepted for presentation at the annual meeting of the Society for Personality and Social Psychology, Portland, OR, USA.

Cooper-White, M., Ames, D. L., Burns, S., Tan, K., Gillespie, G. & Lieberman, M. D. (May 2018) *Investigating the Neural and Cognitive Mechanisms Behind 'Latitudes of Acceptance' For The Opinions of Others.* Poster presented at the annual meeting of the Social and Affective Neuroscience, Brooklyn, NY, USA.

Cooper-White, M. & Lieberman, M. D. (Jan. 2017) *The Influence of Latitudes of Acceptance on Social Decision-Making.* Poster presented at the annual meeting of the Society for Personality and Social Psychology, San Antonio, TX, USA.

RELEVANT PROFESSIONAL EXPERIENCE

2017-2021	UCLA Graduate Writing Center Consultant
2019-2021	Research Scientist/Manager at OpenMind Platform
2018	Neuroimaging Research Intern, Arrow Analytics

SELECTED SERVICE

2016-2020	Board Member, UCLA Association for Women in Science and Engineering
2019	Graduate Student Representative, UCLA Graduate/Professional Student Survey
2016-2017	Volunteer and Workshop Facilitator, DIY Girls

Chapter 1- General Introduction

Background

In a popular cartoon, two people stand on opposite sides of an ambiguous number. From one angle, the number looks like a “6.” From the other side, it resembles a “9.” Both people point at the number, repeating their interpretations out loud with a grimace, unable to understand how the other person could *possibly* see things differently. These fictional characters are experiencing naïve realism (Ross & Ward, 1996). They believe that their view of the world is correct and that the other person must be crazy, stupid, or biased. This may be a toy example, but it is not so far off from how we tend to view people who think differently from us when it comes to much more important issues. For instance, in the United States, partisans have shown an increasing amount of antipathy toward the other side over the past few decades (Iyengar et al., 2019). One recent study found that 42% partisans thought the other side was “downright evil” (Kalmoe & Mason, 2019). Similar effects have been demonstrated around the world, with partisan groups in 26 countries misunderstanding those on the other side (Ruggeri et al., 2021).

In response to this rise in *affective polarization* between people who hold different viewpoints, researchers have attempted to both understand how it affects subjective experience and also develop interventions to reduce it. Recent research has increasingly aimed to “get under the hood” of the partisan mind, attempting to understand the antecedents and consequences of ideological thinking (Zmigrod, 2021; Zmigrod & Tsakiris, 2021; Zmigrod et al., 2021). Studies have found that political polarization manifests neurally, leading liberals and conservatives to have more similar brain responses to members of their political ingroup in comparison to the outgroup (Leong et al., 2020). These *neural polarization* studies have used an approach called intersubject correlation – or neural synchrony – which measures the extent to which people's brains fluctuate in similar ways as they process naturalistic stimuli (Hasson et al., 2004). Leong et al. (2020) found that liberals and conservatives had greater within-group versus between-group synchrony in the dorsomedial prefrontal cortex (DMPFC), a region found in the mentalizing network. Another study found that this synchrony effect was modulated by tolerance to uncertainty, such that partisans who were less tolerant to uncertainty showed greater neural polarization in the brain's mentalizing network (i.e. more synchrony with their ingroup and less synchrony with the outgroup; van Baar et al., 2021).

Therefore, it is evident that the ways in which liberals and conservatives *see* things differently is reflected in their differentiable neural responses. However, less is known with regards to whether it is possible to predict an individual's partisan stance based on their brain data. Further work is needed to develop techniques to predict partisan stance and also to better understand the manner in which partisans polarize neurally.

In addition, the previous neural polarization studies have been conducted primarily using functional magnetic resonance imaging (fMRI). With its high spatial resolution, this neuroimaging modality has provided a rich understanding of where neural polarization occurs in the brain. However, fMRI work is limited in that it is highly expensive and immobile. In contrast, functional near infrared spectroscopy (fNIRS), which detects the same signal as fMRI (i.e. levels of blood oxygenation in the brain), is relatively cheap and portable. Although fNIRS has a lower spatial resolution than fMRI and cannot access subcortical regions, neural polarization research has tended to focus on functional areas of the brain that are large and located in the cortex (such as regions of the mentalizing network), meaning that fNIRS can capture signal in them. As such, fNIRS presents new opportunities for social neuroscience work, allowing researchers to study populations that were previously inaccessible, to collect larger sample sizes, to conduct research in more naturalistic settings, and even to measure neural responses during social interaction. Thus, fNIRS provides a useful modality for the continued study of neural polarization in new populations with larger sample sizes.

Finally, though studies have been dedicated to understanding the partisan brain, less work has been done to use neuroscience to measure the impact of interventions that aim to reduce polarization. This may be because research on neural polarization is still nascent, such that it has been important to describe how polarization manifests before moving to change it. Given this gap in the literature, we propose that more work is needed to understand how interventions affect people's ability to be open-minded toward other viewpoints as opposed to being dogmatic and affectively polarized. Many interventions have been developed to promote open-mindedness. However, it is often challenging to measure their impact due to the limitations of self-report, which can be subject to experimental demand characteristics, social desirability biases, and a lack of introspective ability. Neuroscience may provide a unique avenue for circumventing these limitations in order to test the impact of interventions that aim to reduce polarization.

The Present Dissertation

This dissertation, which contains two studies and one narrative review, aims to advance our understanding of neural polarization in several ways. In Chapter 2, I present the first study of neural partisanship conducted using fNIRS. Thanks to the portability of fNIRS technology, one of my collaborators and I flew to the Middle East to set up a "pop-up neuroimaging lab." We collected neural data from an equal number of pro-choice and pro-life participants as they watched videos of speakers talking about their views on abortion. We hypothesized that we would replicate previous findings that partisans show distinguishable neural responses (Leong et al., 2020). Furthermore, we hypothesized that if partisans clustered into separate groups, we should be able to predict participants' partisanship at the individual level by trying to categorize each participant's brain response as being more similar to the average response from one partisan group versus the other. Thus, to analyze the data, I applied a machine learning technique that we have termed the *neural reference groups* approach in order to classify participants' partisan stance. Other research has classified political orientation based on participants' responses to non-political images (e.g. snakes, chair, baby, an aimed gun; Ahn et al., 2014). However, to our knowledge, this is the first study to classify partisanship using naturalistic stimuli in a real-world setting.

Given that the study presented in Chapter 2 demonstrated that neural polarization can be measured using fNIRS, I designed another study to extend this work, which is presented in Chapter 3. This study had two primary aims. The first aim was to replicate the findings from Chapter 2 with a full cortical setup in order to capture signal in the overall mentalizing network (not just the prefrontal cortex). The second aim of this study was to use the neural reference groups model developed in our prior study to measure the impact of an open-mindedness intervention. We reasoned that if the technique could classify individuals' partisan stance based on their neural responses, it could also be used to classify whether or not a person had undergone an intervention that aimed to alter their subjective construals. Participants in the intervention condition completed a self-affirmation exercise that had previously been found to reduce defensive responding against counterattitudinal information (Cohen et al., 2007). We hypothesized that the intervention would result in a change to activity in participants' mentalizing network. If the intervention changed participants' neural responses, such that they became distinguishable from their control peers, we hypothesized that the neural reference groups approach would be able to classify participants according to whether or not they had undergone an intervention. In summary, this technique allowed us to circumvent the limitations of self-report in order to measure the effect of an intervention on participants' actual cognitive processing. Overall, the two empirical studies presented in Chapters 2 and 3 attempt to extend neural polarization work beyond being merely descriptive to move in a direction that is more predictive and more focused on measuring the efficacy of interventions.

In Chapter 4, I transition to providing an overview of interventions that researchers have tested in order to promote open-mindedness. If future researchers attempt to build upon the study presented in Chapter 3, this narrative review provides a roadmap of future interventions that may be tested, either behaviorally, neurally, or in combination. In this review, I attempt to encourage cross-talk between disparate academic fields and practitioners, all of whom have been attempting to solve the problem of affective polarization in different ways. In addition to reviewing prior research, I also offer a conceptual model that describes open-mindedness as a dynamical system consisting of underlying cognitive, motivational, affective, and social factors. This system exists within the context of environmental factors like social norms and social structures. Therefore, this narrative review attempts to integrate ideas from many fields to provide a resource for future work that aims to reduce affective (and neural) polarization.

In conclusion, this dissertation aims to extend prior work on neural synchrony in new directions that have real-world applications. It describes a method for both measuring neural polarization and predicting how people *see* the world using portable neuroimaging technology. I demonstrate that this 'neural reference groups' approach can be used to predict differences in people's mindsets that are both naturally-occurring (e.g. partisan views) and experimentally-induced (e.g. open-mindedness versus closed-mindedness). This new approach opens up intriguing possibilities for better understanding differences between groups who hold different viewpoints, as well as for measuring the impact of interventions that aim to improve open-mindedness toward alternative viewpoints. Finally, this dissertation concludes with a comprehensive review of open-mindedness interventions, which may be used as a reference by those who wish to extend this dissertation's work on 'neural reference groups' in future research.

REFERENCES

Ahn, W. Y., Kishida, K. T., Gu, X., Lohrenz, T., Harvey, A., Alford, J. R., ... & Montague, P. R. (2014). Nonpolitical images evoke neural predictors of political ideology. *Current Biology*, *24*(22), 2693-2699.

Cohen, G. L., Sherman, D. K., Bastardi, A., Hsu, L., McGoey, M., & Ross, L. (2007). Bridging the partisan divide: Self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. *Journal of Personality and Social Psychology*, *93*(3), 415–430. https://doi.org/10.1037/0022-3514.93.3.415

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science, 303*(5664), 1634-1640.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*, 129-146.

Kalmoe, N. P., & Mason, L. (January, 2019). Lethal mass partisanship: Prevalence, correlates, and electoral contingencies. Presented at the *National Capital Area Political Science Association American Politics Meeting.*

Leong, Y. C., Chen, J., Willer, R., & Zaki, J. (2020). Conservative and liberal attitudes drive polarized neural responses to political content. *Proceedings of the National Academy of Sciences*, *117*(44), 27731-27739.

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In T. Brown, E. Reed, & E. Turiel (Eds.), *Values and Knowledge* (pp. 103-135). Hillsdale, NJ: Lawrence Erlbaum.

Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., ...

& Folke, T. (2021). The general fault in our fault lines. *Nature Human Behaviour*, 1-11.

van Baar, J. M., Halpern, D. J., & FeldmanHall, O. (2021). Intolerance of uncertainty modulates brain-to-brain synchrony during politically polarized perception. *Proceedings of the National Academy of Sciences, 118*(20).

Zmigrod, L. (January, 2021). *A Psychology of Ideology: Unpacking the Psychological Structure of Ideological Thinking.* PsyArXiv. https://doi.org/10.31234/osf.io/ewy9t

Zmigrod, L., Eisenberg, I. W., Bissett, P. G., Robbins, T. W., & Poldrack, R. A. (2021). The cognitive and perceptual correlates of ideological attitudes: a data-driven approach. *Philosophical Transactions of the Royal Society B*, 376(1822), 20200424.

Zmigrod, L. & Tsakiris, M. (2021). Computational and neurocognitive approaches to the political brain: key insights and future avenues for political neuroscience. *Philos. Trans. R Soc. Lond B Biol. Sci.* 376:20200130. 10.1098/rstb.2020.0130

Chapter 2 - Neural Reference Groups:
A Synchrony-Based Classification Approach for Predicting Attitudes Using fNIRS

[Note: This is a preprint copy of the published manuscript: Dieffenbach, M. C., Gillespie, G. S., Burns, S. M., McCulloh, I. A., Ames, D. L., Dagher, M. M., ... & Lieberman, M. D. (2021). Neural reference groups: a synchrony-based classification approach for predicting attitudes using fNIRS. *Social Cognitive and Affective Neuroscience*, *16*(1-2), 117-128.

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. fMRI) 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 (fNIRS) 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 (dmPFC). These results indicate that the neural reference groups approach can be used to investigate naturally-occurring, dispositional differences anywhere in the world.

Keywords: Neural reference groups, neural synchrony, intersubject correlation, fNIRS, dmPFC

Introduction

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, 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 et al., 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 (ISC) 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., 2010; Hasson et al., 2004). 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 (Honey et al., 2012; Jääskeläinen et al., 2008; Regev et al., 2013; Wilson et al., 2007).

Research using the ISC 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 paper, 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 default mode network (DMN; Finnet 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; Nguyen et al., 2019; Nummenmaa et al., 2012; 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, dlPFC, anterior cingulate, amygdala, and insula (Ahn et al., 2014; Jost & Amodio, 2012; Kanai et al., 2011; Kaplan et al., 2007; Knutson et al., 2006; Van Bavel & Pereira, 2018; Westen et al., 2006). 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 functional magnetic resonance imaging (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 functional near infrared spectroscopy (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 and medial prefrontal cortex). 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 lateral and medial prefrontal cortical regions. Lateral prefrontal regions were selected due to previous associations of dorsolateral prefrontal cortex (dlPFC) with counterarguing behavior (Liu et al., 2020; O’Donnell et al., 2018). As part of the DMN, medial prefrontal cortex (mPFC) was selected due its association with social cognitive processes: A large body of evidence suggests that ventromedial cortex (vmPFC) is associated with affective processing, anteromedial prefrontal cortex (aMPFC) 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 paper, 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 IIACSS (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 4-5 minute YouTube-style videos of Arabic speakers discussing their stance on the abortion issue in two 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 & 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 Preprocessing

Acquisition. Participants were scanned using two NIRSport functional near infrared spectroscopy (fNIRS) units (NIRx, Los Angeles, CA), with a layout of 20 channels, comprised 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 U.S. 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).

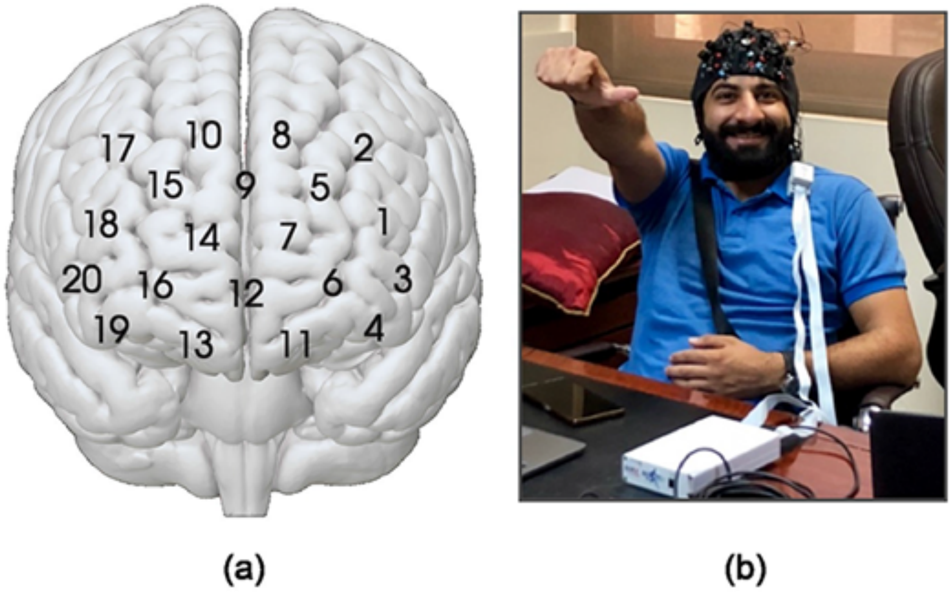


Figure 1. (a) Locations of 20 NIRS channels, which are formed between adjacent sources and detectors. (b) Experimental setup, showing participant fitted with fNIRS cap in the mobile laboratory, which was established in a market research company's office space.

Preprocessing. Prior to data preprocessing, 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 prescreening (*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 preprocessed using a customized fNIRS preprocessing 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, 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; Burns et al., 2019). Research has shown that oxygenated hemoglobin (HbO) has a stronger signal-to-noise ratio compared to deoxygenated hemoglobin (Hb) (Strangman et al., 2002). Furthermore, the HbO 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 uxing 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:

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

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 (# 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.

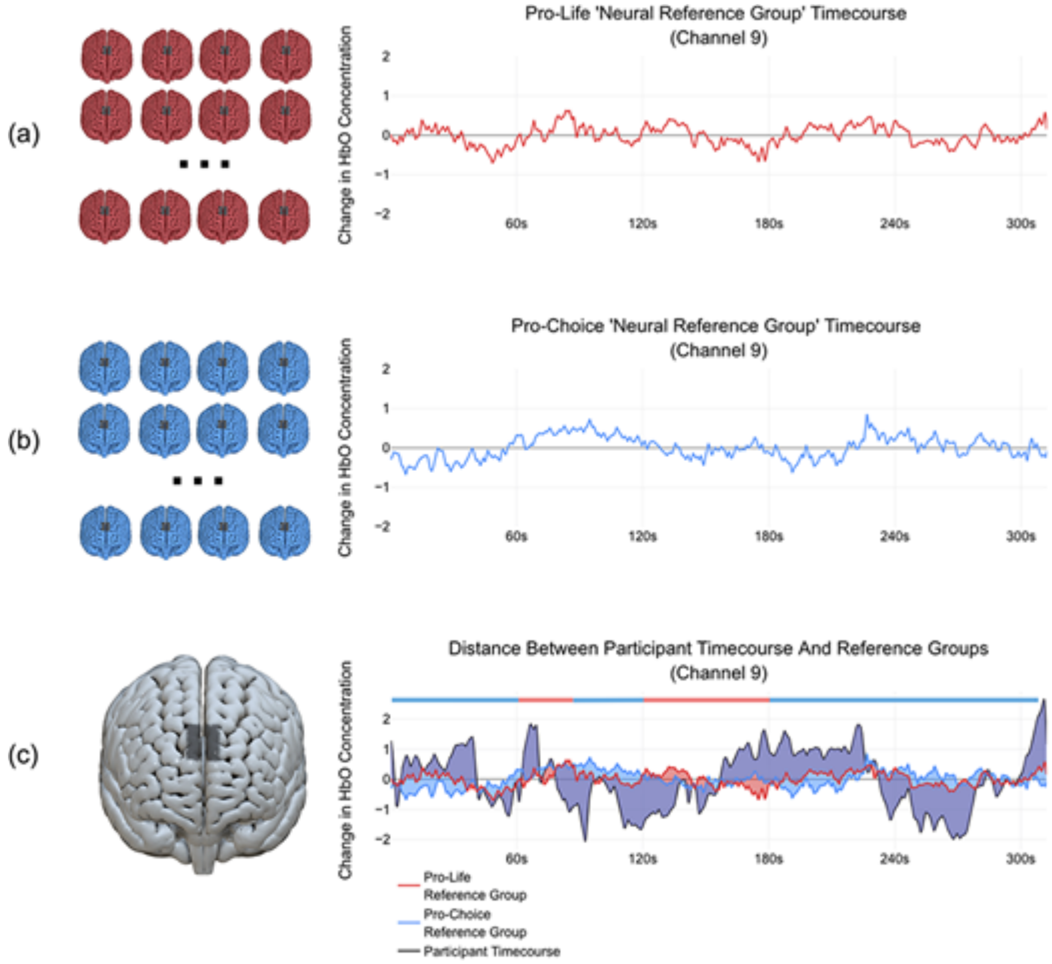


Figure 2. Depiction of the neural reference group classification approach. (a) Neural timecourses from channel 9 for participants holding a pro-life stance are averaged together to form a pro-life neural reference group timecourse. (b) Timecourses for participants holding a pro-choice stance are averaged 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.

Following Yeshurun et al. (2017), classification analyses were conducted on a channel-by-channel basis in regions of interest 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, 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 (# of participants correctly classified/# 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 1,000 times within each channel. Final classification accuracy scores were computed as the average accuracy score from all 1,000 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 ANOVAs, 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, η^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 FDR correction with a *q* criterion of 0.05 (Benjamini & Hochberg, 1995), which was used due to the large number of tests across videos and channels (2 videos x 20 channels = 40 tests). No other channels for either video showed significantly different Euclidean distances between the two groups.

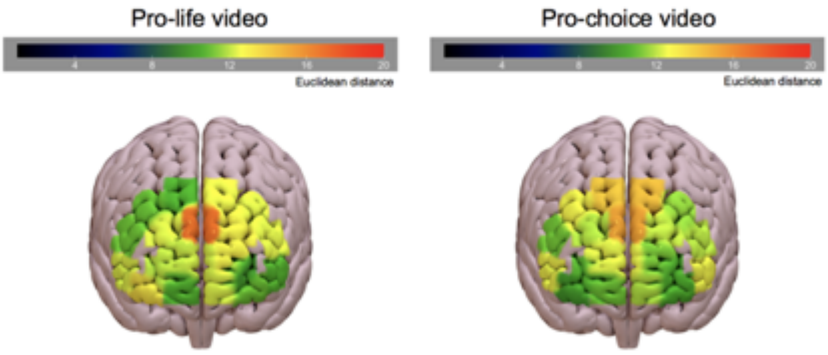


Figure 3. For each video within each channel, group-level differences between pro-life and pro-choice participants were computed as the Euclidean distance between the mean timecourse for each group. These

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.

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 + 1SD) to serve as regions of interest (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 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=.106) for channel 9 and 50.33% (p=0.238) for channel 10.

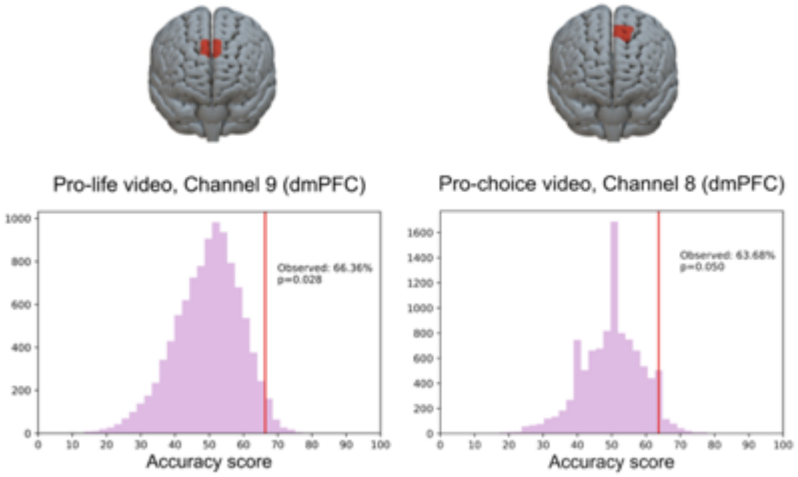


Figure 4. Classification accuracy for channels (in dmPFC) that could distinguish between partisan groups at above-chance levels for the pro-life (left) and pro-choice (right) videos. For each video, the observed classification accuracy is shown relative to a null distribution of accuracy scores generated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

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 a region of interest 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.

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 (Bacha-Trams et al., 2018; Chen et al., 2020; Finn et al., 2018; Finn et al., 2020; Lahnakoski et al., 2014; Nguyen et al., 2019; Yeshurun et al., 2017). 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 & Frith, 2006; Lieberman et al., 2019; Mitchell, 2009). Prior fMRI and fNIRS studies have also demonstrated a positive association between dmPFC activity and perceptions of the effectiveness of persuasive messages (Burns et al., 2019; Falk et al., 2010; Falk et al., 2013; Klucharev et al., 2008). 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 default mode network regions, such as the inferior parietal lobule (IPL) and inferior parietal and temporoparietal junction (TPJ). 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 (Ahn et al., 2014; Jost & Amodio, 2012; Kaplan et al., 2007; Knutson et al., 2006; Van Bavel & Pereira, 2018; Westen et al., 2006).

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 U.S. which has a larger sample size.

Nevertheless, accuracy rates of 66% and 63% in a binary classification is 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 socio-political 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 (Dikker et al., 2017; Dumas et al., 2010). 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.

References

Ahn, W. Y., Kishida, K. T., Gu, X., Lohrenz, T., Harvey, A., Alford, J. R., ... & Montague, P. R. (2014). Nonpolitical images evoke neural predictors of political ideology. *Current Biology*, *24*(22), 2693-2699.

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *NeuroImage*, *145*, 137-165.

Bacha-Trams, M., Alexandrov, Y. I., Broman, E., Glerean, E., Kauppila, M., Kauttonen, J., ... & Jääskeläinen, I. P. (2018). A drama movie activates brains of holistic and analytical thinkers differentially. *Social Cognitive and Affective Neuroscience, 13*(12), 1293-1304.

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., Dreber, A., Huber, J., Johannesson, M., & Avesani, P. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 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), 628-636.

Burns, S. M., Barnes, L. N., McCulloh, I. A., Dagher, M. M., Falk, E. B., Storey, J. D., & Lieberman, M. D. (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.

Burns, S. B., Ames, D. L., Tan, K., Katzman, P., Dieffenbach, M. C., Lieberman, M. D. (2021) ‘I did' versus 'You Should': Exploring Neural Responses to Different Persuasive Attempts. *Manuscript in preparation.*

Chen, P. H. A., Jolly, E., Cheong, J. H., & Chang, L. J. (2020). Intersubject representational similarity analysis reveals individual variations in affective experience when watching erotic movies. *NeuroImage*, 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), 2808-2821.

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), 1742-1752.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., ... & Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, *27*(9), 1375-1380.

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (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*, 261-269.

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– 8424.

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 –1242.

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., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. A. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, 116828.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531-534.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40-48.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634-1640.

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), 15277-15283.

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–D298.

Jääskeläinen, I. P., Koskentalo, K., Balk, M. H., Autti, T., Kauramäki, J., Pomren, C., & Sams, M. (2008). Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. *The Open Neuroimaging Journal*, *2*, 14-19.

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), 677-680.

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,* 353–366.

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), 3863-3868.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & 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., Hyönä, J., Hari, R., Sams, M., & Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, *100*, 316-324.

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), 559-563.

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 & Biobehavioral Reviews*, *99*, 311-328.

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), 1437-1447.

Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1309-1316.

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), 667-685.

Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, *184*, 161-170.

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), 9599-9604.

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., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825-2830.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron, 72*(5), 692-697.

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), 15978-15988.

Saalasti, S., Alho, J., Bar, M., Glerean, E., Honkela, T., Kauppila, M., ... & Jääskeläinen, I. P. (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), 842-851.

Spunt, R. P., & Adolphs, R. (2014). Validating the why/how contrast for functional MRI studies of theory of mind. *NeuroImage*, *99*, 301-311.

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), 719-731.

Tei, S., Kauppi, J. P., Fujino, J., Jankowski, K. F., Kawada, R., Murai, T., & Takahashi, H. (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), 213-224.

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), 1947-1958.

Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2007). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, *18*(1), 230-242.

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same story, different story: the neural representation of interpretive frameworks. *Psychological Science*, *28*(3), 307-319.

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.

Chapter 3 - Leveraging Differences to Bring People Together: Using Neural Synchrony to Detect Polarized Thinking and Evaluate Open-Mindedness Interventions

Abstract

Neural synchrony analyses have traditionally been conducted to identify commonalities between individuals' neural responses. However, recent research shows that synchrony analyses can also be used to detect when groups of people *see* the world differently, which manifests as 'neural polarization' between the groups. In this paper, we report on the first ever functional near infrared spectroscopy (fNIRS) study to measure neural polarization throughout the mentalizing network. Further, we demonstrate how neural synchrony measures can be used to assess the impact of interventions that aim to encourage open-mindedness. During the study, liberal and conservative participants watched political videos about gun control while having their brains scanned. Prior to scanning, some participants completed a self-affirmation intervention, which was meant to increase their open-mindedness. Using neural synchrony analyses, we could detect that the self-affirmation intervention had significantly impacted participants' neural responding in the mentalizing network. Moreover, using a classification procedure called the 'neural reference groups approach,' we could predict whether or not individuals had gone through the intervention at above-chance accuracy levels.

Introduction

Over recent decades, the United States has seen a rapid rise in polarization, with both sides of the partisan divide showing increasing antipathy toward the other (Pew Research Center, 2014). In response to this concerning trend, researchers have attempted to better understand this divide and develop solutions to reduce it. Traditionally, neuroscientists have taken more of a basic science approach, leveraging neuroimaging tools to describe how partisan thinking manifests in the brain. For instance, recent research has discovered that partisans' brains respond differently to the same political information. This phenomenon has been called 'neural polarization' due to the fact that partisans' polarized views cause them to have more similar neural responses to other members of their political group and more distinct responses from the opposing side (Moore-Berg et al., 2020). However, recent advances in neuroimaging technology and analytical techniques have enabled neuroscientists to move beyond merely describing neural polarization; it is now possible for neuroscientists to complement this foundational research with applied intervention work that can tackle the problem as well.

One neuroscience technique in particular, intersubject correlation (ISC), has been fundamental in terms of enabling neuroscientists to contribute to the political polarization literature. This technique was developed as a method for measuring the extent to which individuals' brains fluctuate in similar ways as they have naturalistic experiences such as watching a movie or having a conversation. Studies have found that individuals show robust ISC – or neural synchrony – in a wide range of higher-order cognitive regions and low-level sensory regions (Hasson et al., 2004; Hasson et al., 2012; Nastase et al., 2019). Furthermore, research has found that people who are similar to one another (e.g., who are friends) have more synchronized neural responses (Parkinson et al., 2018). Building on this idea, researchers assumed that if like-minded individuals have greater neural synchrony, then people who have different mindsets should show differences in their neural processing. At first, studies tested this idea by manipulating participants' mindsets experimentally, directing their attention to different components of a scene or framing a narrative differently to lead to different comprehensions of it (Lahnakoski et al., 2014; Bacha-Trams et al., 2017; Yeshurun et al., 2017). Since then, studies have also examined how naturally-occurring trait differences manifest in differential neural processing (Finn et al., 2018; Nguyen et al., 2019; van Baar et al., 2021). In work most relevant to political polarization, researchers have identified how partisans show different neural responses when they are watching the same political videos (Dieffenbach et al., 2021; Leong et al., 2020). Thus, these studies provide neural evidence to suggest that two individuals can look at the same stimulus but ultimately *see* it very differently (Lieberman, under review).

Building upon the discovery that synchrony analyses could reveal group differences, researchers set out to test whether it would be possible to predict group membership at the individual level. Indeed, using a technique called the 'neural reference groups approach,' researchers have been able to categorize individuals as having one mindset versus another (e.g., interpreting a story in a certain way or holding a certain opinion) at above-chance accuracy levels (Dieffenbach et al., 2021; Yeshurun et al., 2017). This technique involves collecting neural responses from individuals who fall into two different 'mindset' groups (e.g., liberals and conservatives) as they watch the same video. Then, an average neural timecourse is computed for each group, which serves as a reference for future individuals to be compared to. Individuals can be classified as having one mindset or the other based on whether their brain responses are more similar to one neural reference group or the other.

In Dieffenbach et al. (2021), we found that individuals' partisan stance could be predicted at above-chance accuracy in the dorsomedial cortex (DMPFC). This region is part of the brain's mentalizing network, a system of neural regions that are involved in thinking about others' mental states (Frith & Frith, 2003). Similarly, other studies have been able to predict participants' experimentally manipulated mindsets at above chance from neural responses in the mentalizing network (Yeshurun et al., 2017; Lahnakoski et al., 2014).

Given that previous research indicates it is possible to detect differences between the neural responses of people who belong to different groups, and even to predict group membership, a natural extension of this work is to use neural synchrony techniques to assess the impact of interventions that aim to reduce group differences. In this paper, we propose that neural synchrony analyses can 'come full circle' – from describing similarities, to describing and predicting polarized responses, to measuring the impact of strategies to reduce polarization between groups. In a first-of-its-kind study, we apply neural synchrony analyses to measure the impact of an intervention that aims to reduce polarized thinking, with particular focus on the mentalizing network.

Our study had two primary aims. First, we attempted to replicate and extend our prior work (Dieffenbach et al., 2021), which involved measuring neural polarization between partisans using functional near infrared spectroscopy (fNIRS), a portable neuroimaging technology that measures blood oxygenation levels in the brain similar to functional magnetic resonance imaging (fMRI). Given that this prior work could only capture brain data in the prefrontal cortex, we wanted to replicate this study in another sample with full cortical (i.e., "whole-brain" in fNIRS terms) coverage. No research to date has examined whether fNIRS can detect partisans' differential neural responding, beyond the prefrontal cortex, when they are watching political videos. We predicted that we would find significant neural polarization between liberals and conservatives. Specifically, we predicted that each partisan group would show greater synchrony with other members of their own political group and less synchrony with partisans on the opposite side.

Our second aim was to explore whether neuroimaging can be used to detect if an open-mindedness intervention has impacted how someone *sees* political information. When it comes to measuring the efficacy of open-mindedness interventions, self-report can be subject to experimenter demand effects, social desirability biases, and also a lack of introspective awareness. Therefore, we reasoned that neuroimaging could add value by providing a signal of whether or not an open-mindedness intervention has impacted a person's mindset directly after it has been administered. Just as partisans' prior beliefs shape how they perceive political information, we reasoned that their present level of open-mindedness would also influence their subjective construals when watching political videos. We predicted that a self-affirmation intervention, which has been shown to reduce defensive responding against counterattitudinal information (Cohen et al., 2007), would change the way that participants interpreted and experienced videos in which members of the opposing party shared their viewpoints. We hypothesized that we would be able to detect this change to their *seeing* (i.e., subjective construals) by detecting a shift in participants' neural responses. We predicted that we would find significant neural polarization between the intervention and control conditions, such that the brain responses of participants in each condition would form separate clusters, which would indicate that the intervention had altered the intervention participants' neural processing.

Throughout this paper, we focus our analyses on the mentalizing network (specifically MPFC = medial prefrontal cortex and bilateral TPJ/IPL = temporoparietal junction/intraparietal lobule). This network has been a fundamental region of interest in the neural synchrony literature. A large body of evidence suggests that the mentalizing network undergirds our ability to effortlessly integrate multiple external and internal inputs to form a pre-reflective, conscious experience (Lieberman, under review). In a recent review, Yeshurun, Nguyen, and Hasson (2021) discuss how the mentalizing network (or default mode network – DMN) gives rise to meaning-making by integrating individuals' prior beliefs and worldviews with contextual information and sensory inputs. According to these researchers, "Shared neural activity at the top of the processing hierarchy, in the DMN, naturally arises from the tendency of social brains to align thoughts and actions. Conversely, the same situation can have markedly different meanings associated with different actions across different contexts. Thus, DMN representations must differ between people who perceive and act differently in a given situation." For this reason, we investigate whether the mentalizing network can reveal neural polarization between partisans and also reflect the impact of a mindset shift intervention.

To do so, we recruited participants into three conditions: a liberal 'control' group, a liberal 'intervention' group, and a conservative 'control' group. Given that we conducted the study at a university where the majority of the student body is liberal, we decided to test the intervention effects on liberals only. All participants watched four YouTube-style videos in which a liberal or conservative shared their political opinions on gun control for approximately five minutes. fNIRS data was collected while participants watched the videos. Prior to watching the videos, participants completed a writing exercise. Participants in the intervention condition completed a self-affirmation task, which involved writing about their most important values.

Our primary analyses involved conducting within- versus between-group neural synchrony analyses on the mentalizing network. We complemented these synchrony analyses by conducting neural reference groups analyses in the mentalizing network in order to attempt to predict group membership at the individual level. Specifically, we conducted analyses comparing conservatives to all liberals (affirmed and control combined), as well as analyses that compared affirmed and control liberals. We hypothesized that across the different analyses, groups would show greater synchrony within their own group and less synchrony with the other group. We reasoned that this neural polarization would reflect differences in how they were experiencing and interpreting the political videos.

Method

Participants

A total of *N*=146 participants were recruited from UCLA's campus and the general Los Angeles community. This sample size was selected in order to have approximately 40 usable participants in each of the three conditions. Eligibility criteria included being over the age of 18, right-handed, and fluent in English. Participants were required to have lived in the U.S. since childhood in order to have had adequate exposure to American politics to understand the context of the stimuli presented. Participants were pre-screened and selected based on their political affiliation and stance on gun control in order to meet the recruitment sample criteria.

For analyses, participants were classified into attitudinal groups based on self-report scales in which they indicated whether or not they agreed with the political opinion advocated in each video. Due to this paper's focus on partisanship, participants were included in this paper's analyses only if their agreement ratings were consistent and clearly partisan (e.g., a participant would be categorized as a liberal if they agreed with both liberal arguments and disagreed with both conservative arguments). *N*=101 participants fulfilled this criterion of demonstrating clear partisan alignment. Further, 10 participants were excluded from the sample due to a period of time when the fNIRS acquisition machine was malfunctioning, such that stimulus timings were not recorded. In total, the final sample for analysis was *N*=91, which included *n*=30 conservatives, *n*=33 control liberals, and *n*=28 affirmed liberals.

fNIRS acquisition

Participants were scanned using a NIRScout fNIRS rig with a layout of 108 channels (comprising 32 light sources and 32 detectors), which achieves 'whole-brain' cortical coverage (**Figure 1**). This layout was created in accordance with the 10-10 UI external positioning system to ensure consistency across head sizes. Participants had their head sizes measured, and then were fitted with caps that affixed the optodes to the scalp. Raw light intensity data was collected at a sampling rate of 1.95 Hz at wavelengths of 760 and 850 nm.

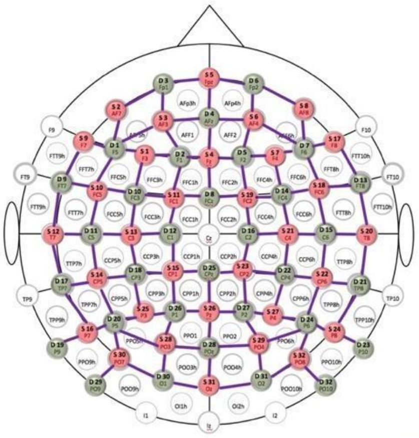


Figure 1. fNIRS optode set-up: a full-head layout in the 10-10 system, with 32 sources and 32 detectors comprising 108 data channels.

Political video stimuli

Participants watched four YouTube-style videos of English speakers who discussed their stances on gun control while facing the camera. In half of the videos, speakers (one male, one female) expressed a liberal, pro gun control stance. In the other two videos, the speakers (one male, one female) expressed a conservative, anti gun control (or pro gun rights) stance. Scripts for the videos were written by the research team (see Appendix) and recorded by young adult actors so that the videos would be of similar lengths and contain similar amounts of both reasonable and logically questionable material. Qualitative participant feedback suggested that they perceived these videos to be real and had strong reactions to them.

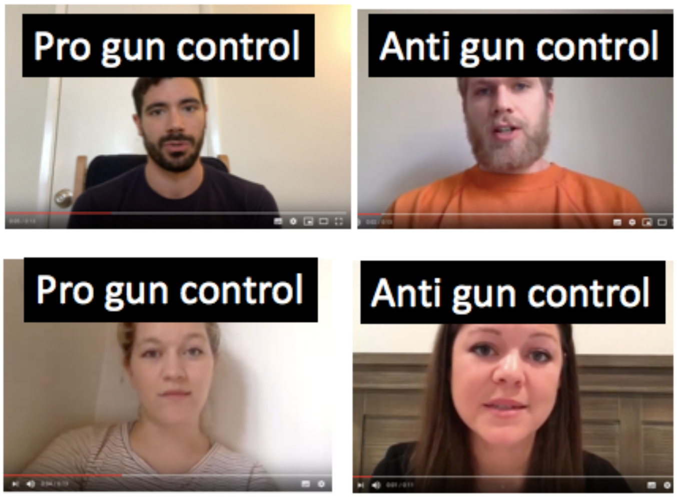


Figure 2. Still frames taken from 4 YouTube-style videos that were created by the research team and recorded by hired actors to serve as naturalistic stimuli for the main task.

Procedure

Prior to the lab session, participants completed a questionnaire that contained demographic questions, individual difference measures, questions about their attitudes toward gun control, and items measuring their openness with regards to alternative viewpoints on gun control (see Appendix for details). Upon arriving at the lab, participants received instructions and did a brief practice of one of the scanning session tasks. Participants were then fitted with the fNIRS cap. Afterwards, participants completed a brief writing task on the computer, which began with a values ranking task followed by a control or intervention writing prompt (see Appendix).

Self-Affirmation Manipulation. Participants in the intervention condition completed a self-affirmation exercise. They described three or four personal experiences in which their top-ranked value was important to them and made them feel good about themselves. Next, participants selected one of those experiences and wrote a short story describing the event and their feelings at the time. They were instructed to focus on their thoughts and feelings and not to worry about spelling, grammar, or how well written it was. This self-affirmation exercise was selected because prior work had shown that it was effective at improving receptivity to opposing political views (Cohen et al., 2000).

Participants in the control condition wrote about what they had eaten in the past 48 hours, which was also used as a control by Cohen et al. Although some affirmation studies ask participants to write about their lowest-ranked value, we suspected that such a control task might cause participants to reflect on other values, which might have an unintended self-affirming effect. Furthermore, we wanted the control condition to be more naturalistic and generalizable, such that differences between the self-affirmation and control conditions could provide confidence that the self-affirmation intervention could be effective in the real world. Following the writing task, participants described their mood and reported on their level of self-esteem as manipulation checks. Then, they completed items that re-assessed their opinion on gun control and open-mindedness toward alternative opinions on gun control.

Scanning Tasks. After completing the writing task, participants watched the political videos while being scanned with fNIRS. Prior to each video, participants were given instructions that they were about to watch a YouTube-style video, and that they should pay close attention, as they would be answering four questions about it afterwards. Each video was then shown in a separate scanning run. Following each video, participants provided ratings on: (1) how much they liked the person in the video, (2) how much the person's argument bothered them or made them angry, (3) how reasonable or unreasonable they thought the argument was, and (4) how logical the argument was. Later in the session, participants also completed a longer questionnaire, during which they indicated whether or not they agreed with the opinions in each video. Other measures included in the full questionnaire were not analyzed in the present study. Following the main task, participants also completed two functional localizer scans, which were not examined in the present study. The localizers included were the "Why-How task," a well-validated localizer of the brain's mentalizing system (Spunt & Adolphs, 2014), and a "counter-arguing task" developed by researchers at University of Pennsylvania and UCLA (O'Donnell et al., 2018).

Data Analysis

Behavioral Analysis. As a manipulation check, we conducted multilevel models to examine whether liberals and conservative evaluated the videos differently (between-subjects factor: partisan group, within-subjects factors: video, with a random intercept for each participant). We conducted analyses for the conservative and liberal videos separately, given that that is also how we examined the neuroimaging data. We expected that liberals would rate the liberal videos more positively than conservatives, and that conservatives would rate the conservative videos more positively than liberals. We also assumed that participants' responses would fall on either side of each scale's midpoint for each video, which would suggest that the video was sufficiently polarizing.

Next, we conducted analyses to measure the impact of the self-affirmation intervention on participants' self-reports. We conducted ANOVAs to examine the effects of intervention on self-esteem and mood, which were used as manipulation check items. We then conducted multilevel analyses to explore the impact of the intervention on participants' evaluations of the videos (between-subjects factor: intervention versus control group, within-subjects factors: video, with a random intercept for each participant). We predicted that self-affirmation liberals would provide more positive ratings with regards to the videos that they disagreed with as compared to the control liberals (i.e., the conservative videos). Finally, we conducted ANCOVAs to explore whether the intervention changed participants' gun control attitudes and openness to other viewpoints, while controlling for baseline attitudes/openness. We expected to see no changes in participants' attitudes toward gun control in response to the intervention; however, we predicted that participants who went through the self-affirmation (versus control) exercise would report being more open following the intervention.

Imaging analysis

Preprocessing. We first truncated the neural timecourses in order to remove timepoints that occurred before stimulus presentation began. Then, we conducted a series of steps in accordance with a typical fNIRS preprocessing pipeline. First, we detected and removed channels in which detector saturation exceeded more than two seconds, or in which the power spectrum variation was too high, as measured by a threshold quartile coefficient of dispersion of 0.6 - 0.03*the sampling rate (see Burns, 2020). Next, the neural data were corrected for motion artifacts using a PCA algorithm, subjected to a bandpass filter of 0.005-0.5 Hz, converted into concentrations of oxygenated hemoglobin using the Modified Beer Lambert Law, and z-scored. The video timecourses were then concatenated by video type, forming one liberal video timecourse and one conservative video timecourse. A mentalizing network region of interest (ROI) was generated from channels in VMPFC, DMPFC, and bilateral TPJ/IPL. (Figure 3). Region of interest locations were then translated into Montreal Neurological Institute (MNI space) for comparability with fMRI findings using a probabilistic mapping method (Singh et al., 2005). fNIRS data were converted to *.img files using xjView (http://www.alivelearn.net/xjview/), and then projected onto a 3D cortical surface using the software Surf Ice.

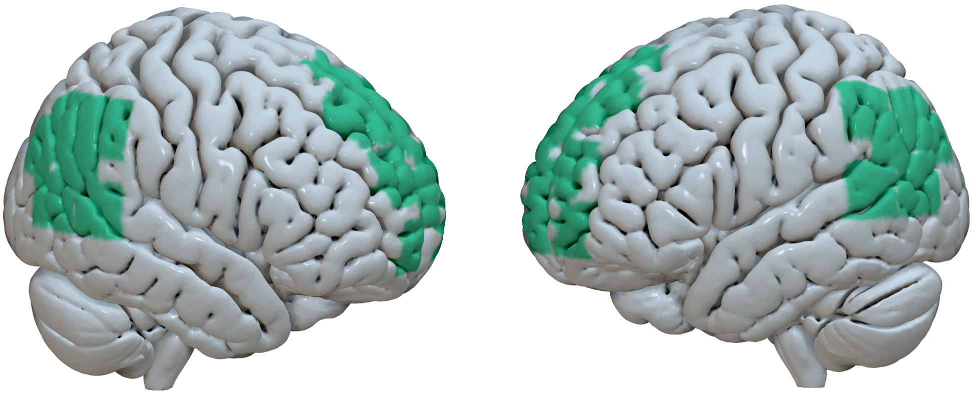


Figure 3. A cortical projection visualizing the mentalizing network ROI, which was defined a priori. This network consisted of DMPFC, VMPFC, and bilateral TPJ/IPL.

Computing neural synchrony values. Neural synchrony, or intersubject correlation (ISC) values, were computed for each video timecourse using the leave-one-out approach (Nastase et al., 2019). To apply this approach to a given channel or region of interest, a Pearson correlation is computed between a participant's timecourse and a 'neural reference group timecourse' — i.e., a timecourse that has been averaged across a group of subjects who share certain similarities (Dieffenbach et al., 2021). When the timecourse of a particular participant (x_P) is compared to the group (*Group),* their timecourse is left out from the group average: $\mathrm{ISC}_{\mathrm{owngroup}}=r(x_P, x_{-Group \neq P})^2$.

Using this approach, we computed ISC values for each participant with other members of their condition (within-group synchrony) and with members of the other condition (between-group synchrony) within the mentalizing network. We applied a Fisher z transformation (the inverse hyperbolic tangent function 'arctanh') to the ISC values prior to running any parametric tests or averaging.

Within-versus-between group intersubject correlation analyses. We then conducted within- versus between-group synchrony (ISC) analyses within the mentalizing network for each video type (liberal and conservative). We first conducted these tests on conservatives and liberals (collapsing across intervention condition) to test for neural polarization between the partisan groups. Then, we conducted analyses on the affirmed versus control liberals. Specifically, we conducted one-tailed, paired t-tests on the transformed ISC values to identify regions of the brain where participants showed greater neural synchrony with their own neural reference group (within-group synchrony) versus another reference group composed of participants in a different condition (between-group synchrony): $H_0: ISC_{within} = ISC_{between}$. Previous research suggests that it is acceptable to conduct paired t-tests on leave-one-out ISC data, despite the fact that neural ISC data follows a power law and is non-independent, as these issues are likely to impact both groups in a similar way (Nastase et al., 2019). According to the researchers, the results produced by paired t-tests are robust and generally analogous to those produced by non-parametric tests.

Based on our prior work conducting these analyses on partisan fNIRS data, we predicted that liberals and conservatives would show significant neural polarization in the mentalizing network for both video timecourses (Dieffenbach et al., 2021). We also predicted that we would see significant neural polarization in the mentalizing network between the control and affirmed liberal groups, which would indicate that the intervention had been effective at changing the affirmed liberals' neural processing. We predicted that the intervention would have a greater impact on how liberals processed counterattitudinal (i.e., conservative) videos, given that self-affirmation is thought to reduce defensive responding against potentially threatening information (Sherman & Cohen, 2002).

Neural reference group classification analyses. Whereas the within versus between group ISC analyses were conducted at the group level, they do not predict the group membership of particular individuals. For this reason, we also conducted classification analyses on the pre-defined mentalizing network ROI using the neural reference groups approach from Dieffenbach et al. (2021). This approach involves using a nearest centroid classifier to categorize a participant into one group or the other based on whichever neural group average they are "closer to" in Euclidean space. We conducted this procedure using a leave-two-out approach, where one member of each group was left out as testing data while the remaining participants were used as training data. We selected this classifier because it requires minimal assumptions in terms of parameters and has been effective at making predictions in prior neuroimaging studies (Dieffenbach et al., 2021; Yeshurun et al., 2017).

We implemented these classification analyses using sci-kit learn, a machine learning library in Python (Pedregosa et al., 2011). We created a pipeline that consisted of an undersampling technique followed by the nearest centroid classifier. We used the imblearn package's ClusterCentroids technique in order to create groups of equal sample sizes (Lemaître et al., 2017). Notably, by including this undersampling technique in a pipeline, it was only performed on the training data and not on the test data for each iteration of the leave-two-out procedure, preventing any 'data leakage' from inflating the classifier's accuracy.

To obtain a measure of classification accuracy, we compared the condition (e.g., affirmed or control) that the classifier predicted against the actual condition that the participant had been in. Classification accuracy scores were computed by dividing the number of participants that were classified correctly by the total number of participants included (# of participants correctly classified/# of total 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 100 times. A final classification accuracy score was computed as the average accuracy score from all 100 repetitions. We then tested the significance of the accuracy score using permutation testing, where we shuffled participants' group membership. Classification accuracy scores were obtained for data shuffled over 1,000 repetitions and compared to the accuracy scores for the real dataset (number of null values larger than the real value + 1/1,000).

Results

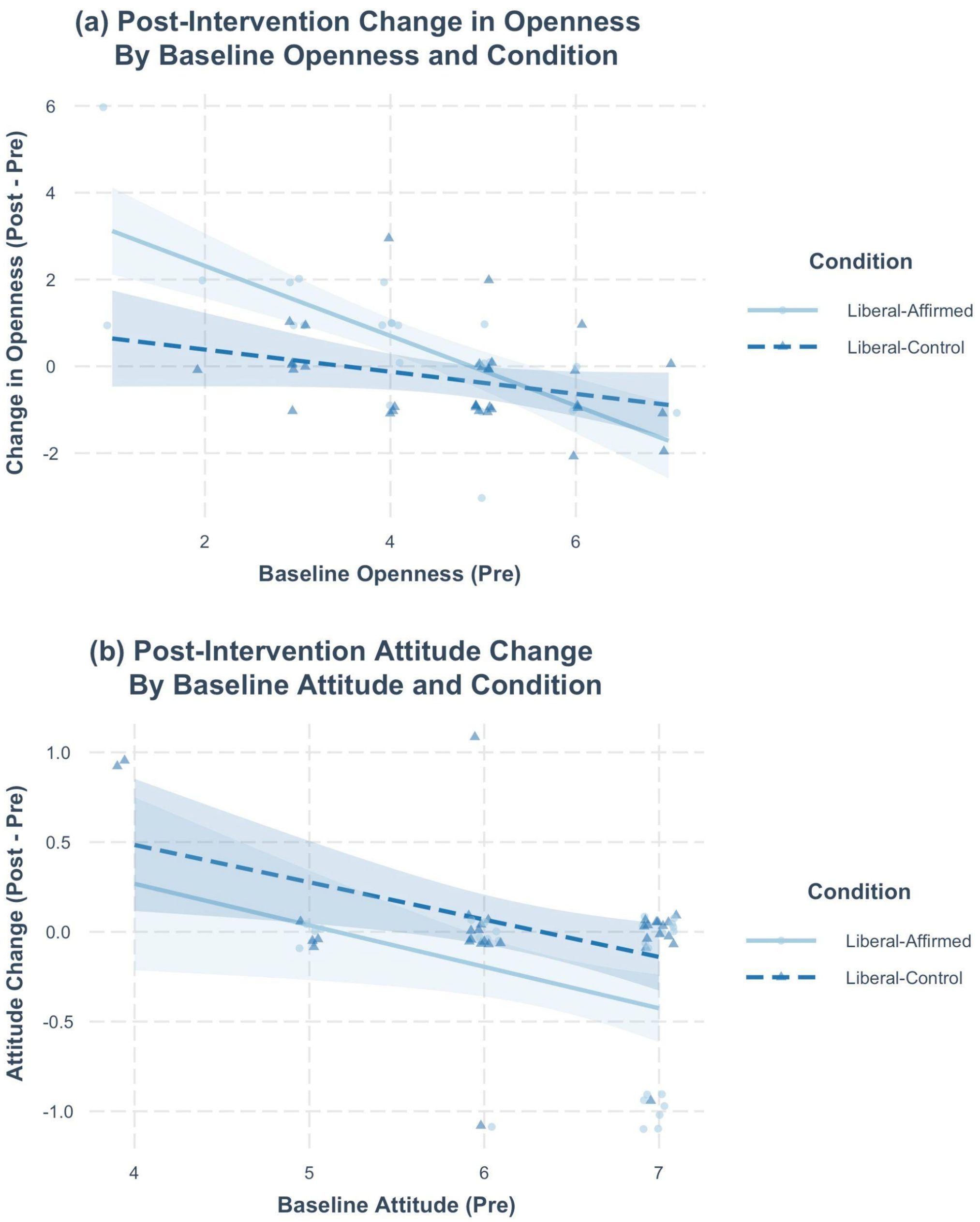
Behavioral Findings

Differences between liberals and conservatives. We first analyzed participants' ratings of the videos to determine whether partisans responded to the videos differently. Given that the four ratings for each video were highly correlated (across the videos,α= 0.89 - 0.93), we computed a composite evaluation score for each video. We found that conservatives and liberals different significantly in their evaluations of the conservative and liberal videos (see Appendix - Figure 1). As was predicted, compared to liberals, conservatives evaluated the conservative videos more positively, B=1.950, t(87)=10.805, p<0.001) and the liberal videos more negatively (B=-1.792, t(87)=-11.666, p<0.001). Liberals' and conservatives' ratings of the video were clearly polarized for the three of the videos, with the ratings of the two partisan groups falling on either side of the midpoint. However, one of the liberal videos (Appendix - LibVid1) followed a slightly different pattern in that participants from both partisan groups provided positive ratings (i.e. ratings that fell above the midpoint).

Differences between affirmed and control liberals. First, we tested whether the two liberal groups differed in their ratings of the videos. We found no significant difference between affirmed and control liberals in their ratings of the liberal or conservative videos (respectively: B=-0.291, t(59)=-1.468, p=0.147; B=0.011, t(59)=0.074, p=0.941). Then, we examined the effect of self-affirmation on two variables that have been used as manipulation checks in prior self-affirmation studies: self-esteem and mood. We found no significant difference between affirmed and control liberals on either outcome (F(1,59)=0.504, p=0.481; F(1,59)=1.375, p=0.256).

Next, we examined whether affirmed and control liberals showed changes in their self-reported openness to alternative viewpoints on gun control after completing the writing exercise (Figure 4a). We had predicted that liberal participants who completed the self-affirmation would show greater increases in openness as compared to the liberal controls. We conducted an ANCOVA, regressing change scores in openness (post-score - pre-score) on condition, while controlling for baseline openness (pre-score). There was a significant interaction between condition and pre-session openness (B=0.550, t(59)=-2.757, p=0.008). In addition, there was a significant main effect of condition (B=3.027, t(59)=3.243, p=0.002), such that affirmed liberals showed greater increases in openness on average ($M\Delta_{\text{Affirmed}}$=0.5, SD=1.58) as compared to control liberals ($M\Delta_{\text{Control}}$=-0.303, SD=1.05). There was also a significant main effect of pre-session openness, such that liberals who were less open to begin with were more likely to increase their openness (B=-0.805, t(59)=-5.715, p<0.001). Post hoc tests of the significant interaction revealed that the relationship between pre-session openness and changes in openness were specific to affirmed liberals. Affirmed liberals who were less open-minded at baseline were more likely to show increases in openness, whereas affirmed liberals who were more open-minded at baseline showed no change or decreases in openness (*B*=-0.81, *t*(60)=-5.72, *p*<0.001). In contrast, for control liberals, who did not show significant changes in openness, the relationship between baseline openness and changes in openness was not significant (*B*=-0.26, *t*(60)=-1.81, *p*=0.08).

In addition to examining the effect of self-affirmation on participants' *openness* toward other attitudes on gun control*,* we also examined whether the two liberal groups showed changes in their *own attitudes* toward gun control after completing the writing exercise (Figure 4b). There was no significant interaction between condition and pre-session attitudes (*B*=0.023, *t*(59)=0.188, *p*=*n.s.*), and no significant main effect of condition (*B*=0.124, *t*(59)=0.157, *p*=*n.s.*). There was a significant main effect of baseline gun control attitudes, such that liberals who began with more extreme attitudes shifted toward more moderate attitudes, and those with moderate attitudes became more extreme (*B*=-0.231, *t*(59)=-2.416, *p*=0.019). The relationship between baseline attitudes and attitude change is likely an artifact of how attitudes are measured and fluctuate naturally over time. However, post hoc visual inspection of the data showed a small trend toward a greater number of affirmed liberals becoming more moderate in their views as compared to control liberals.



(a) Post-Intervention Change in Openness By Baseline Openness and Condition

(b) Post-Intervention Attitude Change By Baseline Attitude and Condition

Figure 4. Changes in self-reported measures from pre to post intervention by condition. (a) There was a significant interaction between condition and participants' baseline self-reported openness toward alternative viewpoints on gun control. Some participants in the self-affirmation condition showed increases in openness, whereas participants in the control condition showed no change. Among participants who completed the self-affirmation intervention, those who were closed-minded at baseline showed increases in openness toward alternative viewpoints. Those who were more open-minded at baseline did not benefit from self-affirmation. (b) There was no significant interaction between condition and participants' baseline attitudes on gun control. For both liberal groups, participants with more moderate attitudes were more likely to shift toward more extreme attitudes, whereas those with extreme attitudes were more likely to shift toward moderate attitudes. Out of the liberals who showed the most extreme views and shifted toward more moderate views, the majority (8) were affirmed liberals in comparison to a minority (2) of control liberals.

Neural polarization between conservatives and liberals

Neural synchrony analyses – concatenated videos. We conducted within- versus between-group synchrony analyses on the conservative and liberal video timecourses separately to identify neural polarization between conservatives and liberals. We conducted our primary analyses in the mentalizing network, and then followed up with post hoc analyses in bilateral TPJ and MPFC.

First, we conducted synchrony analyses for the conservative videos (Figure 5). When we conducted a contrast that included both conservatives and liberals in the same model, there was a marginally significant difference, such that participants had greater within-group versus between-group synchrony (*t*(90)=1.372, *p*=0.087). Next, we looked at the two partisan groups separately. We found that liberals showed significantly greater within-group versus between-group synchrony in the mentalizing network, *t*(60)=2.212, *p*=0.015. In other words, their neural responses clustered together as they were watching the conservative videos. On the other hand, there was no significant within- versus between effect for conservatives, *t*(29)=-0.695, *p*=0.754. In addition to conducting analyses in the mentalizing network, we also conducted post hoc analyses to examine the bilateral TPJ and MPFC. In examining both partisan groups in the same model, there was a marginal difference in synchrony for MPFC (*t*(89)=1.479, *p*=0.071) and no significant difference in synchrony for bilateral TPJ (*t(*89)=1.23, *p*=0.101). In post hoc tests to examine the participant groups separately, liberals had significantly greater within- versus between-group synchrony for both MPFC (*t*(60)=2.169, *p*=0.017) and TPJ (*t*(50)=1.707, *p*=0.047). There were no significant synchrony differences for conservatives.

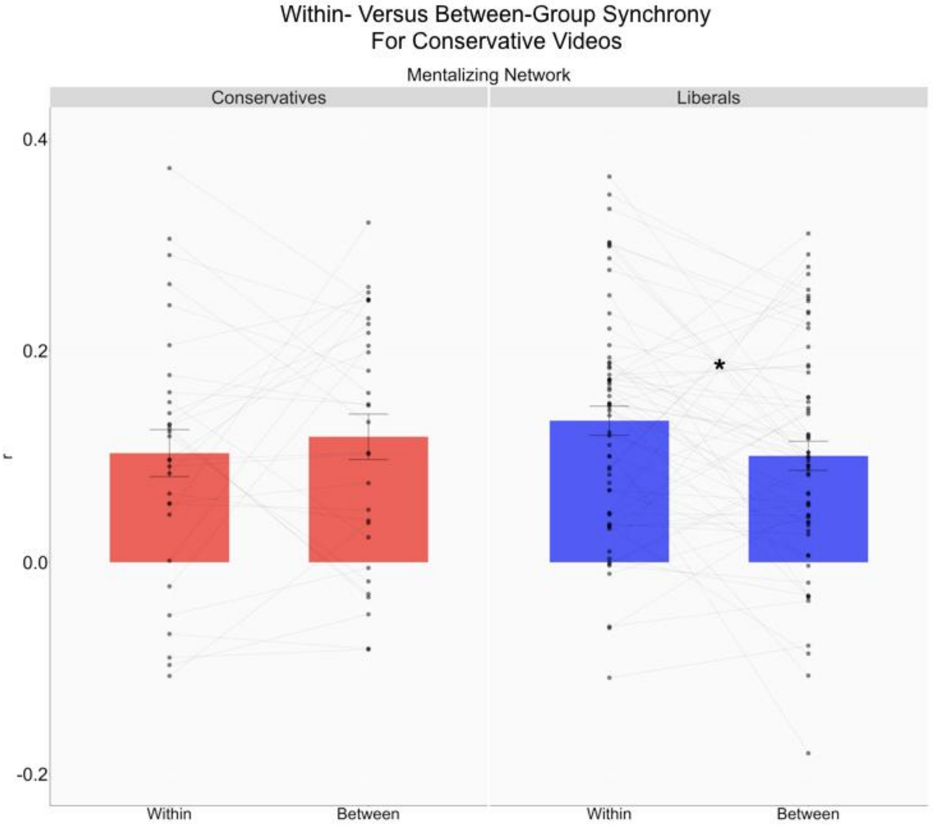


Figure 5. Within-group versus between-group synchrony analyses in the mentalizing network for conservatives (left) and liberals (right) as they were watching conservative videos. Whereas liberals showed greater within-group than between-group synchrony, conservatives showed no difference.

Then, we examined neural synchrony during the liberal videos. We found that there was no significant within- versus between-group effect in the mentalizing network, *t*(90)=-0.278, *p*=0.609. In addition, there were no significant effects for each of the individual group contrasts (for liberals, *t*(60)=0.133, *p*=0.447; for conservatives, *t*(29)=-0.650, *p*=0.740). We found no significant effects in the TPJ or MPFC for the combined or individual group analyses (*p*s=*n.s.*). Therefore, neither liberals nor conservatives showed greater within-group versus between-group synchrony while viewing the liberal videos.

Neural synchrony analyses – most polarizing videos. Given that liberals and conservatives showed more polarized responses for some of the videos (see Appendix), we also conducted a post hoc analysis to examine whether the two partisan groups would show significant neural polarization on the most polarizing video from each video type (conservative and liberal). For the most polarizing conservative video (Appendix - ConVid1), liberals and conservatives showed significant neural polarization in the mentalizing network (*t*(90)=2.268, *p*=0.019). In post hoc analyses of the individual ROIs, there was significant polarization for both MPFC (*t*(89)=1.665, *p*=0.0127) and TPJ (*t*(89)=1.665, *p*=0.050). In analyzing the partisan groups separately, this effect appeared to be driven by liberals, who showed greater within- versus between-group synchrony in the mentalizing network (*t*(60) = 2.382, *p* = 0.010) and MPFC. (*t*(60) = 2.515, *p* = 0.007). In contrast, conservatives did not show significant neural synchrony in the ROIs.

Similarly, for the most polarizing liberal video (Appendix - LibVid2), liberals and conservatives showed significantly greater within versus between group synchrony in the mentalizing network (*t*(90)=1.869, *p*=0.032). In post hoc analyses of the individual ROIs, there was a significant difference for MPFC (*t*(89)=2.496, *p*=0.007), but not TPJ (*t*(89)=0.682, *p*=*n.s.*). This effect seemed to be driven by the liberal group, just as it was for the conservative video. In examining the partisan groups separately, liberals showed marginally greater within- versus between-group synchrony in the mentalizing network (*t*(60) = 1.451, *p* = 0.076), and a significant difference in MPFC (*t*(60) = 2.31, *p* = 0.012). Conservatives showed no significant within-versus-between group differences.

Classification analyses. To complement the neural synchrony analyses, we also conducted classification analyses in the mentalizing network in order to quantify the percentage of participants whose partisanship we could classify at the individual level. From the conservative videos, we could classify participants' partisan stance with an average accuracy of 55.08% in the mentalizing network, which was not significantly different from chance (*p*=*n.s.*). From the liberal videos, the classifier had an average accuracy of 48.685%, which was also not significantly different from chance (*p*=*n.s.*). Furthermore, post hoc analyses of the two polarizing videos did not yield accuracy levels that were significantly greater than chance.

The impact of self-affirmation on neural polarization. After examining the neural polarization between partisan groups, we also looked to see if affirmed and control liberals were neurally distinguishable. We reasoned that if the intervention had impacted participants' neural processing, the two groups should show greater within-group versus between-group synchrony, with members of each group exhibiting neural responses that 'clustered together.' If the two groups showed no significant neural polarization, this would indicate that the self-affirmation intervention had not altered the brain responses of the intervention group sufficiently enough to detect a difference neurally. In particular, we predicted that the self-affirmation intervention would impact participants' responses to the conservative videos, given that prior work has shown self-affirmation can attenuate defensive responding to counterattitudinal viewpoints.

Neural synchrony analyses. For the conservative videos, which all liberal participants disagreed with, as predicted, we found evidence for neural polarization between the intervention and control groups. Participants had significantly higher levels of synchrony in the mentalizing network with other participants in their same condition (e.g., affirmed liberals with other affirmed liberals, control liberals with other control liberals) as compared to participants in the opposite condition (*t*(60) = 2.5471, *p* = 0.007, shown in Figure 6). In conducting contrasts on the groups individually, we found that the control liberals showed significantly greater within-group versus between-group synchrony (*t*(27)=2.936, *p*=0.003), whereas affirmed liberals did not show this effect (*t*(32)=0.890, *p*=0.190). We then conducted post hoc analyses on all liberal participants to investigate the subcomponents of the mentalizing network ROI (MPFC and bilateral TPJ) separately. We found significant within- versus between- group effects in both bilateral TPJ (*t*(59)=2.307, *p*=0.013) and MPFC (*t*(60)=1.835, *p*=0.036), suggesting that both ROIs had contributed to the effect seen in the overall mentalizing network analysis.

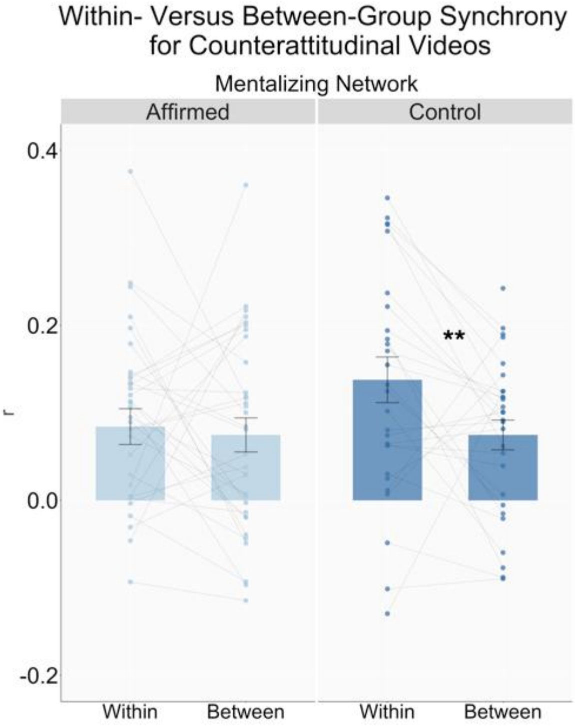


Figure 6. Within-group versus between-group synchrony for affirmed (left) and control (right) liberals in the mentalizing network when watching counterattitudinal (conservative) videos.

Next, we explored whether affirmed and control liberals demonstrated neural polarization during the liberal video timecourse, which all liberal participants agreed with. There was a marginally significant difference between participants' within- versus between-group synchrony in the mentalizing network for the liberal videos (*t*(60)=1.606, *p*=0.057). Post hoc, we examined the MPFC and bilateral TPJs separately. Participants did show significantly greater within-versus between-group synchrony in bilateral TPJ/IPL (*t*(59)=1.997, p=0.025), and a marginally significant difference in MPFC (*t*(60)=1.376, *p*=0.087).

Classification analyses. In addition to conducting synchrony analyses, we also applied the neural reference groups technique to determine if we could predict participants' experimental condition from their neural data as they watched the conservative (counterattitudinal) videos. In line with our predictions, we found that we could classify liberals into the appropriate condition with a mean accuracy score of 67.512%, which was statistically greater than chance (p=0.019; Figure 7a). We found that the classifier was more accurate in categorizing control liberals (76%) than affirmed liberals (64%) (Figure 7b). We also tested the classifier for the liberal videos, finding that it could not classify participants' stance at above chance in the mentalizing network (mean accuracy over 100 iterations = 55.06%, *p*=n.s.) This was in line with our hypotheses that the intervention would likely change neural responses to counterattitudinal but not proattitudinal information.

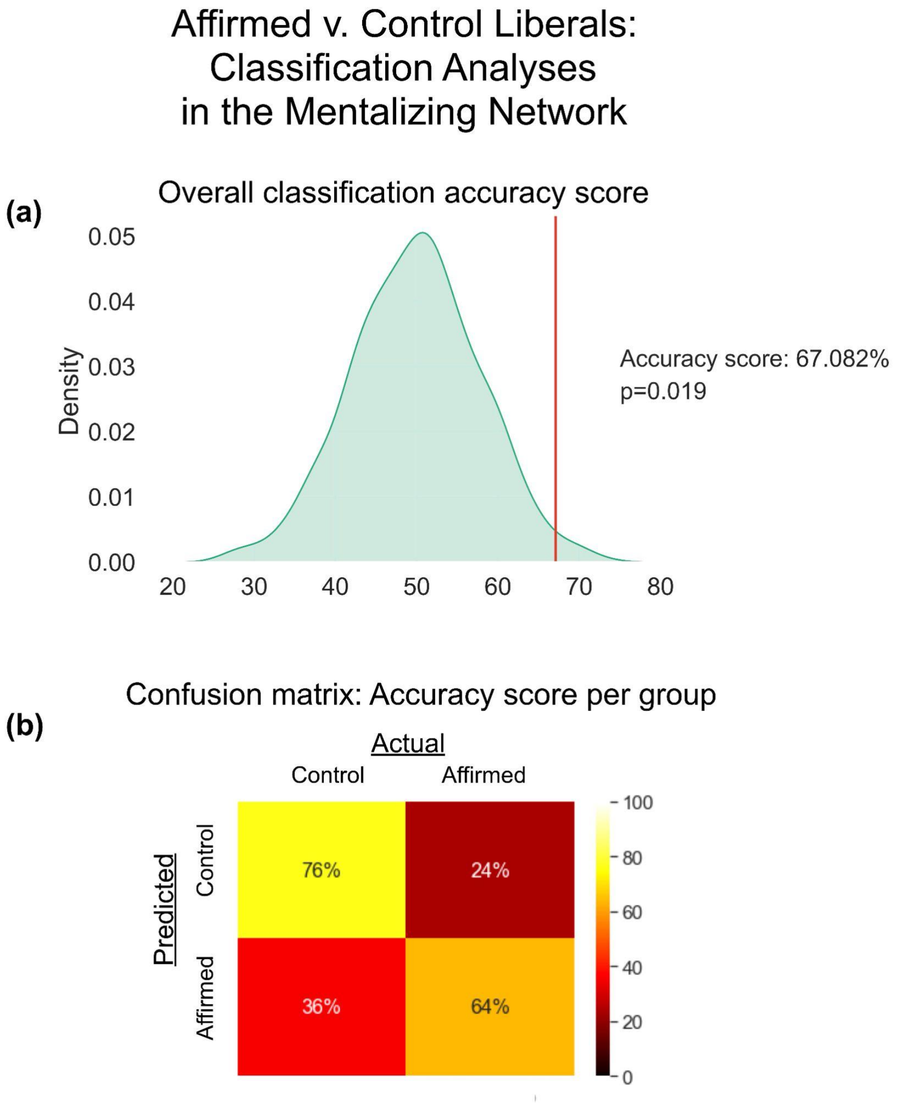


Figure 7. Classifier performance for machine learning analyses categorizing liberals as belonging to the affirmed or control conditions. (a) Displays the overall accuracy score and permutation testing that was conducted to identify a null distribution to which the actual classifier performance (the red line) could be compared. (b) Shows the classifier performance for each group separately, displaying the percentage of each group that was classified correctly versus incorrectly.

Discussion

Political ideology serves as a lens through which we see the world (Lieberman, under review; Van Bavel & Pereira, 2018; Westen et al., 2006). When people hold different viewpoints, this manifests in how their brains process political information, leading them to show greater neural synchrony with others who share their political views, and less synchrony with members of the political outgroup (Leong et al., 2020). Furthermore, preliminary work indicates that political partisanship can even be predicted using a synchrony-based approach (Dieffenbach et al., 2021). The current study, while attempting to replicate these effects in a whole-brain fNIRS study, extends them in a critical direction. With this research, we demonstrate that synchrony analyses can assess the impact of interventions that aim to shift partisans toward becoming more open-minded.

Specifically, we leveraged the neural reference groups approach (Dieffenbach et al., 2021) to determine whether a self-affirmation intervention had shifted liberal participants' subjective construals of political videos. Previous research found that self-affirmation reduced defensiveness toward others with opposing viewpoints, which is why it was selected as this study's intervention. We focused our analyses on a mentalizing network ROI that we had defined a priori, which consisted of MPFC and bilateral TPJ/IPL. We selected this network due to a large body of synchrony-based studies that suggest that it plays a key role in helping individuals to form subjective construals based upon their beliefs, present mindsets, and external inputs (Lieberman under review; Yeshurun et al., 2021).

Self-report results confirmed that the self-affirmation intervention increased openmindedness for participants who were more closed-minded at baseline. Participants who were already open-minded to begin with did not show changes in their self-reported open-mindedness. This finding is in line with previous work that found that participants who were ideologically 'hawkish' (or closed-minded toward outgroup opinions) were most likely to become more open-minded after undergoing an intervention in which they learned about cognitive biases (Nasie et al., 2014). Thus, it is possible that including participants in our sample who were already open-minded to begin with could have reduced our power to detect neural polarization between the control and affirmed groups. In the future, researchers may consider recruiting specifically for participants who are most likely to benefit from an intervention in order to increase power.

However, it is also possible that self-affirmation impacted the neural processing of all liberals in the intervention group, even if these changes were not reflected in the self-report data. The neural analyses may have been able to detect differences between the groups that could not be captured by self-report. Indeed, analyses of participants' brain data revealed neural polarization between the affirmed and control liberal groups. Liberal participants showed greater within-group synchrony with other members of their same condition (affirmed or control), and less synchrony with liberal participants in the opposite condition when they were watching videos that they disagreed with. This effect seemed mainly to be driven by the control participants, who showed greater alignment within their group. Moreover, using a machine learning classifier, we could predict whether or not individual participants had gone through the intervention at an above-chance accuracy level. We found that we could classify participants in the control condition (who had shown more within-group alignment) with a higher degree of accuracy.

This finding opens up intriguing possibilities. It suggests that neuroimaging may be able to help 'get under the hood' in assessing whether or not an intervention has had a meaningful psychological impact moments after it has been administered. It provides an opportunity for measuring individuals' subjective construals as they create them as opposed to asking them to reflect upon them after the fact. In addition, it provides an avenue for assessing whether particular individuals have benefitted from interventions. Furthermore, using neuroscience to measure the impact of interventions can also address some of the limitations of using self-report alone. Due to self-enhancement motives and/or experimenter demand effects, participants may report that they have become more open-minded after an intervention even when they have not (Meagher et al., 2015). Moreover, they might not have introspective access to be able to determine whether an intervention has affected their mindset. Their responses to alternative viewpoints, and their neural processing of them, may provide better insight as to these interventions' efficacy.

Of course, with the present analyses, we were only able to detect that self-affirmation had caused a change to the intervention participants' neural processing. We could not determine the directionality of the intervention effect from the neural data, only that it had occurred. The behavioral findings complemented the neural findings to indicate that self-affirmation led to greater open-mindedness, rather than closed-mindedness. In the future, it may be possible for synchrony-based research to yield effects without requiring any self-report. For instance, in using the neural reference groups approach, researchers could collect brain data from individuals who demonstrate a desired response (e.g., respond in an 'open-minded' way by providing positive ratings to videos depicting people sharing alternative viewpoints), and also those who demonstrate a non-desired response (e.g., respond in a 'closed-minded' way by providing negative ratings). Then, an intervention could be administered to participants who demonstrate a need for it (e.g., participants who are identified as being closed-minded) prior to having them watch videos of people with opposing viewpoints. Using out-of-sample prediction, the neural reference groups approach could identify whether specific individuals' neural responses matched the open-minded or closed-minded reference group, which would indicate whether or not the intervention had an impact in shifting their thinking patterns in the desired direction.

In addition to demonstrating that neural synchrony analyses can be used to measure the impact of interventions, another aim of the present work was to replicate previous effects that have found neural polarization between liberals and conservatives in the MPFC

(Dieffenbach et al., 2021; Leong et al., 2020; van Baar et al., 2021). Although we failed to replicate the original findings when analyzing the concatenated videos, we did replicate the original effect in conducting analyses on the most polarizing video of each video type. For the most polarizing liberal and conservative video, participants showed greater within- versus between-group synchrony in the overall mentalizing network and in MPFC. When the partisan groups were analyzed separately, it appeared that liberals were driving the neural polarization effect. Furthermore, the classification analyses did not yield above-chance accuracy scores.

There are a few reasons that we might have seen significant effects for liberals only. First, given that we had twice as many liberals as conservatives due to a feature of the study design, one possibility is that we were underpowered to detect those effects. Also, due to conducting this study on a liberal campus, it is possible that the conservative participants had greater prior exposure to opposing viewpoints than the liberal participants. It is possible that they did not experience the same level of arousal and/or negative valence when watching opposing viewpoints, having been more 'inoculated' to them through their experiences. Furthermore, conservatives reported having neutral reactions to one of the liberal videos and only reported having strong reactions to one of them.

Given that conservatives' neural responses did not become synchronized during either video type, it is perhaps unsurprising that we could not predict participants' partisan stance from either of the video timecourses. These results would suggest that future analyses that employ the neural reference groups approach should consider that the neural alignment may be heterogeneous across groups, which should impact the selection of appropriate stimuli and machine learning classifiers. According to Finn et al. (2020), certain synchrony data can follow what they call an 'Anna Karenina' (AnnaK) pattern in which people at one end of a spectrum cluster together and people at the other end show idiosyncratic responses. For these kinds of differences, it may be more appropriate to model the data looking at absolute rather than relative scores on scale items. If researchers aim to create distinguishable neural reference groups that form two distinct clusters, then it may be beneficial to select stimuli that produce similar psychological states in both groups that occur at different timepoints. For instance, videos that contain viewpoints from conservatives and liberals (e.g., a political debate or a concatenated series of many short clips) would serve to synchronize conservative and liberals with other in-group members in similar ways. Even though they would enter into similar psychological states, they would do so at different times, which would still allow the groups to be neurally distinguishable.

Conclusion

In conclusion, this study identified a new application for synchrony-based analyses: measuring the impact of interventions that aim to shift people's mindsets. In this study, we used a self-affirmation intervention to induce participants to construe opposing viewpoints in a more open-minded way. We found that the intervention and control groups showed distinguishable neural responses in the mentalizing network when they were listening to opinions that they disagreed with, such that we could detect differences between their brain responses using traditional, group-based synchrony analyses. Furthermore, we could predict whether particular individuals had gone through the intervention at above-chance levels by applying a classification technique called the neural reference groups approach to their brain responses.

REFERENCES

Bacha-Trams, M., Glerean, E., Dunbar, R., Lahnakoski, J. M., Ryyppö, E., Sams, M., & Jääskeläinen, I. P. (2017). Differential inter-subject correlation of brain activity when kinship is a variable in moral dilemma. *Scientific Reports*, *7*(1), 1-16.

Burns, S. (2020). *Neural and psychological coordination in social communication and interaction* (Order No. AAI28000230). Available from APA PsycInfo. (2431953398; 2020-51431-114). https://www.proquest.com/dissertations-theses/neural-psychological-coordination-social/docview/2431953398/se-2?accountid=14512

Cohen, G. L., Sherman, D. K., Bastardi, A., Hsu, L., McGoey, M., & Ross, L. (2007). Bridging the partisan divide: Self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. *Journal of Personality and Social Psychology*, *93*(3), 415-430. https://doi.org/10.1037/0022-3514.93.3.415.

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, *26*(9), 1151-1164.

Dieffenbach, M. C., Gillespie, G. S., Burns, S. M., McCulloh, I. A., Ames, D. L., Dagher, M. M., ... & Lieberman, M. D. (2021). Neural reference groups: a synchrony-based classification approach for predicting attitudes using fNIRS. *Social Cognitive and Affective Neuroscience*, *16*(1-2), 117-128.

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), 1-13.

Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., &

Bandettini, P. A. (2020). Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, *215*, 116828.

Frith, U. & Frith, C.D. 2003. Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.,* 358, 459–73

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science, 303*(5664), 1634-1640.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, *16*(2), 114-121.

Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., & Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, *100*, 316-324.

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), 559-563.

Leong, Y. C., Chen, J., Willer, R., & Zaki, J. (2020). Conservative and liberal attitudes drive polarized neural responses to political content. *Proceedings of the National Academy of Sciences*, *117*(44), 27731-27739.

Lieberman (under review). CEEing+: A Neurocognitive Model of Pre-Reflective Construal.

Lyons, B., Farhart, C., Hall, M., Kotcher, J., Levendusky, M., Miller, J., . . . Zhao, X. (2021). Self-Affirmation and Identity-Driven Political Behavior. *Journal of Experimental Political Science,* 1-16. doi:10.1017/XPS.2020.46

Meagher, B. R., Leman, J. C., Bias, J. P., Latendresse, S. J., & Rowatt, W. C. (2015). Contrasting self-report and consensus ratings of intellectual humility and arrogance. *Journal of*

Research in Personality, *58*, 35-45.

Moore-Berg, S. L., Ankori-Karlinsky, L. O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences*, *117*(26), 14864-14872.

Nasie, M., Bar-Tal, D., Pliskin, R., Nahhas, E., & Halperin, E. (2014). Overcoming the barrier of narrative adherence in conflicts through awareness of the psychological bias of naïve realism. *Personality and Social Psychology Bulletin, 40*(11), 1543-1556.

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), 667–685. https://doi.org/10.1093/scan/nsz037

Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage*, *184*, 161-170.

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), 9599-9604.

Nummenmaa, L., Saarimäki, H., Glerean, E., Gotsopoulos, A., Jääskeläinen, I. P., Hari, R., & Sams, M. (2014). Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. *NeuroImage*, *102*, 498-509.

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

Ost, D. (2004). Politics as the Mobilization of Anger: Emotions in Movements and in Power. *European Journal of Social Theory*, *7*(2), 229-244.

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., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830.

Pew Research Center, (June, 2014). *Political Polarization in the American Public* [Report]. https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/

Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self–Affirmation and the reduction of defensive biases. *Current Directions in Psychological Science*, *11*(4), 119-123.

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), 842-851.

Spunt, R. P., & Adolphs, R. (2014). Validating the why/how contrast for functional MRI studies of theory of mind. *Neuroimage*, *99*, 301-311.

van Baar, J. M., Halpern, D. J., & FeldmanHall, O. (2021). Intolerance of uncertainty modulates brain-to-brain synchrony during politically polarized perception. Proceedings of the National Academy of Sciences, 118(20).

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences, 22*(3), 213-224.

Warner, J. (2014). 'Heads must roll'? Emotional politics, the press and the death of Baby P. *British Journal of Social Work*, *44*(6), 1637-1653.

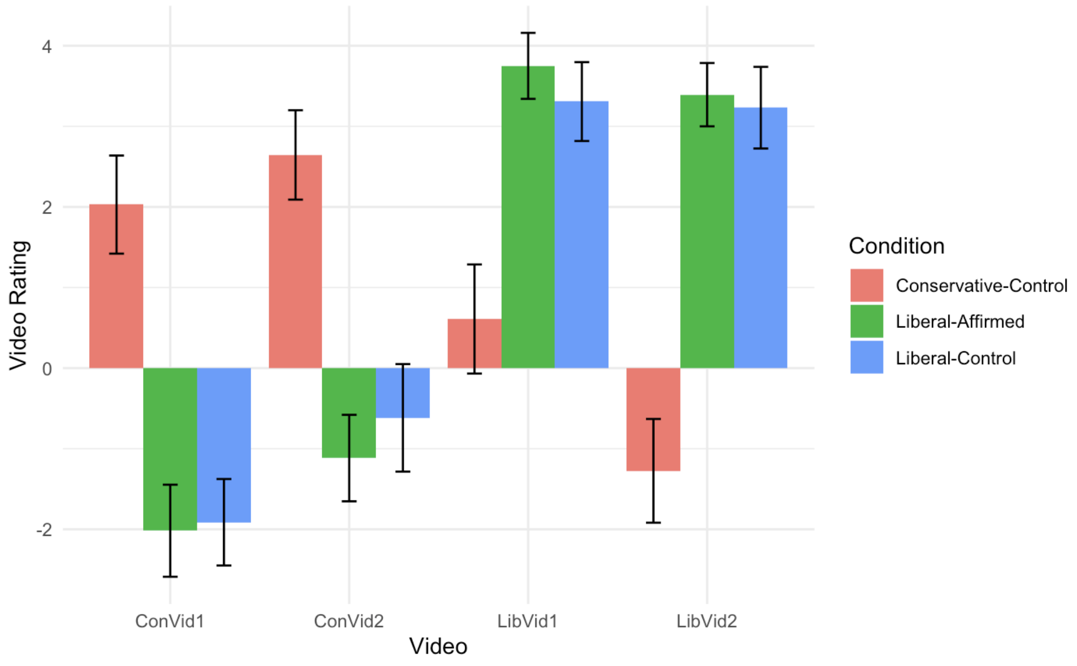
Westen, D., Blagov, P. S., Harenski, K., Kilts, C., & Hamann, S. (2006). Neural bases of motivated reasoning: An fMRI study of emotional constraints on partisan political judgment in the 2004 US presidential election. *Journal of Cognitive Neuroscience*, *18*(11), 1947-1958.

Woods, M., Anderson, J., Guilbert, S., & Watkin, S. (2012). 'The country (side) is angry': emotion and explanation in protest mobilization. *Social & Cultural Geography*, *13*(6), 567-585.

Yeshurun, Y., Nguyen, M., & Hasson, U. (2021). The default mode network: where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience*, 1-12.

Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same story, different story: the neural representation of interpretive frameworks. *Psychological Science*, *28*(3), 307-319.

Appendix



Appendix - Figure 1. Participants' evaluations of the four YouTube style videos, where ratings represented a composite score of how much participants liked the speaker in the video, felt bothered by the speaker's argument, thought the argument was reasonable, and thought the argument was logical. Ratings are displayed for each video by condition. All videos produced polarizing responses between conservatives and liberals, such that one partisan group had a positive evaluation while the other group had a negative one. There were no significant differences between the ratings of the affirmed and control liberal groups.

Appendix - Table 1. fNIRS channel - mappings

Channel	**S-D Pair**	**10-20**	**Nearest MNI Coordinate (Neurosynth)**			**Anatomical Label**	**Anatomical ROI**
			x	y	z		
1	1-1	F3-F5	-48	42	26	L Middle Frontal Gyrus	1
2	1-2	F3-F1	-32	44	42	L Middle Frontal Gyrus	1
3	1-10	F3-FC3	-46	30	40	L Middle Frontal Gyrus	1
4	2-1	AF7-F5	-50	48	0	L Middle Orbital Gyrus	2
5	2-3	AF7-Fp1	-32	62	-8	L Middle Orbital Gyrus	2
6	3-1	AF3-F5	-44	58	12	L Middle Frontal Gyrus	2

7	3-2	AF3-F1	-26	56	30	L Superior Frontal Gyrus	3
8	3-3	AF3-Fp1	-24	66	4	L Superior Orbital Gyrus	4
9	3-4	AF3-Afz	-14	68	20	L Superior Frontal Gyrus	4
10	4-2	Fz-F1	-10	48	48	L Superior Medial Gyrus	3
11	4-4	Fz-Afz	2	60	38	L Superior Medial Gyrus	5
12	4-5	Fz-F2	16	48	48	R Superior Medial Gyrus	5
13	4-8	Fz-FCz	2	30	62	L Superior Medial Gyrus	3
14	5-3	Fpz-Fp1	-20	68	-4	L Superior Orbital Gyrus	4
15	5-4	Fpz-Afz	2	70	12	R Superior Medial Gyrus	4
16	5-6	Fpz-Fp2	14	70	-6	R Superior Orbital Gyrus	6
17	6-4	AF4-Afz	16	66	22	R Superior Frontal Gyrus	6
18	6-5	AF4-F2	26	56	34	R Superior Frontal Gyrus	5
19	6-6	AF4-Fp2	28	66	4	R Superior Orbital Gyrus	6
20	6-7	AF4-F6	44	58	14	R Middle Frontal Gyrus	7
21	7-5	F4-F2	34	46	40	R Middle Frontal Gyrus	8
22	7-7	F4-F6	50	44	22	R Middle Frontal Gyrus	8
23	7-14	F4-FC4	46	30	40	R Middle Frontal Gyrus	8
24	8-6	AF8-Fp2	36	64	-8	R Middle Orbital Gyrus	7
25	8-7	AF8-F6	50	50	2	R Middle Orbital Gyrus	7
26	9-1	F7-F5	-52	40	0	L IFG (p. orbitalis)	9
27	9-9	F7-FT7	-52	18	-12	L IFG (p. orbitalis)	9
28	10-1	FC5-F5	-56	26	16	L IFG (p. triangularis)	9
29	10-9	FC5-FT7	-60	8	2	L temporal pole	9

30	10-10	FC5-FC3	-56	14	36	L IFG (p. triangularis)	10
31	10-11	FC5-C5	-62	-2	24	L postcentral gyrus	11
32	11-2	FC1-F1	-26	30	56	L Superior Frontal Gyrus	3
33	11-8	FC1-FCz	-12	16	68	L Superior Frontal Gyrus	12
34	11-10	FC1-FC3	-40	16	56	L Middle Frontal Gyrus	12
35	11-12	FC1-C1	-30	0	68	L Middle Frontal Gyrus	12
36	12-9	T7-FT7	-66	-8	-14	L Middle Temporal Gyrus	13
37	12-11	T7-C5	-68	-18	8	L Superior Temporal Gyrus	14
38	12-17	T7-TP7	-68	-30	-10	L Middle Temporal Gyrus	13
39	13-10	C3-FC3	-52	0	50	L Precentral Gyrus	10
40	13-11	C3-C5	-62	-16	42	L Postcentral Gyrus	11
41	13-12	C3-C1	-44	-16	64	L Precentral Gyrus	10
42	13-18	C3-CP3	-56	-30	56	L Inferior Parietal Lobule	11
43	14-11	CP5-C5	-68	-30	28	L Supramarginal Gyrus	15
44	14-17	CP5-TP7	-68	-44	12	L Middle Temporal Gyrus	14
45	14-18	CP5-CP3	-66	-44	44	L Supramarginal Gyrus	15
46	14-20	CP5-P5	-62	-56	30	L Middle Temporal Gyrus	15
47	15-12	CP1-C1	-28	-28	74	L Precentral Gyrus	10
48	15-18	CP1-CP3	-42	-44	64	L Postcentral Gyrus	16

49	15-25	CP1-CPz	-14	-46	78	L Superior Parietal Lobule	16
50	15-26	CP1-P1	-26	-58	70	L Superior Parietal Lobule	16
51	16-17	P7-TP7	-64	-56	-4	L Middle Temporal Gyrus	13
52	16-19	P7-P9	-54	-66	-20	L Inferior Temporal Gyrus	17
53	16-20	P7-P5	-58	-68	12	L Middle Temporal Gyrus	13
54	17-7	F8-F6	54	40	2	R IFG (p. orbitalis)	18
55	17-13	F8-FT8	52	18	-12	R Temporal pole	18
56	18-7	FC6-F6	58	26	16	R IFG (p. triangularis)	18
57	18-13	FC6-FT8	62	8	4	R IFG (p. opercularis)	18
58	18-14	FC6-FC4	56	14	34	R IFG (p. opercularis)	19
59	18-15	FC6-C6	66	-2	24	R postcentral gyrus	20
60	19-5	FC2-F2	26	32	56	R Superior Frontal Gyrus	5
61	19-8	FC2-FCz	14	20	66	R Superior Frontal Gyrus	21
62	19-14	FC2-FC4	38	20	56	R Middle Frontal Gyrus	21
63	19-16	FC2-C2	30	2	68	R Superior Frontal Gyrus	21
64	20-13	T8-FT8	68	-8	-12	R Middle Temporal Gyrus	22
65	20-15	T8-C6	70	-18	8	R Superior Temporal Gyrus	23
66	20-21	T8-TP8	70	-30	-8	R Middle Temporal Gyrus	22
67	21-14	C4-FC4	54	0	50	R Precentral Gyrus	19

68	21-15	C4-C6	64	-16	42	R Postcentral Gyrus	20
69	21-16	C4-C2	44	-16	64	R Precentral Gyrus	19
70	21-22	C4-CP4	56	-30	56	R Inferior Parietal Lobule	20
71	22-15	CP6-C6	68	-30	28	R Supramarginal Gyrus	24
72	22-21	CP6-TP8	68	-44	12	R Middle Temporal Gyrus	23
73	22-22	CP6-CP4	62	-44	44	R Supramarginal Gyrus	24
74	22-24	CP6-P6	62	-56	28	R Middle Temporal Gyrus	24
75	23-16	CP2-C2	30	-30	72	R Precentral Gyrus	19
76	23-22	CP2-CP4	44	-44	64	R Superior Parietal Lobule	25
77	23-25	CP2-CPz	14	-48	78	R Superior Parietal Lobule	25
78	23-27	CP2-P2	26	-56	72	R Superior Parietal Lobule	25
79	24-21	P8-TP8	64	-54	4	R Middle Temporal Gyrus	22
80	24-23	P8-P10	54	-66	-20	R Inferior Temporal Gyrus	26
81	24-24	P8-P6	58	-68	12	R Middle Temporal Gyrus	22
82	25-18	P3-CP3	-48	-60	52	L Inferior Parietal Lobule	27
83	25-20	P3-P5	-52	-70	38	L Angular Gyrus	27
84	25-26	P3-P1	-34	-70	56	L Superior Parietal Lobule	16

85	26-25	Pz-CPz	0	-52	70	L Precuneus	28
86	26-26	PZ-P1	-12	-68	66	L Precuneus	28
87	26-27	Pz-P2	12	-68	66	R Precuneus	28
88	26-28	Pz-Poz	-2	-72	60	R Precuneus	28
89	27-22	P4-CP4	44	-56	58	R Inferior Parietal Lobule	29
90	27-24	P4-P6	52	-66	42	R Angular Gyrus	29
91	27-27	P4-P2	38	-70	54	R Superior Parietal Lobule	25
92	28-20	PO3-P5	-46	-80	32	L Angular Gyrus	27
93	28-26	PO3-P1	-32	-76	52	L Superior Occipital Gyrus	30
94	28-28	PO3-Poz	-16	-90	42	L Superior Occipital Gyrus	30
95	28-30	PO3-O1	-32	-94	20	L Middle Occipital Gyrus	30
96	29-24	PO4-P6	46	-78	32	R Angular Gyrus	29
97	29-27	PO4-P2	26	-80	52	R Superior Occipital Gyrus	30
98	29-28	PO4-Poz	14	-90	40	R Superior Occipital Gyrus	30
99	29-31	PO4-O2	26	-98	20	R Middle Occipital Gyrus	31
100	30-20	PO7-P5	-52	-78	16	L Middle Occipital Gyrus	31
101	30-29	PO7-PO9	-44	-86	-14	L Inferior Occipital Gyrus	17
102	30-30	PO7-O1	-34	-96	6	L Middle Occipital Gyrus	31
103	31-28	Oz-Poz	-2	-98	26	L Cuneus	30
104	31-30	Oz-O1	-14	-106	10	L Middle Occipital Gyrus	31

105	31-31	Oz-O2	10	-104	10	R Middle Occipital Gyrus	31
106	32-24	PO8-P6	50	-78	14	R Middle Temporal Gyrus	31
107	32-31	PO8-O2	36	-96	4	R Middle Occipital Gyrus	31
108	32-32	PO8-PO10	44	-88	-14	R Inferior Occipital Gyrus	26

Appendix - Table 2. **Significant anatomical ROIs in whole-brain ISC contrasts**

	Concatenated liberal videos	**Concatenated conservative videos**	**Most polarizing liberal video**	**Most polarizing conservative video**
All liberals & conservatives	None	7, 8, 15, 31	2, 19	1, 4, 5 ,7 8 , 11, 14 15, 27, 31, 32, 34
Affirmed & control liberals	None	None	None	37, 58, 67, 70, 88, 89

* Whole-brain analyses were FDR corrected for multiple comparisons across the multiple anatomical ROIs

Self-Affirmation Intervention/Control Task:

Value ranking:

Below is a list of characteristics and values, some of which may be important to you, some of which may be unimportant.

Please rank these values and qualities (by dragging them) in order of their importance to you: from 1=most important to 11=least important.

______ artistic skills/aesthetic appreciation (1)
______ sense of humor (2)
______ relations with friends/family (3)
______ spontaneity/living life in the moment (4)
______ social skills (5)
______ athletics (6)
______ musical ability/appreciation (7)
______ physical attractiveness (8)
______ creativity (9)
______ business/managerial skills (10)
______ romantic values (11)

Self-Affirmation Condition:

On the previous page, you indicated that **[insert value]** is most important to you.

Describe three or four personal experiences in which **[insert value]** was important to you and made you feel good about yourself.

Focus on your thoughts and feelings, and don't worry about spelling, grammar, or how well written it is.

____

____

____

____

____

Now pick one of these experiences and write a short story describing the event and your feelings at the time.

Again, focus on your thoughts and feelings, and don't worry about spelling, grammar, or how well written it is.

____

____

____

____

____

Control Condition:

Please list, in as much detail as you can, everything you have had to eat and drink in the past 48 hours.

Don't worry about spelling, grammar, how well written it is, or things you find yourself unable to remember.

____

____

____

____

____

Manipulation check:

Mood: Take a moment to consider how you are feeling. How would you describe your mood right now? (1=extremely bad mood to 9=extremely good mood).

Self-esteem: In general, how do you feel about yourself? (1=extremely negatively to 7=extremely positively)

Attitude Measure:

Please indicate *your true opinion* on whether or not there should be stricter gun control laws.

"I am ___________________ stricter gun control laws." (1=completely against, 2=against, 3=sort of against, 4=neither for nor against, 5=sort of for, 6= for, 7=completely for)

Latitude of Acceptance (i.e. issue-specific open-mindedness) measure:

What is your latitude of acceptance for opinions on <u>stricter gun control laws</u>?

Your latitude of acceptance is the range of opinions that DON'T bother you or make you angry. Your latitude of acceptance includes your own opinion (presumably it does not bother you that you hold the opinion that you do!) and any other opinions that also don't bother you or make you angry.

Please use the sliders to mark the "boundaries" of the range of opinions that do not bother you or make you angry. The sliders will start at your own position. You should move them so that they mark the farthest opinions away from your own opinion that do not bother you or make you angry.

(Note: If your own opinion is already at the end of the scale, you will only move one of the sliders. If all opinions other than your own bother you or make you angry, you can leave both sliders where they are.)

"It does NOT bother me or make me angry when a person is _____________ stricter gun control laws."

(1)	(2)	(3)	(4)	(5)	(6)	(7)
completely against	against	sort of against	neither for nor against	sort of for	for	completely for

move to the LEFT of your own opinion <<<< ()	
move to the RIGHT of your own opinion >>>> ()	

Video scripts:

Pro-Gun Control Video 1 Script

Hey everyone, welcome back. Today's video is a doozy because we're going to be addressing one of the biggest controversies in America right now. It's been called one of the greatest public health concerns and one of the most divisive issues in the country, and that is gun control.

Let's first start with a story. In April of 1996 in Tasmania, Australia, a crowd was gathered at the historic Port Arthur prison colony monument. Suddenly, a man opened fire at the busy tourist attraction, killing 35 people. Sounds familiar, doesn't it? But here's something that's not so familiar to Americans - within just a few months, the Australian ruling center-right political party responded to this atrocity by enacting effective gun control legislation that banned automatic and semiautomatic guns, created new licensing requirements, established a national firearm registry, made a waiting period for gun purchases, and fairly compensated anyone who wanted to turn over the guns they had for responsible disposal. Since then, mass shootings in the country have dropped 13%, gun-related homicides 59%, and gun suicides by 65%. Academic studies estimate that the buyback program took enough guns off the street to save at least 200 lives every year.

This is an impressive and encouraging accomplishment, and I bring it up to show how society can get better when it has the political will to do what is right and enact common sense gun control. Even in a country as big as Australia, which is more rural than us and had a similar rate of gun ownership prior to 1996, gun control works. You can look at the rest of the developed world as well – the US is a huge outlier with the amount of gun violence we have, and in our lack of gun laws. For instance in Japan, they had 6 gun deaths last year. Six, total. According to some studies in the US, gun control methods like background checks and gun registration can reduce deaths by up to 83%.

That's an improvement we desperately need. Today is the 318th day of 2017, and so far this year, there have been 317 mass shootings. There are 100,000 gun deaths and injuries in the country every year. And did you know, simply owning a gun greatly increases the risk of violent death in your home? The risk of homicide is three times higher in homes with firearms, and the risk of suicide is three to five times higher. A gun is also 11 times more likely to be used for a suicide, 7 times more likely for a criminal assault, and 4 times more likely for an accidental injury than it is to be used in self-defense. Guns also raise our healthcare and tax costs, due to the legal services, medical emergencies, policing need, and lost earnings that result from people using guns on each other. We're sadly so used to all of this being a reality but to someone who's never heard of guns before, it's like, "why the heck are people allowed to have these things in the first place?"

Gun advocates will say, ok, I guess it's sad that some people die, but you need guns to prevent even more death! Except that is a myth. Armed citizens are actually very unlikely to stop crimes. Of the 6 million violent crimes that happen a year, less than 1% involve the victim defending themselves with a gun, compared to the much higher rate of using the gun to commit the crime in the first place. In the Dallas shooting in 2016, where a guy started sniping police officers during a protest, more than 100 police and dozens of Texans with guns couldn't take down this highly mobile and lethal adversary. They needed a bomb robot to eventually do it only after five people were murdered. Killers may still choose to use a knife or a truck to kill people, that's true, but having a gun readily available to them makes it possible to kill many more people in a shorter amount of time. And actually, civilians with guns usually make situations worse. The presence of a gun makes an argument or a conflict more likely to become violent, statistically. And if a dangerous situation is already happening? Think about it – if a bunch of people decided to fire back at the Las Vegas shooter, who was firing from a 32-story window in the dark, how would that even work? It'll just make more panic and chaos. And you think the police are going to automatically know you're the good guy? But the NRA wants you to think an armed civilian can take out a threat without any forewarning or professional training (since, you know, the NRA, an organization originally created to spread firearm education, even lobbies against funding gun training classes now). They want you to believe in the good guy with a gun myth, but really they just want your money and to hell with your well-being. Alright, I know that was a lot of statistics and facts, so I'll stop there to let all that soak in for you, but I hope this helps move the conversation forward, even if it's just a little bit. Thanks for listening, everybody.

Pro-Gun Control Video 2 Script

Hi folks. Thanks for tuning in today. Now I know that I don't usually don't get into politics much, but given the current political climate, it's hard for me to stay completely silent about my views. Today I wanna talk about gun control – specifically, the 2nd Amendment, how it gets misused by people today, and how we *should* be using it.

Ok, so what does the 2nd Amendment *really* mean for today's society? I've noticed that whenever any mention of gun control gets brought up, conservatives love arguing back with: "But what about our 2nd Amendment rights?!" Gun nuts cling to that one like an old woman clutches her pearls. But they're total hypocrites when it comes to that! They never cared about the 2nd Amendment until the 1970's. Go look in any law library and try to find references to it from before then -- it's one of the least cited amendments of any of them. And it's not like those people care about the other amendments as much, such as freedom of religion in the *1st* Amendment, since they want to ban gay marriage and give preference to Christian refugees.

And get this – the 2nd Amendment was never actually meant to be used the way it gets used now. The way that conservatives understand the 2nd Amendment, and throw its name around as a justification, is far different from what the Founding Fathers intended it to be used for. The text of the 2nd Amendment reads, "*A well regulated militia*, being necessary to the security of a free State, the right of the people to keep and bear arms, shall not be infringed." With that clause about a "well regulated militia" coming first, the passage is not saying individuals should bear arms, but that the People with a capital P have the right to a standing militia that can protect the state using firearms. Many constitutional scholars agree on this, and they have throughout history. No Federal court until 2008 ever ruled against a gun restriction law, because individual gun ownership is *not* what the 2nd Amendment is talking about. For instance there's the famous United States vs. Miller Supreme Court case that ruled you could regulate sawed off shotguns because they did not have quote "some reasonable relationship to the preservation of efficiency of a well regulated militia."

The Founding Fathers included the Second Amendment to ensure we had a national guard. That's it. Even modern conservatives like Scalia have said quote "nothing should be taken to cast doubt on longstanding prohibitions on the possession of firearms by felons and the mentally ill, or laws forbidding the carrying of firearms in sensitive places, or laws imposing conditions and qualifications on the commercial sale of arms." Everyone agreed on this until the 1970's, when the NRA figured out it could make more money by pushing the individualist interpretation of the right to bear arms. This would make lots more people want to buy guns and pay them membership dues. When this individualist interpretation of the 2nd Amendment started becoming popular, *Republican* Supreme Court Justice Berger said that, "this new interpretation of the 2nd Amendment is one of the greatest pieces of fraud on the American public by special interest groups that I have ever seen."

Plus, no matter which interpretation is ultimately correct, the Constitution is a product of the time at which it's written, and is free to be amended as times change. I'm not even going as far to say that we need to get rid of the 2nd Amendment entirely, but let me remind you that John Paul Stevens, a *conservative* former Supreme Court Justice, think it's time to repeal it, just sayin'...

In the 18th century, it was reasonable to think you needed a slow action rifle because you had bears wandering around all the time. We don't have those same threats anymore, and today's guns are much faster and more lethal. And yet, we're required to register our dogs, but not our guns! To claim guns are necessary, and that people have the right to have any type of gun they want, just doesn't stand up anymore. There's just no need for automatic weapons and high capacity magazines unless you want to go kill a bunch of people. A lot of veterans, who were trained to use military-style weapons like AR-15's *in combat,* say that civilians have no business owning these guns. Even the majority of gun owners believe we should increase background checks and create a national database. We just need more common sense here.

I want to end by leaving you with one more thought. Maybe you still think that despite all the constitutional law, you really believe there is this inalienable right for everyone to own as many guns as they want and to carry them wherever they want. But then you also have to admit that in the Declaration of Independence, the right to "life" was listed before "liberty" and "the pursuit of happiness." The freedom of an individual to not find themselves on the receiving end of a firearm is more important than your right to own a firearm. Your rights end where my face begins. Widespread gun ownership eventually leads to people losing the rights to their life. Even if accidentally, that's just not acceptable.

Anti-gun control video 1

Hello gents and ladies, hope all is well. Today's Tuesday folks, so of course, it's time for a current events video for ya. Today I'm going to be debunking some of the myths that are out there about guns. Yet again, the topic of gun control has been tossed around in the media because of recent news and I've seen people debating this in the comments, so I think it's important to have a discussion about why gun control is actually an incorrect and dangerous notion. There's a lot of misinformation out there, bandied about by people who are probably well intentioned, but just don't understand guns and haven't done the research. So here are the important facts for you so that you can understand this issue better.

First things first, saying I don't care about the lives of other people because I own a gun is completely false. In fact, buying a gun is me saying that I definitely *do* care about preserving life, by defending myself, my family, and my community from people who would do harm. There's always going to be criminals that are trying to hurt you. That's just life. But having a gun means you can protect yourself from them. If you make guns illegal, all the law-abiding citizens wouldn't have guns, and all the criminals would. Because they're *criminals,* who don't listen to laws to begin with. And those are exactly the sort of people we need to be defending ourselves against. Remember, guns don't kill people, people kill people. Plus, without the threat of other people with guns to stop the criminals, they'll commit even more crimes to boot. Murderers go after soft targets.

And even if you can, somehow, prevent people from getting guns, they'll still find other ways to hurt people. All the recent truck rammings by terrorists prove that point. But when we have guns, it's at least easier to stop those guys. So in the case where someone is trying to break in, or you're being mugged, or your coworker has decided you're an infidel, it doesn't work to just call the police and hope for the best. *Maybe* I'm wrong, but I don't think a terrorist is going to patiently wait 20 minutes for the police to get there, so if you want to come out of the situation with your life, you need to be able to protect yourself.

My main point here is that more guns doesn't equal more gun violence – and more gun *control* doesn't equal less gun violence. Just take a look at Chicago. The city has some of the strongest gun control laws in the country, but the highest homicide rate. Last year there were more than 4,000 reported shooting incidents in the city alone. Someone is shot every two and a half hours on average. And of all the gun crimes in the city, 97% of them were committed with an illegally obtained weapon. That's a terrible record for a supposed "gun control" law.

In the recent mass shooting in Texas where some guy attacked a church, it wasn't the police that stopped him, but a civilian with a concealed carry permit. If gun laws had been in place to ban carrying weapons like a lot of liberals want, this murderer could have killed a lot more people. In 9 other states, gun crime *rose* after gun control laws were put in place. And for the guy in Vegas, he was on no watch list with no previous criminal record. No gun law would have stopped him, but it would deprive millions of legitimate gun owners of their self-protection rights. People feel safer when they have their own gun, and today with a record number of gun sales going on, there's also a record low of violent crime across the country. Don't ignore that correlation.

Sure, there are bad people out there with guns – that's not something that's up for debate. But even the most fervent gun control advocate will admit that the vast majority of gun owners are responsible and good people. A full one-third of the country owns a gun, according to the Pew Research Center – of course the majority of them are decent people. And two-thirds of them say they own the gun just because of self-defense. For the last third of gun owners, millions are farmers or ranchers who have to defend their crops and animals from predators. Even more people are hunters who use guns to literally put food on the table, the most basic of needs. Using guns is their entire way of life and without them, they have no way to support their families. Taking away guns then is an overreaction to just a problematic minority of individuals, who would be criminals anyways, and punishes all the innocent people who legitimately need guns for their livelihood and sense of safety. All right, so thanks for hearing me out folks. Hope you have a good rest of your week, and see you guys next time.

Anti-Gun Control Video 2:

Hey there everyone! Sorry for the brief hiatus, but not to worry… I'm back! Since I've been gone for a bit, I decided that in my first video back, I wanted to tackle something that many of you know is really important to me. Gun rights. Now, guns are pretty much always in the news on and off, no thanks to the mass shootings that happen in our country on a regular basis, which of course, are devastating. And you'll notice that every time something happens, the liberal media uses these events to drive forward their agenda on gun control. So that's one thing. But recently, there's been people going so far as to suggest that we should flat out *repeal* the 2nd Amendment, and this has gotten me really spooked.

First of all, with the way liberal politicians and the media discuss gun control, they very clearly demonstrate that they have no idea what they're talking about, and that they haven't fully thought through the practicalities of what they're trying to do. Everything is always an AR-15, or they say banning assault weapons would be easy. But "assault weapon" isn't even a real term! There's no class of guns that are officially assault weapons. And real "automatic" weapons have already been banned since 1986. All the "automatic" guns that the media talks about are really just the automatic-like functions of semi-autos, such as the auto round cycling or burst firing.   And still, semi-autos only account for 2% of all gun violence. The rest is with weapons like handguns that gangs use to attack each other in inner cities, or people use on themselves. Not to mention, Nancy Pelosi and those liberals who want to ban guns are the same people who were ok with the Obama administration supplying Mexican drug cartels with those very weapons. But those aren't the numbers you hear about in the news. You also don't hear about any attempts to ban cars or fast food, even though traffic deaths and heart disease kill far more people every year than mass shootings. So these guys just have no idea about the tools they're trying to regulate. Lots of people with common sense would rather keep guns as a last line of defense than entrust their safety to government incompetence.

The greater issue at hand here, folks, is this: in this country, gun ownership is a constitutionally-enshrined right we all have, that helps preserve our freedom and liberty. When the Founding Fathers wrote the 2nd Amendment, they understood how dangerous a tyrannical government could be to the well-being of a nation. You may think that could never happen now, but look what happened to the Jews in Nazi Germany, or the chaos that is Venezuela right now. In the US, it was the Democratic party that wanted more gun legislation because they wanted to keep guns away from African Americans and the Black Panthers. Did you know that? Nowadays, lots of people accuse the Trump administration of being fascists and Nazis, but if you think your government is becoming totalitarian, getting a gun is *exactly* what you should be doing. The famous 2008 Supreme Court Case DC et al. vs. Heller confirmed that you have that right by saying "The Second Amendment protects an individual right to possess a firearm unconnected with service in a militia, and to use that arm for traditionally lawful purposes like self-defense."

Another thing is that guns are also a deeply ingrained part of American culture.
The elites in Washington think banning guns will be good enough to protect everyone from them, but how do they expect to round up every one of the *300 million* guns in this country, when they can't even protect its borders effectively? It's just not feasible in a country of this size. You remember when they tried it with Prohibition, and the War on Drugs? You remember when alcohol and drugs were made illegal, how everyone stopped using them? Yeah… I don't either….
Other countries are ok with giving up their guns because it's not a part of their lives as much. But you simply cannot take guns away from Americans because gun ownership is so important to our way of life. People want their guns. I promise they won't willingly give them up if all guns were banned tomorrow, so you'd either have to do an astronomically expensive buyback program that would bankrupt the economy, or you'd have to take them by force. And that's exactly the kind of government overreach that we have guns to protect ourselves from in the first place.

So you don't like people dying? Cool. I don't either! So that's why you should actually *support* gun ownership, not want to get rid of it. People have the right to defend themselves and their families, and as a society we can't afford to waste needless time and taxpayer dollars on trying to skirt one of our most fundamental rights with laws that wouldn't work anyways. Guns are just a tool, and now you're a lot more informed about them so you can correct anyone else you talk to about the issue.

Post-video questions:
(1) How much do you like the person in the video? (-5=dislike a lot… 0=neither like nor dislike … 5=like a lot)
(2) To what extent do the person's argument bother you or make you angry? (0=not at all … 5=neutral … 10=very much)
(3) How reasonable is the person's argument? (-5=extremely unreasonable… 0=neither reasonable nor unreasonable … 5=extremely reasonable)
(4) How logical is the person's argument? (0=not at all logical … 5=neutral … 10=extremely logical)

Chapter 4 - Open-Mindedness Interventions: An Integrative Review and Roadmap for Future Research

Abstract

Partisan animosity has been growing in the U.S. and around the world over the past few decades, fueling efforts by researchers and practitioners to help heal the divide. Many studies have been conducted in order to test interventions that aim to promote open-mindedness; however, these studies have been conducted in disparate literatures that do not always use the same terminology. In this narrative review, we attempt to integrate research on open-mindedness in order to facilitate cross-talk and collaboration between disciplines. Moreover, we offer a conceptual model to help guide the further development of interventions and research to understand open-mindedness. We propose that open-mindedness is dynamic, such that interventions should focus on both inducing and sustaining it. Specifically, we suggest the interventions that target cognitive and/or motivational pathways can induce open-mindedness in the first place. Then, training in emotion regulation and/or social skills can help to sustain open-mindedness once individuals are in a social context in which they focus more potential challenges to their open-mindedness. We conclude with a discussion of potential future directions for research on open-mindedness interventions.

“Philosophy should be piecemeal and provisional like science; final truth belongs to heaven, not to this world.” - Bertrand Russell, 1927

Introduction

When someone disagrees with us, it is easy to conclude that the other person is downright crazy, stupid, or biased. We assume that if the person were sane, intelligent, and clear-headed, they would share our perspective. When we adopt this mindset, we are being so-called ‘naïve realists’ — we are assuming that we have a monopoly on the ‘final truth’ and that anyone who disagrees with us needs to adjust their way of thinking (Ross & Ward, 1996). When we engage in naïve realism, we struggle to learn from new perspectives or reach mutual understanding with those who disagree with us. In other words, we fail to be *open-minded.* In response to these challenges, many scholars have developed interventions to promote open-mindedness; however, little work has been done to compile and examine these interventions. Thus, in this narrative review, we synthesize empirical work from multiple disciplines that have aimed to improve different aspects of open-mindedness, often using their own distinctive terminologies. To provide a framework for the review, we present a conceptual model of the primary pathways that are targeted by open-mindedness interventions. Furthermore, we report on the quality of the evidence to support different intervention types and suggest goals for future research.

Motivation for The Present Review

Previous work aiming to increase open-mindedness has been conducted across multiple fields, including social psychology, moral psychology, political psychology, positive psychology, conflict resolution/peace-making, education, political science, sociology, philosophy of education, communication studies, virtue epistemology, negotiation, and organizational behavior. Many practitioners and bridge-building coalitions have also attempted to open minds. For instance, as of May 2021, more than 6,500 groups were catalogued in the Princeton University's Bridging Divides Initiative's database, which does not even include groups that operate outside of the United States (Bridging Map, 2021). However, cross-talk among these disparate academic fields and practitioners has been limited, which has resulted in a range of terms that have been used to describe open-mindedness as well as a vast array of intervention approaches that have rarely been integrated.

Scholars who view open-mindedness through a lens of virtue and/or epistemological development, such as moral psychologists, philosophers, and educators, tend to emphasize how being open-minded affects individual learning (Baehr, 2011). On the other hand, social psychologists, political scientists, conflict resolution scholars, sociologists, and organizational behavior experts tend to focus more on the interpersonal and group-level consequences of being closed-minded. For example, the United States has seen a sharp increase in 'affective polarization,' or reported antipathy between its two political parties over the past decade (Iyengar et al., 2019). Liberals and conservatives think that people on the other side are closed-minded and do not share their values and goals (Pew Research Center, 2019). Moreover, liberals and conservatives have segregated themselves physically by moving to different neighborhoods (Bishop, 2008) and also virtually into social media 'echo chambers' (Cinelli et al., 2021). Therefore, when partisans fail to be open-minded toward one another, they can engage in behaviors that can reshape social structures, which can serve to further reinforce their closed-mindedness.

Thus, failing to be open-minded can have pernicious individual and group-level consequences. With this review, we hope to encourage further inter-disciplinary dialogue and collaboration among researchers and practitioners who aim to develop integrative, high-impact interventions to promote open-mindedness. We selected the format of a narrative review rather than a meta-analysis because the studies included are highly methodologically diverse, examining different outcomes and employing different interventions that target distinct mechanisms (Baumeister & Leary, 1997). Although we could examine the quality of the evidence for the different interventions, highlighting when the evidence was strong, mixed, or emergent, we did not directly compare the interventions' effect sizes due to the methodological diversity of the included literature. In the future, it may be beneficial for scholars to agree on a shared terminology when discussing open-mindedness and related interventions, which would facilitate better information exchange across academic disciplines. Furthermore, it would be useful for scholars to align on more standardized measures and intervention protocols. This would make it possible to perform similar, more precisely targeted reviews and meta-analyses in the future to determine which open-mindedness interventions have the strongest effects.

Previous reviews have been conducted on interventions that are related to, but conceptually distinct from, open-mindedness interventions. For instance, Paluck et al. (2020) and Paluck and Green (2009) provide thorough narrative reviews of interventions to reduce prejudice and discrimination. These interventions tend to focus on improving attitudes and behaviors toward specific people or groups. In contrast, interventions to improve open-mindedness focus more on encouraging neutral or positive attitudes toward the alternative ideas that others hold. Furthermore, increasing open-mindedness is also distinct from attitude change (for a recent review of the attitude change literature, see Albarracín & Shavitt, 2018). Attitude change interventions attempt to shift people's evaluations (i.e., change their minds); in contrast, open-mindedness interventions attempt to expand people's 'latitudes of acceptance,' or the range of attitudes that they find to be acceptable, without requiring them to shift their attitudes (Sherif & Hovland, 1961).

There is also significant conceptual overlap between open-mindedness and empathy interventions. Empathy can be defined as "the ability of one person (a perceiver) to share and understand the internal states of someone else (a target)" (Weisz & Cikara, 2021). Many definitions posit that empathy has affective, cognitive, and motivational components (Weisz & Cikara, 2021). However, many empathy intervention studies do not separate these three components from one another. Of these components, the concept that comes closest to open-mindedness is 'cognitive empathy' (or 'perspective-taking'), which is regarded as being able to intellectually understand what others feel and think. However, understanding another person's point of view is different from believing that the person's view may be reasonable and worthy of consideration, which is core to being open-minded. Furthermore, the wide-ranging empathy literature has historically focused more on people's ability to share and understand what others *feel* rather than what they *think,* which is more core to open-mindedness. For a review of empathy interventions, see Weisz and Zaki (2018).

Finally, open-mindedness is related to, but distinct from, 'openness to experience' (sometimes abbreviated as openness), which is one of the factors in the Big Five Inventory. Openness to experience is defined as being "seen in the breadth, depth, and permeability of consciousness and in the recurrent need to enlarge and examine experience" (McCrae & Costa, 1997). In other words, openness has more of a focus on people's tendency to be curious and to pursue novel experiences rather than people's openness to alternative ideas. Furthermore, few studies have attempted to shift openness, given that it is considered to be a relatively stable personality trait. Therefore, the present review focuses more on the ability to change people's openness to alternative ideas than on the ability to change their openness in general.

Open-Mindedness as a Dynamic System: A Conceptual Model

For the purpose of this narrative review, we formally define *open-mindedness* as "an individual's willingness and ability to consider alternative viewpoints." Most definitions of open-mindedness have been developed by philosophers who perceive it to be an 'intellectual virtue.'[1] Our conceptualization of open-mindedness is a simplified version of an idea from the virtue epistemologist John Baehr (2011), who defines an open-minded person as being "characteristically (a) willing and (within limits) able (b) to transcend a default cognitive standpoint (c) in order to take up or take seriously the merits of (d) a distinct cognitive standpoint." Our definition attempts to retain these ideas while also simplifying them so that they are accessible for scientists and practitioners alike. Baehr's definition closely aligns with the philosopher Bertrand Russell's ideal of 'critical receptiveness,' which encourages welcoming new ideas while also being appropriately skeptical of them (Russell, 1928; see Hare, 2001, 2009). Psychologist John Lambie (2014) refers to this idea as 'critical open-mindedness,' distinguishing it from what he calls 'anything goes open-mindedness' to address concerns by critics that open-mindedness gives equal weight to all opinions, including evil ones (p. 16). This view of open-mindedness is also similar to the idea of 'moral pluralism,' which posits that there may be multiple acceptable moral viewpoints, but also that there can be some views that are unacceptable (Graham et al., 2013).[2] Thus, people can

[1] Recently, open-mindedness has been classified by philosophers under the umbrella of 'wise reasoning.' For a recent review of wisdom science, which contains definitions of constructs and related scales, see Grossman et al. (2020).

[2] Graham et al. share a helpful quote from philosopher Isaiah Berlin that helps distinguish 'moral pluralism' from 'moral relativism': "If I am a man or a woman with sufficient imagination (and this I do need), I can enter into a value system which is not my own, but which is nevertheless something I can conceive of men pursuing while remaining human, while remaining creatures with whom I can

be open-minded without giving a platform to hateful rhetoric, becoming brainwashed by misinformation or conspiracy theories, or even changing their minds.

Now that we have provided a basic definition of open-mindedness, we will dive deeper into the multiple, interdependent factors that contribute to having an open mind. Many open-mindedness interventions tend to focus on targeting individual mechanisms in isolation, which has been important for developing a thorough understanding of the underlying psychology of open-mindedness. However, creating interventions that can have maximal real-world impact may require a more holistic, integrated approach that considers individual, social, and broader cultural and structural factors that influence one another. For this reason, we present a conceptual model that depicts an individual's open-mindedness as existing within a dynamic and interconnected system (Figure 1).[3]

This model draws on ideas from Kurt Lewin, a founder of social psychology, who was heavily influenced by Gestalt theorists. According to Lewin's field theory, "the person and his environment have to be considered as one constellation of interdependent factors" (Lewin, 1946, p. 338). According to Lewin, at any given moment, a person's behavior could be considered to be a product of internal and external factors operating within a field of forces called their 'lifespace.' Furthermore, just as individuals are nested within situations and environments, researchers have proposed that we can also conceive of interaction partners (e.g., pairs or groups of people) as existing within one integrated system (Dale et al., 2018). Thus, since individuals' thoughts and behaviors occur within the context of a system, their

communicate, with whom I have some common values—for all human beings must have some common values or they cease to be human, and also some different values else they cease to differ, as in fact they do. That is why pluralism is not relativism—the multiple values are objective, part of the essence of humanity rather than arbitrary creations of men's subjective fancies." (Berlin, 2000)

[3] For an in-depth review of how dynamical systems theory applies to research in social psychology, see Richardson, Dale & Marsh (2014). See Thelen et al. (1994) for a general overview of cognition and dynamical systems theory.

ability to be open-minded depends on factors within that system. External factors like social norms, incentive structures, and the design of certain discussion forums may promote or discourage open-minded thoughts and behaviors.

In addition to conceiving of open-mindedness as existing within a system of factors, we also emphasize that it is dynamic: it can have both trait- and state-like qualities. Even if a person becomes more open-minded toward an opposing viewpoint after undergoing a brief intervention in isolation, once they begin interacting with someone who actually holds that viewpoint, they may quickly become defensive and clam up again. For instance, imagine that you enter into a conversation with your 'crazy Aunt Mildred' with an open mind. However, as soon as she starts 'spouting nonsense,' your blood starts to boil, you go on the defense, and you lash out at her. Although you began the conversation with an open mind, your mind closes up again once you are in a social situation that triggers automatic emotional and behavioral responses. Thus, people start out with a baseline level of open-mindedness that can then shift in different situations. Because it exists within a dynamic system, we might think of open-mindedness as a sort of *candle in the wind*. Once an open mind is 'lit,' outside forces can either shield it or snuff it out.

Researchers have used similar dynamical system models to describe the structure of attitudes and the process through which conflict resolution occurs. According to Albarracín and Shavitt (2018) attitudes "are partly memory based and partly constructed on the fly." To support this idea, they point to recent computational research in psychology that models attitudes as neural networks that engage in 'constraint satisfaction' (Monroe & Read, 2008). This model proposes that attitudes form by satisfying different constraints, including pre-existing associative weights (explaining attitudes' trait-like nature) and situational constraints (explaining attitudes' state-like nature). Similarly, researchers have described conflict resolution as a dynamical system of factors in which conflict can serve as a strong ‘attractor state’ that people tend to settle into, despite it being avoidable (Coleman et al., 2007).

We propose that the cognitive underpinnings that explain changes in open-mindedness might be described using a similar dynamical systems framework that involves constraint satisfaction. This conceptual model, which has yet to be simulated or tested empirically, suggests that a person’s open-mindedness in a given situation depends on pre-existing individual traits and external factors (Figure 1). Specifically, the model proposes that we can first develop the ability to be more open-minded through interventions that target (1) cognitive and (2) motivational processes at the individual level. These interventions help us to engage in less biased thinking, become more aware of our thought processes, and become motivated to be receptive to alternative ideas. However, we also benefit from training in (3) affective and (4) behavioral/social skills so that we can maintain an open-minded stance when we enter into a social context by regulating our emotions and engaging in constructive behaviors. We believe that psychological interventions can target these different levels, or some combination of them, to promote and sustain open-mindedness across different situations.

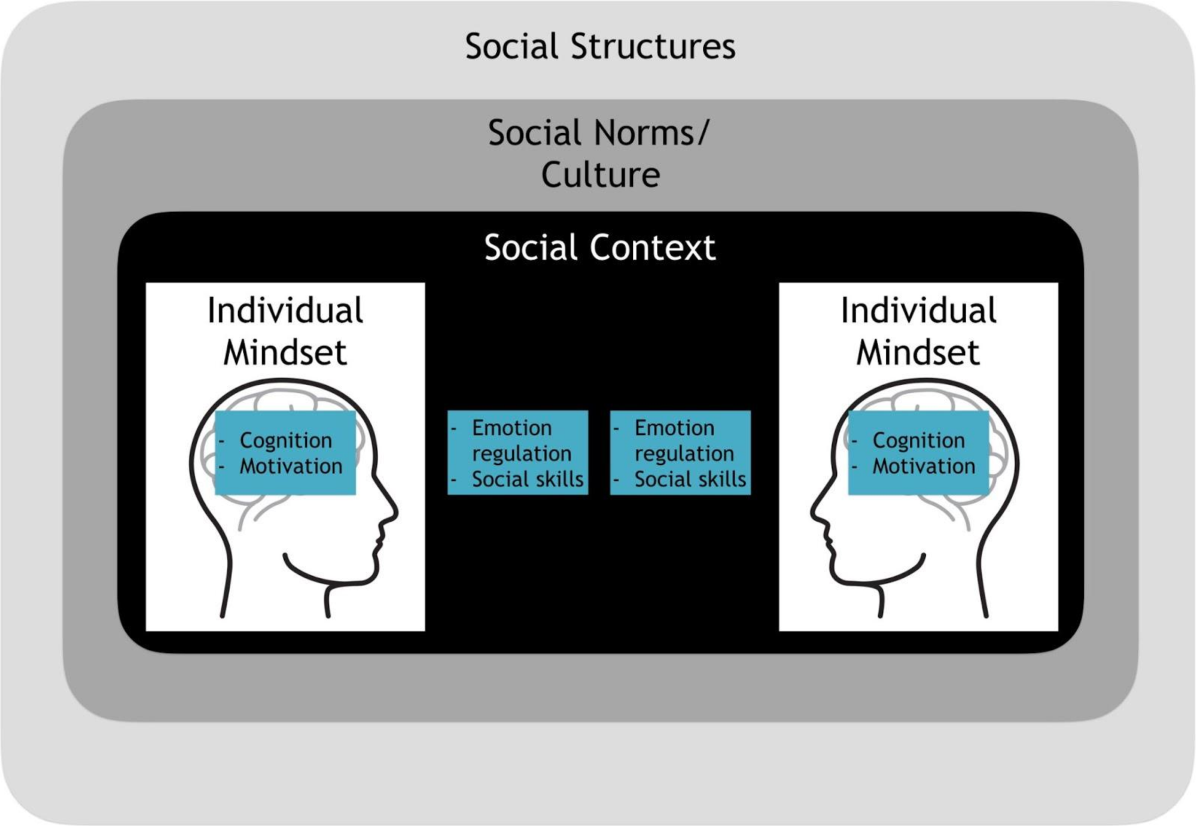


Figure 1. A conceptual model of the components that contribute to promoting and sustaining open-mindedness. Interventions can promote open-mindedness at the individual level through cognitive and motivational pathways. Interventions can also teach emotion regulation and social skills that can help to sustain open-minded thinking in social contexts. Furthermore, interventions can target social norms/culture and social structures, which influence the extent to which individuals are willing and able to engage in open-minded thinking and behavior.

In addition to the four factors that are the focus of this review, social norms (a group’s shared understanding of what behavior is appropriate in a given context) and social structures (the way that people and institutions within a society are organized) also exert a strong influence on people’s ability and desire to be open-minded. For instance, until 2005, same-sex marriage was a divisive issue in the United States, with the majority of Americans opposing it. However, by 2015, the majority of Americans *supported* it, and it was legalized in all 50 states (Pew Research Center, 2015). In turn, research has shown that this legalization led to changes in people’s perceptions of social norms (Tankard & Paluck, 2017) and further increases in individuals' open-mindedness toward same-sex marriage (Ofosu et al., 2019).

One strategy that has been proposed for influencing norms is targeting social referents — whereby interventions are strategically administered to individuals who are the most influential in their social mechanisms (see Paluck et al., 2016). Another strategy involves weakening the grip of certain social norms by educating people with regards to the inaccuracy in their beliefs about group norms (i.e., correcting pluralistic ignorance). In addition, there are likely to be several relevant strategies that aim to alter social structures in order to promote open-minded behavior. For instance, according to a theory that is popular among behavior scientists, choice architecture has a strong influence on people's behavior. Thus, changing choice architecture can 'nudge' people to engage in certain behaviors (Thaler & Sunstein, 2008). For instance, the architecture employed on social media platforms may influence the extent to which people engage in open-minded discussions with one another. Beyond choice architecture, the incentives that are embedded in institutions such as in the workplace or higher education can also be powerful drivers of behavior. Organizations may consider how they can realign their incentive structures with aims to promote open-minded thinking and dialogue.

Thus, we acknowledge that interventions may also consider shifting social norms and social structures in order to promote open-mindedness. A thorough exploration of interventions that impact these environmental factors is beyond the scope of this review due to the review's primary focus on psychological change within individuals and social interactions. Scholars in fields such as cultural psychology, anthropology, sociology, political science, organizational behavior, and group dynamics may be able to offer further insight into best practices for developing interventions that target those broader societal factors (see Blankenship et al., 2006; Bolman & Deal, 2017; Carnall, 2007; Hernández-Mogollon et al., 2010; Nielsen & Abildgaard, 2013; Schein, 1990; Shapiro, 2006; Steward, 1972; Tankard & Paluck, 2016; Valente, 2012).

Open-Mindedness Interventions

Now that we have provided a definition and conceptual model for open-mindedness, we will discuss interventions that have been developed to increase open-mindedness. According to the conceptual model laid out above, we have organized these interventions based on the primary pathway that they aim to target. First, we will review interventions that induce open-mindedness in the first place by targeting cognitive and motivational pathways within the individual. Then, we will discuss interventions that target emotion regulation and social/behavioral skills, which help to sustain open-minded thinking in real-world interactions.

As we outlined in our conceptual model, these four pathways are interconnected, such that intervening on one mechanism can have downstream effects on others. For this reason, we have categorized interventions based on the primary outcomes that they target and measure, although some could arguably be placed into multiple categories. Furthermore, some of these interventions target open-mindedness directly, while others target related constructs or underlying mechanisms. By highlighting interventions in each category, we hope to provide a roadmap that future researchers and practitioners can use to create integrated interventions that both induce and sustain open-mindedness in the long-term.

Inducing Open-Mindedness Through Cognitive Pathways

The majority of open-mindedness interventions target cognitive pathways. In particular, these interventions aim to reduce biased thinking and/or promote a more expansive mindset. Some of these interventions take a targeted and direct approach by teaching individuals about cognitive biases and giving them strategies to avoid biased thinking. Another targeted approach involves teaching people to embrace certain beliefs, or implicit theories, that lead them to engage in more open-minded thinking and behavior. Yet another targeted, more socio-cognitive, approach involves training people to take the perspectives of others. Alternatively, some interventions are less targeted, influencing more domain-general cognitive mechanisms. For instance, many interventions aim to broaden and/or complexify people's thinking through priming. Other interventions use cognitive training that operates on low-level cognitive processes underlying open-ended thinking, such as meta-cognitive awareness. We will begin by reviewing the more direct and targeted cognitive approaches, followed by the more domain-general approaches.

Targeted Cognitive Approaches

Teaching About Biases. One targeted approach for increasing open-mindedness through a cognitive pathway involves teaching individuals about the existence of biased thinking and then training them on how to engage in alternative thought processes. Over the past decades, researchers have documented several cognitive biases that reliably alter human judgment and decision-making in a variety of domains (Nisbett & Ross, 1980; Tversky & Kahneman, 1974; Vallone et al., 1985). Research on correcting biases in social judgment shows that participants must learn about the bias, identify how the bias has affected their own judgments, and then be motivated and able to correct for it (Wegener et al., 1995; Wegener, Petty, & Dunn, 1998). In general, these approaches tend to focus on shifting people from engaging in heuristic-based, System 1 thinking to more controlled, System 2 thinking (Lilienfield et al., 2009; Stanovich & West, 2000).

In three studies conducted among Israelis and Palestinians, Nasie et al. (2014) first taught participants about naïve realism, defining it as "the human tendency to form one's own worldview regarding various subjects, perceived by an individual as the only truth." Then, they taught participants how naïve realism is related to conflict and provided an example of it occurring during a specific conflict. The researchers found that this intervention was most successful among participants who were initially more authoritarian — or 'hawkish' — as these individuals had started out with more biased thinking to begin with, which they could then recognize themselves engaging in and correct for. These hawkish participants who learned about naïve realism reported being more open to the views of the opposing side (e.g., Hawkish Palestinians were more open to Israeli attitudes), and more open to learning about those alternative views from movies, media, and/or meeting with a member of the opposing group. On the other hand, less authoritarian — or 'dovish' — participants were more likely to be open-minded to begin with, and thus, the intervention was less effective for them. The researchers note that further research on the long-term effects of this intervention is warranted, in addition to testing the intervention in larger samples.

Though Nasie et al. (2014) taught participants about how naïve realism can lead to conflict, they did not instruct participants on how to avoid engaging in naïve realism. However, some researchers warn that teaching participants about bias may not be not enough. Lord et al. (1984) argue that it is also important to provide people with specific strategies for overcoming bias. In their study, the researchers exposed proponents and opponents of capital punishment to two essays: one suggesting that the death penalty reduces crime rate and one suggesting that the death penalty is ineffective. Participants read instructions that: (1) were general, (2) told them that they should "be unbiased" and consider all evidence in an impartial manner, or (3) taught them about biased assimilation of evidence and instructed them to consider how evidence supporting an opposite conclusion would affect their evaluations ('consider-the-opposite'). Whereas participants in the first two conditions displayed more extreme attitudes after reading the essays, participants in the 'consider-the-opposite' condition did not show attitude polarization (i.e., remained more impartial). Thus, the 'consider-the-opposite' intervention was thought to be most successful because it incorporated education on a specific bias and also provided tools to reduce the bias.

Some researchers have found that incorporating gamification and personalization into debiasing interventions can boost effect sizes. For instance, Morewedge et al. (2015) created a computer game called *Missing: The Pursuit of Terry Hughes* (Symborski et al., 2014)*,* which teaches participants about three cognitive biases related to open-mindedness (the bias blind spot, confirmation bias, and the fundamental attribution error). In the game, participants are primed to first engage in these biases. Then, they learn the definitions of the biases and receive personalized feedback about the extent to which they engaged in the biases while making decisions during the game. Finally, they learn about another example in which the biases affected a situation in the real world. Then, they have the opportunity to practice making unbiased judgments. The researchers compared this 60-minute video game intervention against a 30-minute instructional video about the biases. They found that both interventions were effective in reducing the three types of biases in the short- and long-term. However, the computer game was more effective than the instructional video. Whereas the video produced small to medium effect sizes (d=0.38-0.69 from pre-to-post intervention; d=0.49-0.66 at follow-up), the computer game produced mostly large effects (d=0.98-1.168 from pre-to-post; d=0.72-1.11 at follow-up). Therefore, the researchers argue that brief, one-shot interventions can be powerful at debiasing, especially when they incorporate gamification, personalized feedback, and opportunities for practice.

Other researchers have developed debiasing interventions that reinforce concepts over multiple training sessions. For instance, Hudley and Graham (1993) developed a 12-session “attribution retraining program” in order to reduce attributions of “hostile intentions” and reduce aggressive behavior in 10-12-year-old boys. This intervention focused on increasing participants’ open-mindedness to alternative explanations for their peers’ behavior. The program taught participants how to (1) identify intent accurately, (2) make non-hostile attributions when intent was ambiguous, and (3) learn how to generate decision rules for how they should behave in response to non-hostile intent. The researchers found that the intervention was successful for individuals who were identified as being aggressive at the pre-intervention stage. Compared to participants in two control conditions, formerly aggressive participants were less likely to perceive ambiguous intentions as hostile and less likely to exhibit a preference for aggressive behavior in response to ambiguous intent. They were also rated as being less aggressive by their teachers. Thus, by increasing participants’ awareness of negative attributions and shifting their thinking to make more positive attributions, the researchers were able to increase receptivity and decrease hostility.

Importantly, bias reduction is most likely to occur when participants are motivated to change. Levy and Maaravi (2018) point out that bias awareness interventions can backfire, or ‘cause a boomerang effect,’ if their recipients perceive the intervention as a threat to their self-image. These researchers attempted to replicate Nasie et al.’s (2014) findings by teaching participants about two different cognitive biases: the ‘halo effect’ and the ‘powerful women’ bias. The halo effect refers to people’s tendency to evaluate someone’s traits based on an initial (usually positive) evaluation that they make of a different trait. For example, when a person finds someone else to be attractive, they might also assume that that person has other positive qualities, such as being smart or friendly. The powerful women bias is a perception that powerful women are less competent than their equivalent male counterparts. The researchers found that teaching participants about the halo effect was successful, whereas teaching participants about the powerful women bias was unsuccessful. Their explanation for why this occurred was that biases that are perceived to be universal (such as the halo effect) are non-threatening, such that participants can acknowledge that they engage in them without facing social consequences. In contrast, biases that can have negative social implications can be threatening to participants’ self-image and therefore harder for them to acknowledge. For instance, admitting to engaging in the powerful women bias may be perceived as tantamount to admitting prejudice against women (i.e., chauvinism). Thus, the researchers argue that it may be necessary to combine awareness training with a complementary intervention that addresses motivational processes simultaneously (e.g., using self-affirmation to reduce the need to preserve one’s self-concept).

Overall, these studies show that debiasing can be effective so long as certain factors are taken into consideration. The effects of debiasing interventions can be moderated by factors such as personality traits, developmental level, cognitive style, and culture. They can also interact with affective and motivational processes in unintended ways, potentially prompting a ‘backfire effect,’ as was found in the study by Levy and Maaravi (2018). These boundary conditions provide evidence to support the idea that researchers should implement more integrative debiasing techniques that consider motivational, affective, and social factors in addition to purely cognitive processes.

Changing Implicit Theories (Mindsets). In addition to teaching people about the downsides of engaging in biased thinking, researchers have also developed interventions that attempt to alter individuals' ‘implicit theories,’ which are beliefs that we hold about the world and human nature (Dweck, 2012a). The most common interventions in this category attempt to change how people think about whether certain human attributes are fixed or malleable. Specifically, they try to shift people from holding an ‘entity theory,’ in which they believe a certain attribute is fixed and/or finite, to an ‘incremental theory,’ in which they believe an attribute is changeable and/or unlimited. In particular, it is thought that fixed mindsets are associated with being motivated to defend or affirm one’s identity, whereas malleable mindsets are more associated with learning goals (Nussbaum & Dweck, 2008). Such interventions can affect a variety of outcomes and tend to be used most often to help boost academic achievement. However, people’s implicit theories about attributes including intelligence, intellectual humility, and empathy can also affect the extent to which they engage in open-minded thinking and behavior (Dweck, 2012b).

Beliefs About Intelligence. A large body of literature has shown that changing people’s implicit theories of intelligence from a fixed mindset to a growth mindset can result in many positive outcomes, including ones related to open-mindedness. For instance, in a study conducted by Porter and Schumann (2018; see Study 4), participants read an article that either described intelligence as a static trait (fixed mindset condition) or a trait that can be developed (growth mindset condition). They found that participants in the growth mindset condition reported being more intellectually humble (i.e. more willing to acknowledge the limitations of their own knowledge) and were more likely to make respectful attributions about a hypothetical classmate who disagreed with them. Relatedly, they found that people who reported being more intellectually humble said that they would be likely to engage in more open-minded behaviors when interacting with the classmate who disagreed with them. Based on their findings, the researchers argue that interventions that aim to boost intellectual humility can improve social interactions between people who disagree, although this has yet to be tested beyond a hypothetical scenario.

Another study, conducted by Yeager et al. (2013; studies 2 and 3), also found that inducing a growth mindset led to a reduction in hostile attributions. In Study 2 of the paper, participants completed a three-part intervention in which they read an article that described the brain's ability to change (neuroplasticity), read notes from older classmates who described the potential for people to change their characteristics, and completed a writing exercise in which they were asked to write to future students about how people's characteristics can change. The control group completed a writing activity in which they described how academic skills can change. The study found that participants in the growth mindset condition were less likely to make hostile attributions about a classmate's behavior in a hypothetical, ambiguous scenario. They were also less likely to want to engage in aggressive behavior toward others. In Study 3, they found that participants who went through the intervention showed significant effects on the same outcomes at follow-up eight months later, proving that these interventions can have long-lasting effects.

Beliefs About Empathy. In addition to changing people's perceptions of intelligence, research has also investigated the effect of changing people's implicit beliefs about empathy. For instance, Schumann et al. (2014) found that participants who reported having a malleable theory of empathy were more likely to try to expend 'empathic effort,' or behave in an open-minded manner, toward someone with opposing views (Studies 2-3). They also used an intervention to manipulate people’s implicit theories about empathy (Studies 4-7). They had participants read an article that either described empathy as malleable or fixed. They found that participants in the ‘malleable’ condition were more willing to listen to outgroup members’ views (i.e., engage in empathic effort) and even to volunteer to participate in empathy training. Schumann and colleagues argue that interventions that focus on changing people’s theories about empathy are more likely to be more effective than simply teaching them skills like perspective-taking, which they might not spontaneously use unless they have the motivation to do so.

Another study tested the effects of a similar, but more in-depth, intervention (Weisz et al., 2020). In this study, participants came into the lab for three separate sessions. They were sorted into one of four conditions: viewing empathy as malleable, learning about social norms around empathy, malleable mindset + social norms, and control (growth mindset of intelligence). During the three sessions, participants engaged in a variety of activities that employed the ‘saying-is-believing effect’ (Hausmann et al., 2008), which included reading articles, reflecting on their own experiences, writing letters to other students, and giving a speech. The researchers found that participants in the two mindset conditions were more likely to believe that empathy is malleable, and that this effect persisted after an eight-week delay. Participants in all three intervention conditions showed greater empathic accuracy for others’ positive emotions right after the intervention and after an eight-week delay. However, none of these interventions increased empathy toward political outgroup members, empathic accuracy for others’ negative emotions, or empathic effort relative to the control condition. The researchers propose that the intervention may have had these mixed results because it focused on strengthening empathic approach motives but not on reducing empathic avoidance motives. Further work is needed to disentangle these varying effects.

Another implicit belief about empathy that seems to be effective in promoting open-mindedness is the idea that empathy is unlimited. In a creative set of six studies, researchers used performance art experiences to manipulate the extent to which individuals perceived empathy as being a limited or unlimited resource (Hasson et al., 2020). The six studies were conducted across many different types of group differences, including ethnic, religious, political, and national. Given that these studies were conducted during performance art experiences, the researchers were able to capture a wide range of self-report, other-report, and behavioral outcomes. The researchers found that participants who were taught to believe that empathy is unlimited experienced greater empathy toward outgroup members, supported prosocial actions toward the outgroup, and displayed more empathic behavior toward the outgroup during face-to-face interactions.

Beliefs About Intellectual Humility. Studies have also found that teaching people that it is beneficial to be intellectually humble can boost self-reported intellectual humility. Such an intervention is similar to growth mindset interventions in that it involves changing the belief that people have about certain human characteristics, which then informs subsequent motives and behavior. Porter et al. (2020; see Study 5) had participants in the intervention condition read a news article about the personal benefits of being intellectually humble (i.e., being able to admit what you do not know). Participants in the opposite condition read an article that touted the benefits of intellectual certainty (i.e., being vocal about showing how much you know). They found that participants in the intellectually humble condition reported themselves as being more intellectually humble than those in the intellectual certainty condition. Participants were also more likely to want to receive further training on a task at which they had previously failed. One limitation to this study is that the participants' intellectual humility was measured using self-report, which may have been subject to demand characteristics. Further research that assesses intellectual humility using other types of measures, such as other-report and/or natural language processing, may be beneficial. However, this study was encouraging in that it suggested that interventions, even small ones, can shift intellectual humility.

Beliefs That Groups Can Change. Other research has attempted to shift people's theories about the ability for entire groups of people to change (group malleability). Halperin et al. (2012) had participants read an article that described groups as either being able or unable to change due to factors like having new leadership. They found that Israeli Jewish participants who were in the malleable mindset condition had more positive attitudes toward Palestinians and were more likely to be willing to compromise (Study 2). Similarly, they also found that the intervention was effective on these outcomes for Palestinian Israelis (Study 3) and Palestinians living in the West Bank (Study 4). Furthermore, in Study 4, they included an additional outcome measure, finding that the West Bank Palestinian participants were 70% more likely to be willing to meet and listen to the viewpoint of an Israeli Jew.

Perspective-Taking. In contrast to interventions that focus on changing people's implicit theories, other interventions focus more on helping participants practice the cognitive skill of taking other people's perspectives. A large body of literature is dedicated to the technique of 'perspective-taking,' which attempts to help individuals adopt a new perspective, or put themselves 'in another person's shoes' (for a review, see Todd & Galinsky, 2014). For this reason, we will provide a more high-level overview of how perspective-taking interventions tend to be implemented, along with their potential pitfalls and boundary conditions.

In most perspective-taking interventions, participants review a photograph, video, or recording of a specific individual. Then, they write about a day in the life of that person, imagine the person's mental states, and/or imagine what it would be like to think like that person or experience their situation. Sometimes, participants are asked to 'put themselves in the other person's shoes,' imagining the feelings and thoughts that they would have if they were in the other person's situation. Other times, they are asked just to imagine what the other person thinks and feels. Recently, researchers have also incorporated more advanced augmented or virtual reality (AR or VR) technology that allows participants to experience the world from another person's perspective, which some researchers have argued is more powerful (Herrera et al., 2018; Van Loon et al., 2018; Yee & Bailensen, 2009). Given that these interventions essentially 'give' participants a perspective to understand rather than requiring them to imagine it, it may be appropriate to group them with other 'perspective-getting' interventions, which tend to focus on helping people to more accurately understand another person's perspective by asking them about it (Eyal et al., 2018; see section on Sustaining Open-Mindedness Through Social Skills).

Perspective-taking interventions have been used to manipulate many concepts that are related to open-mindedness, including prejudice and intergroup empathy. In their comprehensive review of literature on intergroup perspective-taking, Todd and Galinsky (2014) describe how perspective-taking improves explicit and implicit evaluations of outgroup members, strengthens approach-oriented reactions, increases non-verbal positivity and rapport, facilitates intergroup contact experiences, and undermines stereotype maintenance. Todd and Galinsky suggest that perspective-taking reduces outgroup bias and improves intergroup relations through multiple mechanisms. It reduces biased attributions (Regan & Totten, 1975; Todd et al., 2012; Vescio et al., 2003), increases perceptions of self-other overlap (Davis et al., 1996; ; Galinsky et al., 2005; Todd & Burgmer, 2013), increases empathy toward outgroup members (Batson et al., 1997; Dovidio et al., 2010), and decreases stereotype accessibility and ingroup favoritism (Galinsky & Moskowitz, 2000).

Although perspective-taking can improve attitudes and behavior toward outgroup members, it can also backfire. Perspective-taking can serve to highlight 'unbridgeable' differences between people (Okimoto & Wenzel, 2011). It can also expose individuals to alternative viewpoints that they perceive as threatening and desire to distance themselves from (Catapano et al., 2019; Paluck, 2010). Furthermore, taking another person's perspective can activate meta-stereotypes, making participants more aware of how their views are likely to be perceived by the other person. Sassenrath et al. (2016) argue that individuals are likely to assume that others who have limited information about them and with whom their group has had conflict in the past are likely to form negative evaluations of them. Thus, concerns about negative self-evaluations can reduce the effectiveness of perspective-taking.

Research has shown that boomerang effects are most likely to occur among individuals who identify strongly with their ingroup (Tarrant, Calitri, & Weston, 2012; Zebel et al., 2009). Non-dominant group members are especially likely to exhibit strong identification with their group, and therefore, they tend not to benefit, and can even suffer adverse consequences, from perspective-taking exercises (Bruneau & Saxe, 2012). For these individuals, perspective-taking must compete with a strong motivation for maintaining their social identity (Jetten et al., 2004), and thus, these individuals are resistant to increased self-other overlap. These findings suggest that interventions that employ perspective-taking should also employ techniques that address motivational concerns for self-integrity. Such interventions include self-affirmation (to be described further in a later section), and emphasizing shared values or similarities (Catapano et al., 2019; McDonald et al., 2017).

Many perspective-taking interventions have been based on the assumption that when simulating the minds of others, individuals are likely to be accurate in their perceptions. However, given a rich understanding of the role of cognitive biases in social perception, researchers have begun to question this assumption. For instance, in a series of 25 studies, Eyal, Steffel, and Epley (2018) found no evidence that perspective-taking improves accurate understanding of another person's viewpoint, despite individuals' intuition that it would. They argue that perspective-taking does not give individuals access to new information; instead of gaining access to accurate information, individuals must rely on their stereotypes of others, which can be biased.

Thus, perspective-taking can backfire by increasing biased perceptions. For example, Skorinko and Sinclair (2013) found that perspective-taking increased reliance on stereotypes during decision-making by making stereotypes more salient when individuals simulate the mind of an individual who displays stereotype-consistent traits. Aside from increasing stereotyping, in competitive contexts, perspective-taking can also lead to reactive egoism, or increases in selfish behavior (Epley et al., 2006; Pierce et al., 2013). Individuals assume that their competitors have selfish motives, so when they imagine what it is like to be in a competitor's shoes, individuals defend themselves against being taken advantage of by acting selfishly in return. However, Epley et al. also found that highlighting shared goals can promote a more cooperative environment, which can facilitate reduced egoistic behavior in conjunction with perspective-taking.

Clearly, it is important that people are accurate in their understanding of others' perspectives when engaging in perspective-taking. Certain perspective-taking interventions, such as those that employ AR/VR, may be more effective at giving people an accurate view of alternative perspectives. However, research has found that these interventions tend to be person-specific rather than generalizable (Van Loon et al., 2018). Furthermore, they may not be scalable given cost and accessibility concerns. Another technique that has been shown to promote a more accurate exchange of information, which is referred to as perspective-getting, encourages people to ask another person what they believe rather than making assumptions (Eyal et al., 2018).

Yet another technique that has been used to promote more accurate perspective-taking is holding people accountable to the target of their perspective-taking. For instance, Tuller et al. (2015) conducted four studies to examine how perspective-taking might change people's views on controversial issues (e.g., weight discrimination and abortion). In all four studies, they had participants engage in 'relationship forming' with someone with opposing views (either in person or through reviewing a previous participant's responses). Then, they had participants articulate the other person's opinion on the controversial issue. The researchers found that perspective-taking was only successful in reducing the extremity of people's views when participants met the person who they thought had opposing views and were also told that the other person would be reviewing what they wrote for accuracy purposes. By holding participants accountable, this approach induced accuracy motives to complement the perspective-taking intervention. We will discuss other accuracy-inducing interventions in the section on inducing open-mindedness through motivational pathways.

Paradoxical Thinking. Another strategy to induce open-mindedness involves asking people leading questions or presenting them with arguments that contain exaggerated versions of their beliefs (Knab et al., 2021). According to the researchers, this technique is effective because it proposes an attitude that falls within a person's 'latitude of acceptance' (the range of opinions that they consider to be acceptable), and therefore does not raise a defensive response. However, given that the attitude is extreme, this surprises participants, and ultimately leads them to reflect on and reconsider their own stance. Knab et al. propose that the underlying mechanism that causes this effect is increased cognitive flexibility.

In one study, researchers presented Israeli Jews with the following leading question: "Why do you think that the real and only goal the Palestinians have in mind is to annihilate us, in a manner that transcends their basic needs such as food and health?" (Hameiri et al., 2018). While Jewish Israelis might believe that Palestinians have been causing conflict with them, most would not agree with this extreme version of that opinion. When participants responded to this question, they were more likely to report that they had reconsidered their beliefs regarding the Israel-Palestine conflict. In another study, Israeli Jews in the intervention condition watched video clips that argued that they centered their identity around experiencing conflict and that they could not afford to end the conflict with Palestine (Hameiri et al., 2014). The researchers proposed that this idea was attitude-congruent but also extreme, explaining that most Israeli Jews tend to think of the conflict as necessary, but not core to their identity. Participants in the control condition watched videos about tourism in Israel. The study found that participants in the paradoxical thinking condition reported that they had reevaluated their opinions and reported that they were willing to endorse compromising with Palestine. Furthermore, a greater percentage of participants in the intervention condition voted for ‘dovish’ parties who support a peaceful resolution of the conflict as compared to the control condition.

However, research in this domain shows that paradoxical statements of questions cannot be too extreme if they are to be successful. If they are exaggerated enough, they fall into a participant’s latitude of rejection (the range of opinions that the participant considers to be unacceptable), whereby they immediately dismiss them and do not reevaluate their own views. Hameiri et al. (2020) argue that paradoxical statements should aim for a ‘sweet spot’ in which they are only slightly exaggerated. They tested this with regard to Israeli Jews’ opinions about refugees and asylum seekers. They divided participants into four conditions and asked them to read a news article that was consistent with their views (i.e. it proposed that Israel should not provide refugees with health care). In two of the conditions, participants read articles that were not exaggerated. In the third condition, the article made an argument that was slightly exaggerated. In the fourth condition, the article’s argument was extremely exaggerated. They found that the only condition in which participants reported reevaluating their beliefs was in the third condition (the ‘sweet spot’).

Puncturing the Illusion of Explanatory Depth. Another intervention that aims to encourage participants to reflect on their thoughts and opinions involves taking participants through an exercise in which they realize they know less than they thought. In the literature, researchers refer to this as ‘puncturing’ a bias called ‘the illusion of explanatory depth,’ whereby people think that they know more about complex phenomena than they really do (Rozenblitz & Keil, 2002). These studies first ask participants to rate their level of understanding with regards to a complex phenomenon (e.g., how toilets flush, how the brain coordinates behavior, or how the United States Supreme Court determines the constitutionality of laws). Subsequently, participants are asked to write a detailed and step-by-step, causal explanation of how the phenomenon works. Then, they read an article that actually explains how it works, which tends to reveal that the participant knew less about the phenomenon than they thought. Finally, participants rate how well they actually understood the phenomenon prior to learning about it from the article.

Studies have found that puncturing the illusion of explanatory depth reduces participants' overconfidence in their own knowledge (Rozenblitz & Keil, 2002; Fernbach et al., 2013; Voelkel et al., 2017; Crawford & Ruscio, 2021). Yet findings have been mixed with regard to the impact of this intervention on political attitudes. Fernbach et al. (2013) had participants explain complex political policies in detail. They found that participants reported having less extreme political attitudes after going through the intervention, as compared to participants who were told to enumerate the reasons for their political position. Another study had participants merely "reflect on how well you could explain to an expert, in a step-by-step, causally-connected manner the details of … [a sociopolitical] issue. (Johnson et al., 2016; see Experiment 9). Similarly, they found that the intervention reduced participants' overconfidence and attenuated the extremity of their attitudes. However, recent research had more mixed results (Crawford & Ruscio, 2021). In attempting to replicate the study by Fernbach et al. (2013), the researchers found that the intervention reduced overconfidence but did not affect attitude extremity. Further work will be required to better understand these effects. However, It may still be possible that puncturing the illusion of explanatory depth is an effective technique for promoting open-mindedness, if not attitude change. Light and Fernbach (2020) propose that the illusion of explanatory depth and other 'knowledge calibration' techniques can help to promote intellectual humility.

Correcting False Meta-Perceptions. Other research has found that giving participants feedback about the accuracy of their meta-perceptions about people with opposing views reduces their bias toward them. For instance, Lees and Cikara (2020) found that participants thought that members of the political outgroup felt more negative toward their ingroup than the outgroup members really did. The researchers also found that a simple intervention was effective at mitigating this bias. They showed participants their own estimates of the outgroup's beliefs alongside data that revealed the outgroup members' actual (more positive) beliefs. They found that showing participants this corrective feedback led to reductions in their negativity bias. Researchers found that this intervention effect replicated in nine out of ten countries (Ruggeri at al., 2021).

Moore-Berg et al. (2020) propose that meta-perceptions may be easier to correct than first-order beliefs: "convincing people that they are wrong about others' minds may be easier than convincing them they are wrong about their own minds." To explain this, the researchers suggest that this is because meta-perceptions are reliably false and pessimistic. Lees and Cikara (2021) propose that people are more open to corrections to their meta-perceptions because they are motivated to manage their reputation. In order to manage the impression that others have of them, people need to have an accurate understanding of what that impression is. Overall, this is a nascent area of research that demonstrates promising effects for promoting open-minded thinking.

Domain-General Cognitive Approaches

In addition to debiasing training, mindset interventions, and perspective-taking, other interventions have attempted to improve open-mindedness through more domain-general cognitive pathways. These interventions involve broadening and complexifying thinking patterns (again, promoting System 2 over System 1 processing) through techniques such as cognitive disfluency, self-distancing, priming creativity, mood inductions, and cognitive training.

Cognitive Disfluency. Researchers have found that cognitive disfluency (i.e. making text difficult to read) can induce analytical (System 2) thinking, which tends to be less prone to cognitive biases (Alter et al., 2007). Given this effect, studies have tested whether disfluency can improve open-mindedness toward others. For instance, in a study conducted by Yang et al. (2013), the researchers had participants read a passage that was either easy or hard to read (Study 3). Following the manipulation, participants read about a proposal to build a mosque near the 9/11 Ground Zero site and then provided reactions to the proposal (a composite of behavioral, affective, and cognitive measures). Both conservatives and liberals who viewed the hard-to-read passage showed less polarized attitudes compared to those who viewed the easy-to-read passage. In a similar study on fluency and the confirmation bias, Hernandez and Preston (2013) found that presenting participants with counter-attitudinal information in a hard-to-read font reduced the extent to which participants evaluated the information in a biased and extreme way. Thus, disfluency might promote more thorough consideration of counter-attitudinal information as opposed to 'knee-jerk' reactions against it.

Self-Distancing. Another approach that induces a more analytical and abstract thinking style, called self-distancing, encourages people to transcend beyond their own egocentric view of the world. According to construal level theorists, taking a psychologically distant perspective induces people to be in an abstract, rather than concrete, mindset (Trope & Liberman, 2010). Moreover, self-distancing is thought to lead people to focus less on the emotionally arousing components of their memories and more on self-reflection (Kross & Ayduk & 2017). This approach involves having people remember events that happened to them and view them from an outsider's perspective. For instance, they might imagine themselves watching the event as a 'fly on the wall' or refer to their past self using third-person pronouns. In contrast, a person who is in a self-immersed mindset might remember past experiences by reliving them 'through their own eyes' from a first-person perspective. Studies have found that self-distancing increases creativity, improves problem-solving, reduces negative affect, reduces physiological stress, and reduces emotional reactivity (Ayduk & Kross, 2011; Förster et al., 2004; Jia et al., 2009; Kross & Ayduk, 2017). Self-distancing has been used as one form of reappraisal, an emotion regulation technique that will be discussed in this review's section on Sustaining Open-Mindedness Through Emotion Regulation. However, in this section, we will discuss how self-distancing can induce open-minded thinking in the first place.

One study found that participants who reasoned about personal issues from a distant (versus immersed) perspective were more willing to express intellectual humility (i.e., 'recognize the limits of their knowledge'), endorse more moderate political opinions, and report being willing to join a bipartisan group that would discuss political issues (Kross & Grossman, 2012). Another study involved training people over the period of one month to reflect on interpersonal challenges from a third (versus first) person perspective (Grossman et al., 2021). The researchers found that participants in the third person (self-distancing) condition showed improvements in intellectual humility, acknowledgement of diverse viewpoints, and search for conflict resolution. In a follow-up study, the researchers demonstrated the same effects over the course of a week, suggesting that self-distancing training can be achieved over a shorter timespan.

Priming Creativity. Another method for broadening thinking styles is priming creativity. Creativity is closely linked to flexible or divergent thinking and the personality trait of openness to experience (Chi, 1997; McCrae, 1997). Much of the literature on creativity has measured creativity as a mere correlate or outcome of other personality factors, but recent work has explored how creativity might foster open-mindedness by disrupting traditional patterns of thinking (e.g., stereotypes). For instance, Sassenberg and Moskowitz (2005) found that compared to control participants who displayed automatic stereotype activation about African-Americans, participants who remembered times they had 'behaved creatively' showed no stereotyping. The researchers argue that a strength of this manipulation, compared to other interventions such as perspective-taking, is that the intervention does not have to be tailored with regard to reducing stereotypes about a specific people or groups — its mechanisms seem to be more domain-general.

Creativity has also been tested in terror management theory research as a defense against the threat of mortality salience, which is thought to restrict one's worldviews and promote defensiveness. Routledge et al. (2004) found that following a mortality salience manipulation, American participants who designed a 'creative t-shirt' reacted less negatively to an essay with an opposing viewpoint criticizing American culture. In a follow-up study, Routledge & Arndt (2009) found that following a mortality salience manipulation, when Americans read that other Americans valued creativity, participants became more open to learning about alternative cultural and religious viewpoints. Although these studies tested how creativity might buffer against mortality threats, it would be useful to examine whether creativity can reduce defensive responding to threatening viewpoints during interaction. Furthermore, it would be useful for future studies to examine potential mechanisms for how creativity reduces defensive responding.

Positive Mood Inductions. Research has found that putting participants into a positive mood can also broaden their thinking. Positive mood inductions include providing participants with refreshments, giving them a small gift, having them watch a short comedy clip, priming them with positive statements, or having them recall a positive memory. Overall, the findings in this domain are mixed. According to Fredrickson's 'broaden-and-build theory' (Fredrickson, 2001; Fredrickson, 2004; Fredrickson & Branigan, 2005), positive emotions broaden individuals' 'thought-action repertoires.' According to this theory, inducing a positive emotion should expand a person's mind so that they can come up with more thoughts and potential actions. In contrast, the 'mood-as-information' approach (Schwarz, 2000) suggests that moods signal information to individuals about their situation and guide them to react accordingly. According to the mood-as-information approach, positive moods should signal the absence of a threat and lead individuals to rely more on heuristic thinking, whereas negative moods should signal that the individual needs to be alert to a potential problem in the environment and result in more deliberative processing.

In support of the 'broaden-and-build' theory, mild positive affect inductions have been shown to enhance cognitive flexibility (Murray, Sujan, Hirt, & Sujan, 1990), promote creativity (Isen et al., 1985; Isen, Daubman, & Nowicki, 1987) and reduce biased anchoring effects (Estrada, Isen, & Young, 1997). Research has also found that positive mood inductions can promote open-mindedness. Nelson (2009) conducted two studies in which she induced participants to be in a positive, negative, or neutral affective state. In Study 1, she had participants either write about their morning routine (neutral condition) or about a time when they were elated, joyful, or proud (positive affect). In Study 2, participants read a series of statements out loud that were positive (e.g., "Most people like me."), negative (e.g., "Nobody understands me or even tries to."), or neutral (e.g., "It snows in Idaho."). In these studies, Nelson found that participants in the positive condition were more likely than participants in the neutral and negative conditions to engage in cognitive perspective-taking and to express more empathic concern for dissimilar others.

In contrast, although these studies found that positive affect led to increased open-mindedness, other researchers have found that positive mood can impair cognitive functioning, such as planning (Oaksford et al., 1996), working memory (Spies et al., 1996), and task switching (Phillips et al., 2002). In the domain of social cognition, Park and Banaji (2000) had participants watch a 10-minute video that was happy, neutral, or sad. They found that inducing happiness led participants to rely more on stereotypes when making social judgments; in contrast, sadness led to less stereotyping. They argued that positive affect led to heuristic processing and negative affect led to detail-oriented thinking, in line with the 'affect-as-information' approach.

To reconcile the research showing that positive mood improves cognitive functioning in some situations and inhibits it in others, Mitchell and Phillips (2007) propose that positive mood generally leads to heuristic thinking, but that motivational factors can modify its effects, such that positive mood can be beneficial when situations involve novel information-seeking. Further studies are needed to tease apart the contexts in which positive affect is beneficial for promoting open-mindedness. Researchers should consider combining a positive mood induction with other manipulations that might encourage information-seeking. It is also necessary for researchers to determine how much positive affect should be 'administered,' and how long the effects of positive mood inductions last. In summary, affect inductions may be useful tools in promoting greater cognitive flexibility, but also might increase reliance on heuristics; indeed, further work is required to better understand these effects.

Cognitive Training. Other more domain-general approaches for inducing open-minded thinking include adaptive, cognitive training and neurofeedback. These approaches attempt to improve low-level cognitive mechanisms that underlie open-mindedness. For instance, one study focused on improving participants' 'metacognitive awareness' – which is also referred to as 'confidence calibration' or 'introspective ability' (Carpenter et al., 2019). Prior work has found that people who are dogmatic have impaired metacognitive abilities, which suggests that cognitive training that focuses on improving these abilities may be beneficial for boosting open-mindedness (Rollwage et al., 2018; Rollwage & Fleming, 2021). In the study by Carpenter et al., participants completed 8 sessions during which they completed a perceptual discrimination task and received feedback. In the intervention condition, participants received feedback with regards to the accuracy of their metacognitive judgments (i.e., the extent to which their confidence ratings aligned with their performance). In the control condition, participants received feedback about their task performance alone. The study found that only participants in the intervention condition showed improvements in their introspective abilities.

Another cognitive training study used video games to improve participants' cognitive flexibility (Glass et al., 2013). This work may be relevant for increasing open-mindedness given that prior research has found that individuals with more extreme attitudes also exhibit cognitive inflexibility (Zmigrod et a., 2020). In the video game training study, participants in the intervention condition played a 'real-time strategy' video game called StarCraft, while participants in the control condition played a 'life simulation' game called the Sims 2 over the course of 40 hours. The study found that participants in the intervention condition showed improvements in their cognitive flexibility.

The studies reviewed in this section thus far were conducted in highly controlled laboratory settings in which academics followed rigorous experimental protocols. However, when it comes to commercial brain training — which tends to consist of brief cognitive games that have been developed by companies — many researchers disagree about the extent to which they are effective, primarily because their effects do not often transfer to improved cognitive performance on other tasks (Owen et al., 2010; Simons et al., 2016). Furthermore, to our knowledge, no published research has measured the impact of these low-level cognitive trainings on open-mindedness specifically (only on its underlying mechanisms). Thus, further work will be needed in order to determine whether these approaches are effective at boosting open-mindedness.

Inducing Open-Mindedness Through Motivational Pathways

When people are processing viewpoints that challenge their opinions, motivational factors impact their open-mindedness toward those viewpoints in addition to cognitive factors. In their wide-ranging review of 'wise interventions,' which they define as interventions that aim to increase human flourishing, Walton and Wilson (2018) focus on three primary motivations that are relevant to open-mindedness: 'the need to be accurate,' 'the need for self-integrity,' and 'the need to belong.' Similarly, according to Van Bavel and Pereira's (2018) 'identity model of beliefs,' individuals balance accuracy goals against identity goals (e.g., belonging, epistemic, existential, status, system justification, and moral goals) when they process information. The researchers argue that the mind places a 'weight' on all accuracy and identity goals as a function of an individual's disposition and their social context.[4] They argue that people try to process information in an accurate and unbiased

[4] The researchers developed this conceptual formula to illustrate their model, where V represents the value placed on holding accurate beliefs and w represents the weight put on each goal. V =

manner when the weight placed on accuracy is larger than the net weight of the identity goals. However, people tend to engage in biased thinking when their identity goals supersede accuracy goals. In particular, political partisanship tends to satisfy identity goals, and when there is a large weight placed onto those goals, it can distort information processing and lead to rigid, dogmatic thinking. Based on these models, we propose that interventions can take three routes to promote open-mindedness toward alternative viewpoints: upweight accuracy motives, preemptively satisfy the need for self-integrity, and/or leverage the need to belong.

Promoting Accuracy Motives

Studies have shown that promoting the goal of processing information accurately can help to reduce biased thinking (Kunda, 1990; Lener & Tetlock, 1999). Incentivizing people to be accurate (e.g., through monetary rewards) can help to mitigate bias. For example, Waytz, Young, and Gingest (2014) offered participants incentives to provide accurate evaluations of the opposing political party. They found that incentivizing accuracy mitigated a bias called ‘motive attribution asymmetry,’ wherein participants tend to attribute positive motives to their political ingroup and negative motives to the political outgroup. Furthermore, incentives also improved participants’ willingness to negotiate with the opposing party, improved their optimism around being able to reach a compromise, and reduced their tendency to hold negative, essentialist beliefs about the other party. Similarly, Bullock et al. (2013) found that paying participants when they answered accurately (or admitted that they did not know the answer) reduced party differences in response to questions about politics. For example, without incentives, liberals and conservatives provided different responses to the following question: “Compared to January 2001, when President Bush first took office,

$w_1\text{Accuracy} - \sum(w_2\text{Belonging} + w_3\text{Epistemic} + w_4\text{Existential} + w_5\text{Status} + w_6\text{System} + w_7\text{Moral}...w_n\text{OtherGoals})$

has the level of inflation in the country increased, stayed the same, or decreased?” With incentives, this gap between the parties shrank.

Another method for promoting accuracy is holding people accountable for being accurate. According to Lerner and Tetlock (1999), informing people that they will be held accountable can boost open-minded thinking under specific conditions. Methods for convincing people that they will be held accountable include telling them that they will be evaluated, telling them they will have to justify their responses, and telling them their responses will be made public (Kunda, 1990). Lerner and Tetlock argue that accountability interventions are most effective when people are told “they will be accountable to an audience (a) whose views are unknown, (b) who is interested in accuracy, (c) who is interested in processes rather than specific outcomes, (d) who is reasonably well-informed, and (e) who has a legitimate reason for inquiring into the reasons behind participants’ judgments.” According to their review of a wide range of accountability studies, accountability interventions can attenuate a wide range of cognitive biases that are related to open-mindedness, including making biased attributions and stereotyping.

Research has also shown that priming accuracy goals with subtle manipulations can reduce the extent to which people report that they would share political misinformation (Pennycook et al., 2021). The researchers effectively primed accuracy using two methods. One method involved showing participants a politically neutral headline and asking them to rate the headline’s accuracy. The other method involved asking participants to indicate whether or not they agreed that “it is important to only share news content on social media that is accurate and unbiased.” Van Bavel and Pereira (2018) have also suggested that priming people to think “like scientists, jurors, or editors” might also help to promote accuracy motives.

Satisfying the Need for Self-Integrity

In addition to upweighting accuracy goals, researchers have also tested the efficacy of satisfying identity motives to help reduce biased information processing. In particular, interventions can either try to satisfy a need to defend one's self-integrity or leverage one's need to belong. The thinking behind interventions that focus on self-integrity is that alternative viewpoints can serve as potential threats to a person's need to be accurate and consistent in their beliefs. Instead of being open to being wrong when someone disagrees with them, many people will double down on their own beliefs in order to preserve the idea that they are accurate and consistent. In an individual context, this can translate to engaging in motivated reasoning, confirmation bias, and selective exposure. Interpersonally, people tend to defend their views rather than consider that they could be wrong. Interventions can try to preemptively buffer against these threats by satisfying those needs beforehand. Once individuals' needs are satisfied, they do not need to interpret information in a biased way in order to fulfill their needs.

One of the primary techniques that has been developed to fulfill individuals' need for self-integrity is self-affirmation (Steele, 1988). Self-affirmation is a process by which individuals affirm important values (e.g., family, friendship, etc.). Typical manipulations involve asking participants to rank their values and then to write about times in the past when those values were important. The effects of self-affirmation are transferable, such that an individual can receive self-affirmation with regard to a specific set of values, and that manipulation can buffer against threats to self-integrity in a different domain (Steele, 1998).

In reducing defensive responding, participants are able to engage in less biased, more objective consideration of information. For instance, Cohen et al. (2000) found that affirmed participants were more persuaded by evidence contrary to their own political views, as well as more critical of an argument put forward by an individual who shared their political views. Relatedly, Correll et al. (2004) found that participants were more critical of views expressed by an ingroup member and more sensitive to the strength of arguments from ingroup and outgroup members, though they did not find that participants were more receptive to views expressed by out-group members. Ward et al. (2011) found that self-affirmation reduced the extent to which students derogated a concession from their professor (i.e., reduced their 'reactive devaluation' of the professor's concession), an effect which was not explained by distraction or explicit mood enhancement. Finally, Binning et al. (2015) found that American participants who self-affirmed were more convinced by factual evidence about the nation's economy rather than national polling data (normative information) when it came to evaluating President Obama's policies. These studies provide some evidence to suggest that self-affirmation leads to more objective evaluation, and in some cases, more open consideration of alternative views.

In addition to measuring the effects of self-affirmation on information processing via self-report, recent research has involved using neural measures to assess the impact of these interventions. Using a technique called the 'neural reference groups approach' (Dieffenbach et al., 2021), Dieffenbach et al. (in prep) found that self-affirmation altered participants' neural processing in the brain's mentalizing network when they were watching videos containing political opinions that they disagreed with. This study suggests that neuroimaging can help to 'get under the hood' in assessing the impact of self-affirmation interventions, especially when self-report is found to be unreliable. However, this approach has not yet been used to determine whether self-affirmation reduces biased information processing, but only to show that some change has occurred.

Despite evidence to support that self-affirmation is effective, other work has shown that self-affirmation interventions may be delicate and only work under certain boundary conditions. For instance, Cohen et al. (2007) found that self-affirmation was only effective when an individual's beliefs about the issue were made salient. This effect occurred in the context of individuals reading a counter-attitudinal report and also engaging in a negotiation with a confederate who purportedly held opposing views. Participants who were affirmed and had their identity made salient provided more positive evaluations of the opposing viewpoint and made more concessions during the negotiation. In contrast, participants who were affirmed and who were instructed about the importance of compromise made fewer concessions. According to Cohen and colleagues, identity salience serves to alert individuals about the stakes of coming to a compromise with another party. In contrast, simple instructions about the virtues of compromise and rationality may make the individual focus more on behaving according to those virtues, but less focused on the outcome of their interaction.

Self-affirmation interventions that do not manipulate identity salience may be less effective. A recent paper found that several previously unpublished studies had failed to find significant effects in terms of self-affirmation on political outcomes (Lyons et al., 2021). They found that self-affirmation did not affect a wide range of outcomes related to open-mindedness, including belief superiority, affective polarization, evaluation of news sources, and endorsement of factual beliefs. However, none of these studies manipulated identity salience. The researchers posit that it is possible that these studies were not effective because of this. They also argue that it is possible that self-affirmation does not actually work, and that the findings of previous self-affirmation studies may have been spurious due to small sample sizes, inconsistent methods, and different contingent effects.

Thus, further work is required in order to understand how self-affirmation affects open-mindedness and under what conditions.

Leveraging the Need to Belong

In addition to satisfying identity motives as a way to downweight their effect on cognition and behavior, some interventions focus on leveraging identity motives to encourage people to be open-minded. In particular, given people's need to belong, social norms can be a powerful force for encouraging open-mindedness. This effect has been demonstrated in terms of people's self-reported empathy toward others. Weisz et al. (under review) found that students who received a norms-based intervention reported being more motivated to empathize with others and ultimately engaged in more prosocial behavior. Another study found that people expressed feeling more empathy toward people in distress after they had seen that on average, most participants who had completed the study before them had reported feeling empathy toward those people (Nook et al., 2016). In these studies, individuals felt and expressed more empathy toward others because of their desire to belong and conform to group norms that highlighted others engaging in empathy.

Other research has examined the effects of activating affiliative motives in interpersonal contexts. In a study by Chen et al. (1996), the researchers told participants that they were going to be interacting with another participant (although they did not actually end up interacting with the partner). Beforehand, they primed participants with accuracy motives or impression-based motives. To do this, they asked participants to imagine they were in hypothetical scenarios and to write out what they would do. For instance, in one of the accuracy prompts, participants were asked to imagine being a reporter trying to identify the facts of a story. In one of the impression prompts, participants were asked to imagine that they had been set up on a blind date with their friend's cousin who they were not attracted to. Participants also learned about the attitudes of their future interaction partner on a particular social issue. Then, they indicated what their own attitudes were on the issue. The researchers found that participants in the impression condition were more likely to conform their attitude toward their partner's attitude, whereas participants in the accuracy condition seemed to show no conformity effect. This study is relevant to promoting open-mindedness for two reasons. On the one hand, it shows that priming a desire to belong or conform can be effective at promoting agreement between participants. On the other hand, it also shows that accuracy primes lead participants to not shift their attitudes toward a partner's. However, the goal of open-mindedness interventions is not to change people's opinions, but to increase the respect that they have for alternative opinions. Therefore, the decision of whether to promote accuracy or promote affiliation may be dependent on the context (i.e., social or non-social) and what outcomes are desired.

Finally, another technique that has been used to promote social goals is through changing people's perceptions of time. According to socioemotional selectivity theory, those who perceive their time to be limited place higher priority on the immediate goals of emotion regulation and social connectedness; in contrast, individuals who perceive time as expansive prioritize future-related goals such as knowledge acquisition instead (Carstensen, Isaacowitz, & Charles, 1999). In applying this theory to promoting open-mindedness, one might make two competing predictions. If perceiving time to be expansive encourages individuals to seek out knowledge and interact more with strangers, then they might become more open to novel and alternative viewpoints through increased exposure. Alternatively, as DeWall, Visser, & Levitan (2006) propose, perceiving time to be limited might encourage individuals to be receptive to others in order to ensure that their social interactions remain positive. Using a simple manipulation in which college students read an essay focusing their attention on their upcoming graduation, DeWall et al. (2006) found that inducing a limited temporal perspective led participants to endorse "going along to get along" over staunchly defending their views, and to change their attitudes to be more in line with a future discussion partner with an opposing view. Thus, inducing individuals to perceive time as being limited could improve motivation to affiliate with others who hold opposing views and to engage in emotion regulation that maintains positive interactions.

Sustaining Open-Mindedness Through Emotion Regulation

In the previous sections, we discussed interventions that attempt to induce open-mindedness by targeting cognitive and/or motivational processes within the individual. In this section, we will discuss interventions that attempt to sustain open-minded thinking by helping people to regulate their emotions. These interventions focus on teaching people to regulate their negative emotions through direct or indirect methods. Researchers have noted that in addition to training people to regulate their emotions, it is also important to ensure that they are motivated to regulate their emotions in intergroup contexts (Halperin, 2014).

Cognitive Reappraisal

While positive mood inductions focus more on making individuals open-minded in the first place, other interventions teach emotion regulation strategies in order to help people remain open-minded when they encounter alternative viewpoints in a social context. Two of the most commonly studied emotion regulation strategies are emotion suppression and cognitive reappraisal. Emotion suppression, which involves inhibiting an emotion from occurring, tends to reduce behavioral expressions of negative emotion but not negative emotional experiences. In contrast, cognitive reappraisal involves reinterpreting the meaning of a stimulus in order to minimize its emotional impact. This strategy is thought to decrease both the behavioral expression and experience of negative emotions (Gross, 2002). In the context of interactions between individuals with different viewpoints, it is important to reduce their expressions and experiences of negative emotions in order to sustain a constructive conversation. Thus, interventions have focused on the relationship between cognitive reappraisal and improving relations between those with opposing views (Halperin & Gross, 2010; Halperin & Tagar, 2017; Halperin, 2014).

To begin with, researchers found a positive correlation between Israelis' tendency to spontaneously reappraise negative emotions and their support for policies that would provide aid to Palestinians (Halperin & Gross, 2011). Next, researchers manipulated cognitive reappraisal experimentally. Halperin et al. (2012) taught Israeli participants to engage in reappraisal by handing them anger-inducing pictures and asking them to respond to the pictures "like scientists, objectively and analytically." Following the training, participants were instructed to apply this technique while they were presented with information about the Israeli-Palestinian conflict. In two studies, the researchers found that participants who were trained in reappraisal supported conciliatory policies between Israelis and Palestinians. Furthermore, this effect was long-lasting — differences between the intervention and control group remained at a five-month follow-up. The researchers also found that the effect was mediated by reductions in anger toward Palestinians, suggesting that the reappraisal training led participants to downregulate their negative affect in relation to the political issue.

Recently, researchers tested the efficacy of a mobile game intervention, ReApp, that teaches reappraisal strategies (Porat et al., 2020). In the experiment, Jewish-Israeli participants learned about the strategy of reappraisal. Then, they were paired with a partner and practiced reappraising one another’s emotions in response to images (e.g., reappraising extreme sadness in response to a picture of a dog in a cage). Participants who played ReApp (as opposed to Connect Four in the control condition) experienced lower levels of disgust and anger and expressed less support for aggressive policies against Palestinians. This study provides encouraging evidence that reappraisal can be taught in a scalable manner that does not require in-person training or feedback.

Although training in cognitive reappraisal can be effective, researchers argue that this intervention may only be successful when people are motivated to regulate their emotions. As Halperin et al. (2014) point out, people who are involved in intractable conflicts are likely to be driven by a motive to maintain their group identity, and reacting negatively to an opposing group can make up part of that group identity. In fact, Tamir et al. (2019) suggest that most cognitive reappraisal interventions involve simultaneously activating an emotion goal — such as decreasing negative emotions — and also providing participants with the means to achieve that goal. They conducted a study among Israeli participants to test the effects of three conditions in decreasing anger toward a video depicting Palestinians: control (watch video naturally), emotion goal (telling participants to decrease their emotion reaction), and emotion goal + reappraisal training. They found that activating the emotion goal alone was as effective as the emotion goal + reappraisal condition in terms of decreasing self-reported negative emotions and angry facial expressions (as measured by lower corrugator activity) in comparison to the control condition. They concluded that may be sufficient to activate emotion goals, rather than teaching the technique of reappraisal, in order to reduce negative emotions.

Indirect Emotion Regulation

Given that individuals may not always be motivated to regulate their emotions, some researchers have advocated for the potential of ‘indirect emotion regulation’ strategies as a means to reduce negative emotions in intergroup contexts. Indirect emotion regulation works by targeting an emotion to alter, identifying a cognitive appraisal that underlies that emotion, and then altering that underlying appraisal. In particular, the interventions that have been tested in this domain manipulate a fixed versus malleable mindset. In this review, we have placed these strategies in the cognitive section given that they explicitly target cognitive mechanisms, but it is worth noting here that they have downstream consequences for affect. According to researchers that identify these as indirect emotion regulation strategies, instilling the belief that an opposing group or a conflict situation is malleable rather than fixed leads to reduced anger toward an opposing group (Halperin et al., 2011), greater perceptions of hope (Cohen-Chen et al., 2014), and reduced intergroup anxiety (Halperin et al., 2012). Yet again, this line of research points to the idea that open-minded interventions involve a complex interplay between cognitive, motivational, and affective processes.

Overall, helping individuals learn emotion regulation strategies may be an effective way to ensure that people with opposing views can be open to listening to one another and engaging in active dialogue. Further, guiding the appraisals that people make about concepts such as group malleability can help guide the emotions that they experience toward members of opposing groups. Given that people experience a diverse range of emotions in the context of intergroup relations (and many appraisal dimensions), researchers might explore the downstream effects of shifting other appraisal dimensions, such as intentionality and agency. In addition, since much of the work on improving emotion regulation has been conducted in the specific context of the intractable conflict between Israel and Palestine, it would be useful to conduct studies that replicate these findings in a variety of contexts (e.g., between liberals and conservatives in the United States, between students and professors, between acquaintances, etc.) to ensure their generalizability.

Encouraging Emodiversity

Recent research has proposed that it is not always necessary to focus on downregulating emotion in order to promote open-mindedness. Across five studies, Grossman et al. (2020) found a robust relationship between people's tendency to recognize and experience a range of emotions — emodiversity — and a tendency to engage in wise reasoning. In these studies, they measured emodiversity using sentiment analysis of interview transcripts (Study 1) and a formula developed by Quoidbach (2014) that incorporates multiple self-report items (Studies 2-5). They found that participants with higher emodiversity also reported having higher intellectual humility, were more likely to consider diverse perspectives, were more likely to adopt a distant (rather than immersed) viewpoint, and were more likely to search for compromise. Similarly, participants with more emodiversity scored higher on the situated wise reasoning scale, a measure of state-level wise reasoning.

Although there is compelling evidence for there being a positive relationship between emodiversity and wise reasoning, less work has been done to develop interventions to promote emodiversity. Grossman et al. attempted to do so in three of their studies (Studies 4a-c) using a few different methods. They had participants attempt to appraise their emotions in a differentiated versus simple (good/bad) manner and they had participants focus on multiple emotions that they had experienced versus one strong emotion. However, they did not find that any of these approaches worked to increase emodiversity. Therefore, further work will be needed to help boost emodiversity and to examine the causal effect between emodiversity and wise reasoning.

Sustaining Open-Mindedness Through Social Skills

In previous sections, we discussed how interventions that target cognitive and motivational pathways can induce people to be open-minded in the first place. In addition, training on emotion regulation strategies can help to sustain open-mindedness in social situations. Now, we will discuss the impact of equipping people with the proper social tools that they need to successfully navigate difficult conversations. Even if individuals start off being receptive, if their interaction with one another does not go well, they can quickly move to become closed-minded again. Thus, further training in communication skills may help ensure that individuals remain open-minded during social interaction. Although there are myriad studies on how to improve communication generally, this section maintains a more specific focus on best practices for improving communication between individuals with divergent views.

Building Rapport

In research on communication and negotiation, building rapport is often seen as a key component for maintaining a positive environment and generating mutually beneficial outcomes. For a comprehensive review on the relationship between rapport and conflict outcomes, see Nadler (2003). A simple method for building rapport between strangers is having them engage face-to-face. Drolet and Morris (1999) found that participants who engaged face-to-face as opposed to side-by-side achieved higher joint gains during a negotiation. In a second study, they found that even when participants were separated during a conflict game, if they met face-to-face first, they were more likely to prioritize joint gains.

These effects were mediated by increased rapport between those who interacted face-to-face, which was assessed both via self-report and coding of video footage of the interactions. In this research, rapport was defined as “a state of mutual positivity and interest that arises through the convergence of nonverbal expressive behavior in an interaction (see Bernieri, 1988, and Tickle-Degnen & Rosenthal, 1990).” Other research has found that when face-to-face contact is not possible, even engaging in “small talk” prior to an interaction can serve a similar rapport-building function. In a study of participants who negotiated over email, those who had the opportunity to chat over the phone for five minutes prior to negotiations reported feeling greater rapport, and had more successful negotiation outcomes (Morris, Nadler, Kurtzberg, & Thompson, 2002). To extend these findings, it would be useful for researchers to test whether “get to know you” exercises can be conducted via typing alone in a more anonymous context, to see if these interventions can have similar efficacy. Other methods aside from engaging face-to-face or engaging in small talk have also been developed for building rapport. For instance, engaging in behavioral mimicry has been shown to produce mutually beneficial negotiation outcomes (Maddux, Mullen, & Galinsky, 2008). Another technique for building rapport between individuals is through self-disclosure, which has been shown to increase liking (Collins & Miller, 1994).

Although it may be useful to integrate these techniques in building rapport between individuals with opposing views, careful consideration is required. For instance, interactions between negotiators may seem conceptually similar to interactions between people with opposing views, but there are important conceptual differences as well. In the negotiation literature, participants are usually assigned roles and put into hypothetical scenarios. These individuals do not face potential threats to their own beliefs in the same way that participants with opposing viewpoints might. Individuals who engage with others who hold opposing views might be more reticent to self-disclose for fear of making themselves vulnerable to attack from the other side.

Thus, more research is needed on how rapport-building exercises affect individuals in this specific type of social context. Recent work has shown that compared to typing, engaging face-to-face or via video chat facilitates higher perceptions of humanization, greater conversation responsiveness, and lower conversation conflict (Schroeder, in prep; Lieberman & Schroeder, 2019). Given that much discord between people with opposing views happens between anonymous strangers online, it will also be important to develop interventions that can build rapport without the benefit of face-to-face interaction.

Perspective-getting

To ensure effective communication, it is important that individuals accurately understand one another's point of view to avoid "talking past one another," or having misperceptions about the other's view as being threatening. As was mentioned briefly in the discussion above on perspective-taking, taking on another person's perspective does not ensure accurate understanding of that person's viewpoint (Eyal et al., 2018). One method for improving the exchange of accurate information is perspective-getting, which involves directly asking another person to explain their perspective. For instance, Eyal and colleagues (2018) found that in contrast to participants who engaged in perspective-*taking*, romantic partners who engaged in perspective-*getting* were more accurate at understanding one another's views. In their study, the researchers instructed participants to ask their partner to provide their opinion on a series of specific statements, and then had participants predict their partner's responses to those statements on a 7-point scale.

Outside of the lab, field studies that involve political canvassing suggest that perspective-getting can be effective on a large scale at reducing prejudice toward out-group members. Kalla and Broockman (under review) conducted multiple field studies in which they employed different narrative strategies while engaging in conversations during door-to-door canvassing (see also Broockman & Kalla, 2016; Kalla & Broockman, 2020). In some of these studies, they employed three strategies all at once: analogic perspective-taking, perspective-giving, and perspective-getting. In others, they included only one at a time. They found that an intervention that only employed perspective-getting had an equivalent effect size to ones that paired perspective-getting with perspective-giving and to studies that used all three techniques. In an experimental study, they also found that a perspective-getting exercise had the strongest impact on reducing prejudiced attitudes toward immigrants and transgender people. Based on these findings, they concluded that perspective-getting was the core component that made their earlier canvassing intervention effective.

Other studies have found that encouraging individuals to ask questions and to be generally curious can also be effective at facilitating positive attitudes between communication partners. First, question-asking can improve a speaker's impression of a listener. Huang et al., (2017) found that when participants were instructed to ask at least 9 questions (versus "at most 4"), they were perceived as being more responsive and were better liked by their conversation partners. Applying this to the domain of communication between individuals with opposing viewpoints, Chen, Minson, and Tormala (2010) asked participants to engage with a purported debate partner over chat. They found that participants who received a question from their debate partner (e.g., “But I was interested in what you’re saying. Can you tell me more about how come you think that?”) rated their partner and themselves as being more receptive. Naïve raters also judged participants in the “question” condition to behave more receptive toward their partner in their responses to their partner’s message. Studies have also found that high quality listening (which researchers define as listening that is “empathic, attentive, and nonjudgmental”) can reduce speakers’ social anxiety, improve their self-awareness, reduce defensive processing, and reduce the extremity of their attitudes (Itzchakov et al., 2018; Itzchakov et al., 2017).

In addition to influencing speakers’ impressions of listeners, asking questions can also affect listeners’ impressions of speakers. In a second study, Chen et al. (2010) found participants who were asked to generate questions in response to a message containing an opposing viewpoint reported being more favorable toward and more willing to engage with people who held that viewpoint. Thus, even the process of coming up with follow-up questions can be beneficial in maintaining a positive interaction. Importantly, the researchers emphasize that when individuals ask questions, these should be *elaboration* questions. The researchers state that elaboration questions are “not asked in order to couch an argument in question form, nor to trap the other party into making a contradictory statement, but rather to gain greater understanding of the other’s views.” In their study, they guided participants to ask questions in this manner by telling them to ‘‘come up with three open-ended questions for the speaker that will help you better understand why he feels as he does.” Therefore, it is important to note that question-asking can be effective, but only when the right kinds of questions are asked. Furthermore, it will be important to extend these findings by observing the effect of question-asking in true interactive contexts. Overall, question-asking (or perspective-getting) seems to be a useful tool for both facilitating the accurate exchange of information and maintaining positive relations between interaction partners, such that defensive responding is less likely to occur.

Perspective-giving

In concert with perspective-*getting* on the listener's side*,* the other important process required for promoting effective information exchange and maintaining positive attitudes is perspective-*giving* on the speaker's side. Perspective-giving occurs when a speaker shares their views and feels heard and understood by the listener. Researchers also refer to this as narrative exchange, particularly with regards to deep canvassing (Kalla & Broockman, 2020). In two studies, Bruneau and Saxe (2012) asked members of non-dominant groups (Mexican immigrants and Palestinians) to engage with members of dominant groups (White Americans and Israelis) through text chat and video. They found that perspective-giving led non-dominant group members to express more positive attitudes toward their interaction partners, arguing that the exercise allowed these individuals to "feel heard." However, when non-dominant group members took the perspective of dominant group members, their attitudes toward those individuals became more negative. In contrast, dominant group members benefited from the perspective-getting intervention in one of the studies, with a trending but non-significant effect in the second study; however, for these individuals, engaging perspective-taking was more effective at improving their attitudes.

In these studies, perspective-taking was operationalized as being assigned to accurately summarize a partner's viewpoint after reading about it. This implementation of perspective-taking differs from traditional paradigms, in that it motivates listeners to pay close attention, gives them an opportunity to express their understanding to make the speaker feel heard, and test their understanding of the other person's perspective. Thus, this form of perspective-taking might be more akin to perspective-getting, if there is opportunity for iterative interaction, in which the listener can parrot back the speaker’s argument until they “get it right.” Overall, these studies suggest that researchers carefully consider contextual factors that might introduce a power imbalance between dialogue partners, and tailor interventions accordingly.

More evidence for the effectiveness of perspective-giving, which is sometimes also referred to simply as disclosure, can be found in research on negotiations. Though negotiators might think it is always in their best interest to “hold their cards close to the vest,” full disclosure may be more beneficial. For instance, research conducted by Thompson (1991) found that negotiators who provided or sought information from their partner achieved better joint outcomes, and at no cost to their individual profit. However, in these studies, Thompson noted that it can be a big challenge to get negotiators to engage in this way. A slim percentage of negotiators were willing to seek or disclose information spontaneously. Informing participants that they might have different priorities also did not encourage spontaneous disclosure or information seeking. Furthermore, even some of the participants who were instructed to seek or disclose information refused to do so. This further highlights the challenge in developing interventions that can improve information seeking and disclosure behavior in the long-term, in cases where interaction partners cannot be prompted to do so.

In the context of negotiators with opposing political views, Keltner and Robinson (1993) found that opposing partisans who fully disclosed their views prior to negotiating with one another evaluated each other more positively and were more successful in their negotiations. In comparison, participants who partially disclosed their views were no better off than participants who disclosed no information. The researchers argue that full disclosure allows partners to become more aware of potential points of agreement and to eliminate perceptions of extreme ideological differences (i.e., reduce false polarization). Partial disclosure, on the other hand, increases suspicion that an interaction partner is 'hiding something.' Apart from improving the perceptions of the listener, disclosure might also confer benefits on the speaker, as disclosure has been found to be intrinsically rewarding for the speaker (Tamir & Mitchell), and to increase the listener's liking for the speaker (Collins & Miller, 1994). However, these findings have not been tested in the context of interactions between individuals who know they hold opposing views, and thus, further research on the effects of disclosure on the speaker in this context is warranted. Persuading individuals to disclose in this kind of scenario might also require activities that aim at building trust between interaction partners, such that they feel comfortable disclosing.

Framing Opinions with Receptive Language

Recent research has tested the effects of stating opinions using language that signals receptiveness (Table 1). Hussein and Tormala (2021) tested different kinds of phrases to examine whether they would impact people's ratings of how open-minded and receptive a speaker was. They found that readers perceived speakers to be more open-minded and more receptive when they used phrases that expressed uncertainty, acknowledged mistakes, or highlighted drawbacks. Yeomans et al. (2020) took a more data-driven, bottom-up approach to identify language that can signal open-mindedness. They developed a natural language processing algorithm to determine features that most clearly distinguished between receptive and non-receptive text, and then developed an intervention in which they taught participants to use the 'receptiveness recipe' that was identified by the algorithm. The algorithm identified the following features as signaling receptiveness: using positive statements (rather than negations), acknowledging understanding of the other person's view, using hedges to soften claims, and identifying points of agreement.

Receptiveness technique	Example phrase	Source
Expressing uncertainty	"I cannot be entirely sure, but I believe that…"	Hussein & Tormala, 2021
Acknowledging mistakes	"I used to think X, but I was wrong."	Hussein & Tormala, 2021
Highlighting drawbacks	"One of the disadvantages … is that …"	Hussein & Tormala, 2021
Positive statements rather than negations	"X is true" or "X is good" rather than "Y is not true"	Yeomans et al., 2020
Acknowledging understanding	"I see your point" or "I understand where you are coming from"	Yeomans et al., 2020
Using hedges	"X is partly true…" or "Y is sometimes the case"	Yeomans et al., 2020
Find points of agreement	"I agree that it's a difficult situation, which is why X," rather than "That doesn't work because Y"	Yeomans et al., 2020

Table 1. Examples of language that can be used to signal receptiveness.

Using "I"-statements

A large body of research on close relationships has demonstrated significant differences in how language affects conflict between friends and romantic partners. Studies have found that using 'I-statements' (e.g., "I feel disappointed") versus more accusatory "you-statements" (e.g., "You disappointed me") promotes more positive feelings and more productive interactions (Simmons, Gordon, & Chambless, 2005; Kubany, Richard, Bauer, & Muraoka, 1992). Thus, I-language is less threatening, and less likely to produce negative affective responses that can impede individuals from being able to continue being receptive to an interaction partner. Research has also examined how I-language might be effective at improving communication between strangers. Rogers et al. (2018) asked participants to rate a series of statements that might be used at the beginning of a conflict discussion in terms of the likelihood that they would produce a defensive reaction. They found that statements using I-language, in combination with language that attempted to address both the speaker and listener's perspective, were perceived as being the most likely to ensure a positive interaction. However, this study employed only hypothetical scenarios, and thus further testing in a truly social context is warranted. Furthermore, the receptiveness algorithm that was developed by Yeomans et al. (2021) also found that messages that used more first-person language were more likely to be rated as receptive. Overall, I-statements seem to result in less defensive responding, which helps maintain emotional receptivity between individuals during communication.

Moral Reframing

Another technique that has been tested as a way to help people with different viewpoints communicate effectively is called moral reframing (Feinberg & Willer, 2019). With this technique, people reframe their arguments about ideological issues by reframing them in a way that speaks to the other person's values. For instance, research has shown that American conservatives tend to place high value on loyalty, sanctity, and authority, while liberals value fairness and care (Graham, Haidt, & Nosek, 2009). As such, studies have shown that conservatives become more supportive of pro-environmental policies when they are presented with arguments that suggest that it is their patriotic duty to protect the environment and that the environment is dirty and needs to be purified (Feygina et al., 2010; Feinberg & Willer, 2013). Similarly, liberals are more likely to support military spending when presented with arguments that the military can help to reduce income inequality and racial discrimination (Feinberg and Willer, 2015). Although this technique has been studied more in terms of its ability to promote persuasion, we think that it could also be a useful tool for communicating across political divides.

Interventions that Target Multiple Pathways

While attempting to categorize interventions based upon the primary pathway that they targeted, we had particular difficulty with regards to two popular interventions: mindfulness training and intergroup contact. These interventions can be considered "kitchen-sink" approaches — they often involve multiple components and, as a result, affect multiple pathways. Because of this, we decided to break out these two interventions into their own "spotlight" section. Large bodies of literature have been dedicated to understanding their efficacy. Therefore, we will give a broad overview of the ways in which we believe they impact open-mindedness.

Mindfulness Training

In recent years, there has been a large increase in studies testing the efficacy of mindfulness (for a recent review, see Creswell, 2017). Mindfulness has been operationalized and measured in many different ways (Quaglia et al., 2015). A common definition, provided by Kabat-Zinn (1994), refers to mindfulness as "paying attention in a particular way: on purpose, in the present moment, and non-judgmentally." Most researchers consider mindfulness as aiming to cultivate two primary outcomes: (1) increased present-moment attention and awareness and (2) an open, accepting, and non-judgmental attitude.

Certain mindfulness meditation practices focus more specifically on one of these sub-factors. Some mindfulness programs focus most on training attention to the present moment, whereas others focus more on boosting socio-cognitive or compassion-based skills. As such, these different kinds of programs can have different impacts on open-mindedness. For example, a large-scale study called the ReSource Project examined the effects of different types of mindfulness training modules (Hildebrandt, McCall, Singer, 2017; Bockler et al., 2018; Singer & Engort, 2019; Hildebrandt et al., 2019; Engert et al., 2017). The "Presence" module focused on improving attention to the present moment and bodily awareness. The "Affect" module used a "loving kindness meditation" to help cultivate gratitude, prosocial emotions, and the ability to deal with difficult emotions. The "Perspective" module focused on boosting meta-cognitive and perspective-taking skills. The researchers found that the Presence module was effective at improving participants' self-reported ability to observe, be present, and not react. However, it did not impact participants' ability to adopt an accepting and non-judgmental mindset. Instead, the Perspective module improved acceptance while the Affect module improved the ability to be non-judgmental and increased altruistically motivated social behavior. The Perspective and Affect modules also improved the same outcomes as were impacted by the Presence module, promoted the use of emotion regulation strategies, and attenuated the physiological stress response. With regards to open-mindedness, these findings suggest that mindfulness trainings that focus on teaching socio-cognitive skills and/or compassion-based practices can improve people's ability to process information in an open and unbiased way, to be accepting of others who may hold alternative viewpoints, and to have adaptive emotional responses in reaction to alternative viewpoints. These trainings appear to impact all of the psychological mechanisms included in our conceptual model, including cognitive, affective, motivational, and social processes.

In the most direct test of the effects of mindfulness on receptivity to those with opposing views, Alkoby et al. (2019) investigated the efficacy of a general-purpose, well-established eight-week mindfulness program called "mindfulness-based stress reduction" (Kabat-Zinn, 1990). Israeli Jews were assigned to the mindfulness condition or to a control (i.e., no intervention). At the conclusion of the mindfulness program, a subset of participants from each condition also received training in cognitive reappraisal. All participants then watched a video in which an Israeli-Palestinian politician gave a "harsh speech against the Israeli government's actions." The researchers found that, compared to the control condition, all three experimental conditions (mindfulness, reappraisal, and their combination) effectively reduced negative emotional responding, reduced perceived threat, and increased support for compromise.

Overall, testing the efficacy of mindfulness interventions with regard to relations between individuals with opposing views is new territory. In addition to testing its efficacy in this context generally, it would be useful to test the efficacy of mindfulness training programs of different lengths to determine proper “dosage.” Is it necessary to conduct a several-session training course, or can a brief intervention suffice? In addition, it will be important to understand the long-term effects of open-mindedness interventions. Finally, many studies on mindfulness have been conducted in populations who are motivated to use the treatment to improve their own well-being. This means that intervention samples in open-mindedness studies are often self-selected. In order to develop interventions that can be used and accepted widely, it may be useful to develop mindfulness interventions that consider how to best serve individuals who might be resistant to them.

Intergroup Contact

Decades of research have also examined the beneficial effects of intergroup contact on reducing prejudice (Allport, 1956). For theoretical and meta-analytic reviews of the effects of intergroup contact, see Pettigrew (1998); Pettigrew & Tropp (2006); and Pettigrew et al. (2011). Studies of contact theory tend to focus on reducing prejudice between members of racial and ethnic groups; however, we propose that intergroup contact operates on several mechanisms related to open-mindedness, including affective and cognitive. In particular, Dovidio, Gaertner, and Kawakami (2003) suggest that intergroup contact is effective because it reduces anxiety and alters social categorizations.

Some researchers have argued that intergroup contact can operate as a sort of ‘exposure therapy’ in reducing people’s negative affect in response to out-group members

(Birtel & Crisp, 2012). In a recent meta-analytic review of 45 studies with 60 independent samples, Pettigrew and Tropp (2016) found that the association between intergroup contact and prejudice reduction was mediated by reductions in threat and anxiety responses during interactions between members from opposing groups. In one study included in the review, Blascovich et al. (2001) found that Whites who had more contact with members of other racial and ethnic groups showed lower physiological markers of stress and reported lower levels of anxiety during an interaction with an out-group member compared to those who had less contact. Pettigrew and Tropp also found that perspective-taking/empathy was a significant mediator. Other studies have found that even imagined contact can improve people's attributions and emotions about stigmatized groups and people with opposing views (Warner & Villamil, 2017; Birtel & Crisp, 2012).

Intergroup contact can also operate on cognitive pathways by leading to the formation of new group identities. In particular, it can help to personalize members of the out-group and also correct inaccurate meta-perceptions about what those out-group members are really like. According to social identity theory, people naturally group concepts into categories, which underlies people's tendency to categorize people into an in-group versus an out-group (Turner et al., 1997; Stets & Burke, 2000). Researchers who advocate for the common in-group identity model suggest that changing how people make social categorizations can be effective at reducing intergroup bias (Gaertner & Dovidio, 2011).

For example, Gaertner et al. (2000) reanalyzed data from the classic social psychology experimenter called the Robbers Cave, in which boys at a summer camp formed two social groups that were in conflict who then improved their relationships with one another through intergroup contact (Sherif, 1961). Based on their analysis and other experimental findings from their laboratory, they proposed that Sherif was effective at reducing intergroup conflict between the two groups because of strategies he employed that led to “decategorization, recategorization, and mutual differentiation processes.” Decategorization (or personalization) involves seeing a member of an out-group as an individual rather than a group member. Recategorization involves focusing on a superordinate identity, or some other shared group membership, such that the out-group member is reclassified as an in-group member. And mutual differentiation involves having group members emphasize their group differences as a benefit to their mutual collaboration. Therefore, intergroup contact can help to either lead to depersonalization or facilitate the creation of a shared identity/superordinate goals, both of which can help to reduce prejudice.

It is important to note that there are boundary conditions on the effects of intergroup contact. For instance, some research has also found that intergroup contact can promote rather than alleviate anxiety (Shelton, 2003). Instances of negative contact can have adverse effects through making group membership more salient (Paolini, Harwood, & Rubin, 2010). Since the advent of contact theory, researchers have noted that certain conditions are required in order for contact theory to be successful, including equal status, common goals, cooperative interdependence, support from norms and/or authorities, opportunity for personal acquaintance, and the development of intergroup friendships (Allport, 1954; Dovidio, Gaertner, and Kawakami, 2003). Therefore, intergroup contact must be administered carefully. Further research is needed to determine how effective it is in contexts involving people with different ideological viewpoints, and whether it requires additional prerequisites in order to be successful.

Comprehensive Dialogue Training

The majority of interventions in this section focus on testing the effects of explicit instructions directly prior to interaction. However, few of them incorporate repeated interactions between individuals with opposing views, which may serve as a training ground to improving interactions in the long term. Research on the efficacy of practicing dialogue is limited, though in recent years, psychologists have been moving toward developing such interventions. For instance, Influs et al. (2019) developed the intervention "Tools of Dialogue" to reduce tension between Israeli and Palestinian adolescents. In this intervention, Israelis and Palestinians engaged in an 8-session series. Each session contained an introduction to a specific topic (e.g., empathy, prejudice…), activities and games that would promote synchronous behavior, and opportunities for one-on-one and group dialogue. The researchers obtained a large battery of pre- and post-intervention measures, including recorded dialogues, saliva samples, and individual interviews. Thus far, results show that the intervention increased perspective-taking, which was operationalized as the extent to which participants perceived that the "conflict is complicated and that there is justice on both sides of the conflict." Physiological findings were mixed. Intervention participants showed decreases in cortisol production at post-test versus pre-test, in contrast to control participants who showed no change. (Influs et al., 2019). Furthermore, this effect was mediated by increases in perspective-taking. With regards to oxytocin, there was no main effect on the intervention; however, there was an interaction effect, as those who were high in perspective-taking at pre-test and then went through the intervention showed increases in oxytocin levels (Influs et al., 2019). Overall, this intervention provided opportunities for participants to learn effective communication techniques and to practice them in a guided setting. However, the intervention program also involved repeated exposure to members of the opposing group, learning about alternative viewpoints, learning about cognitive biases, and participating in synchronous activity. Thus, it becomes very difficult to tease apart mechanisms, and to determine whether or not the dialogue component was effective at improving perspective-taking and stress responding. Future work is needed to determine whether training in communication skills and repeated, supervised practice can be effective at improving interactions between individuals with opposing views.

Discussion

In this review, we have attempted to provide a comprehensive overview highlighting interventions that can be used to promote and sustain open-mindedness. We outlined a conceptual model that can be used to understand the different psychological pathways on which these interventions operate. We organized these interventions according to the psychological pathway on which they had the most direct impact. Where possible, we included studies that tested the effects of interventions directly on improving attitudes and/or relations between individuals who hold opposing views; however, many interventions have not yet been tested directly in this domain.

Summary of Current Evidence

In reviewing a broad range of literature across multiple academic fields, we identified four main psychological pathways that open-mindedness interventions can target. First, we reviewed interventions that have aimed to alter cognitive processes using either direct or more domain-general methods. Second, we identified research programs that have induced open-mindedness through motivational pathways, whereby the goal is to promote accuracy goals, satisfy the need for self-integrity, or leverage the need to belong. Third, we discussed how emotion regulation training can help individuals remain open-minded when they encounter viewpoints that may give rise to negative emotions during social interactions. Finally, we explored how interventions that teach social skills can also help to maintain the open-mindedness of an individual and their interaction partner.

Overall, research on the efficacy of open-mindedness interventions is still nascent. Further evidence is required to determine which interventions are most effective on their own and in combination with one another. To our knowledge, this narrative review is the first attempt to consolidate research on open-mindedness interventions across multiple fields. In Table 2, we provide a list of the intervention types included in this review. In addition, for each intervention type, we include qualitative descriptions that indicate whether evidence has been found to support its efficacy and how much effort is required to administer it.

Intervention technique	Intervention type	Evidence Supporting Efficacy	Required effort to administer
Teaching about biases	Cognitive (targeted)	Yes - with boundary conditions	Low or High
Changing implicit theories/mindsets (e.g. growth mindset)	Cognitive (targeted)	Yes - with boundary conditions	Low
Perspective-taking	Cognitive (targeted)	Yes - with boundary conditions	Low
Paradoxical thinking	Cognitive (targeted)	Yes - with boundary conditions	Low
Puncturing the illusion of explanatory depth	Cognitive (targeted)	Mixed - some failures to replicate	Low
Correcting false meta-perceptions	Cognitive (targeted)	Yes	Low
Cognitive disfluency	Cognitive (domain-general)	Yes	Low
Self-distancing	Cognitive (domain-general)	Yes	Low
Priming creativity	Cognitive (domain-general)	NA - no research to date	Low

Positive mood inductions	Cognitive (domain-general)	Yes - with boundary conditions	Low
Cognitive training	Cognitive (domain-general)	Mixed - little research to date	High
Promoting accuracy motives	Motivational	Yes	Low
Satisfying the Need for Self-Integrity (Self-Affirmation)	Motivational	Mixed - requires boundary conditions, some failures to replicate	Low
Leveraging the Need to Belong	Motivational	Yes	Low
Cognitive reappraisal	Emotion regulation	Yes	Low or High
Indirect emotion regulation	Emotion regulation	Little research to date - see “Changing implicit theories/mindsets”	Low
Encouraging emodiversity	Emotion regulation	NA - no research to date	More evidence needed
Building rapport	Social skills	Yes	Low
Perspective-getting	Social skills	Yes	Low
Perspective-giving	Social skills	Yes	Low
Framing opinions with receptive language	Social skills	Yes	Low
Using “I-statements”	Social skills	Yes	Low
Moral reframing	Social skills	Yes	Low
Mindfulness training	Multiple	Yes - specific to certain subtypes of mindfulness	High
Intergroup contact	Multiple	Yes	High
Comprehensive dialogue training	Multiple	Yes	High

Table 2. List of intervention techniques included in this review along with subjective ratings that describe whether or not evidence supports each technique and the effort required to implement each technique.

This table can be used as a reference by researchers and interventionists to understand the current state of the literature and prompt them to consider whether an intervention might work in a certain context. For instance, studies have found that some interventions are effective only when certain boundary conditions, such as personality characteristics or social contexts, are present. Others have failed to replicate in recent work. Furthermore, some interventions appear to have potential based on existing correlational data, but have yet to be tested. Although this table can provide preliminary guidance, it does not contain information regarding effect sizes (i.e. the impact that can be achieved by the intervention) or the longevity of the intervention's effect. In the future, we hope that further research will enable the creation of an all-inclusive 'menu of options' from which researchers and practitioners can select the most appropriate interventions.

Currently, there are barriers that make it difficult to create such a comprehensive list with clear recommendations. First, more research is needed in order to better understand the efficacy of each intervention type. Some interventions are backed by large literatures that support their efficacy in general (e.g. changing implicit theories/mindsets, self-affirmation, cognitive reappraisal, and mindfulness), but have less evidence to support their ability to promote open-mindedness specifically. Other interventions have been developed more recently and tested in few studies, if any (e.g. cognitive training and encouraging emodiversity).

Second, the measures used to assess open-mindedness vary widely across studies. Many studies assess open-mindedness using non-validated measures that are tailored to a particular context. Although the creation of such 'bespoke' measures allows for specificity with regards to particular social and political issues or groups, it also poses challenges for comparing across studies and compiling effect sizes. Thus, it may be beneficial for researchers to develop a more standardized set of measures, and also to refer to the construct of open-minded thinking and behavior using more consistent terminology. As the literature grows and uses a more consistent set of validated measures, it will become possible to conduct meta-analyses that can provide further insight into which interventions have the strongest effects.

Future Directions

Given the nascency of research on open-mindedness interventions, there are many opportunities for future research. First, researchers can consider developing novel techniques for measuring open-mindedness. Most open-mindedness studies rely on self-report, which can be biased by demand characteristics, perceived social desirability, and a lack of introspective ability. Measurement error poses a problem for comparing effect sizes between intervention types and also for concluding that certain interventions that yield null effects are truly ineffective. Thus, further work is needed with regards to developing open-mindedness measures that are both precise and accurate. Recent research in this domain has been promising. For instance, studies have shown that portable neuroimaging techniques such as fNIRS and EEG can be used to study individuals interacting within dyads and even larger groups (Dikker et al., 2017; Burns, 2020). Technological advances in online text-based and video chatting have also made it more feasible to bring people with different viewpoints together to have conversations for research purposes (Binnquist, Dolbier, Dieffenbach, & Lieberman, under review). In tandem, researchers have developed more sophisticated yet accessible techniques for analyzing rich conversational datasets, including natural language processing (NLP) models and tools to analyze facial expressions (Yeomans et al., 2020; Cheong et al., 2021). Therefore, researchers now have the tools to measure the impact of open-mindedness interventions on behavior and the brain, which can help to address the limitations of self-report, especially in this particular field.

Simultaneously, given the large number of bridge-building associations being built in the United States and beyond, there are many opportunities for researchers to team up with practitioners to better understand the impact of their interventions on real-world outcomes. In their review of prejudice reduction interventions, Paluck et al. (2020) discuss the benefits and practicality of conducting field studies. They suggest working with partners who are already conducting interventions, as this allows researchers to test ideas that have already been feasibly implemented in an applied setting. Furthermore, they recommend designing intervention-based research by optimizing for settings that allow researchers to measure certain behavioral or real-world outcomes. Researchers have also found creative methods for measuring the broader impact of being open-minded. For instance, Minson et al. (2018) found that when two people are both receptive to one another's views, the social networks that they belong to become less homogenous. Therefore, researchers can measure the real-world impact of their interventions through teaming up with partners and/or using sophisticated techniques like social network analysis.

Another recent trend in terms of the development of interventions has been examining what 'dosage' is required to create interventions that are maximally impactful and long-lasting, but also feasible to administer at scale. Researchers have found that even brief, light-touch interventions can have large, long-term effects (Yeager & Walton, 2011). They argue that these interventions work because they focus on making small changes to subjective meaning-making — "the working hypotheses people draw about themselves, other people, and social situations" — that can have transformational effects (Walton & Wilson, 2018). Furthermore, they argue that these changes are highly context-specific and can be especially powerful when conducted in contexts like schools or companies that can help to reinforce change. For this reason, Walton (2014) has coined the term 'wise interventions' to describe these light-touch techniques due to the fact that they are precise in terms of the psychological mechanisms that they target and that they are maximally impactful. On the other hand, it is also possible that interventions that apply more of a 'kitchen-sink approach', combining multiple interventions together, are more effective. Little research has directly compared the effects of interventions that are administered in combination versus alone. It may also be beneficial to explore the extent to which interventions require reinforcement through repeated 'boosters' over time.

In conclusion, we believe that there are many open questions remaining as to the individual mechanisms and group-level forces that cause people to be open- or closed-minded. We argue that there is a fertile field for research in terms of understanding what an open mind 'looks like,' how to measure it, how to induce it, and how to sustain it. We hope that this review can serve as a helpful starting point for researchers in both basic and applied settings to develop more impactful interventions. In the long-term, such interventions may be able to address the rising affective polarization that has been seen in America and around the world. While it is unrealistic to expect that people will come to agree on everything, the research reviewed in this paper suggests that it is possible for people to learn to embrace a diversity of viewpoints and respect those who disagree with them.

REFERENCES

Albarracín, D., & Shavitt, S. (2018). Attitudes and attitude change. *Annual Review of Psychology*, *69*, 299-327.

Alkoby, A., Pliskin, R., Halperin, E., & Levit-Binnun, N. (2019). An eight-week mindfulness-based stress reduction (MBSR) workshop increases regulatory choice flexibility. *Emotion, 19*(4), 593-604.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Baehr, J. (2011). The Structure of Open-Mindedness. *Canadian Journal of Philosophy, 41*:2, 191-213, DOI: 10.1353/cjp.2011.0010

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., ... & Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group?. *Journal of personality and social psychology*, *72*(1), 105.

Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology*, *1*(3), 311-320.

Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior*, *12*(2), 120–138.

Binnquist, A. L., Dolbier., S. Y., Dieffenbach, M.C., Lieberman, M. D. (under review). The Zoom solution: Promoting effective cross-ideological communication online.

Bishop, B. (2008)*. The Big Sort: Why the Clustering of Liked-Minded America is Tearing Us Apart*. Boston: Houghton-Mifflin.

Blankenship, K. M., Friedman, S. R., Dworkin, S., & Mantell, J. E. (2006). Structural interventions: concepts, challenges and opportunities for research. *Journal of Urban Health, 83*(1), 59-72.

Bolman, L. G., & Deal, T. E. (2017). *Reframing organizations.* San Francisco, CA: Jossey-Bass.

Bridging Map. (May, 2021). *Princeton University Bridging Divides Initiative. https://bridgingdivides.princeton.edu/bridging-map/map*

Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of 'perspective-giving' in the context of intergroup conflict. *Journal of Experimental Social Psychology*, *48*(4), 855–866. https://doi.org/10.1016/j.jesp.2012.02.017

Carnall, C. A. (2007). *Managing change in organizations.* Pearson Education.

Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously. A theory of socioemotional selectivity. *The American Psychologist*, *54*(3), 165–181.

Catapano, R., Tormala, Z. L., & Rucker, D. D. (2019). Perspective Taking and Self-Persuasion: Why "Putting Yourself in Their Shoes" Reduces Openness to Attitude Change. *Psychological science*, *30*(3), 424-435.

Chi, M. T. H. (1997). Creativity: Shifting across ontological categories flexibly. In *Creative thought: An investigation of conceptual structures and processes* (pp. 209–234). https://doi.org/10.1037/10227-009

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9).

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, *26*(9), 1151–1164. https://doi.org/10.1177/01461672002611011

Cohen-Chen, S., Halperin, E., Crisp, R. J., & Gross, J. J. (2014). Hope in the Middle East: Malleability beliefs, hope, and the willingness to compromise for peace. *Social Psychological and Personality Science*, *5*(1), 67-75.

Cohen, G. L., Sherman, D. K., Bastardi, A., Hsu, L., McGoey, M., & Ross, L. (2007). Bridging the partisan divide: Self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. *Journal of Personality and Social Psychology*, *93*(3), 415–430. https://doi.org/10.1037/0022-3514.93.3.415

Coleman, P. T., Vallacher, R. R., Nowak, A., & Bui-Wrzosinska, L. (2007). Intractable conflict as an attractor: A dynamical systems approach to conflict escalation and intractability. *American Behavioral Scientist*, *50*(11), 1454-1475.

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: a meta-analytic review. *Psychological Bulletin*, *116*(3), 457.

Correll, J., Spencer, S. J., & Zanna, M. P. (2004). An affirmed self and an open mind: Self-affirmation and sensitivity to argument strength. *Journal of Experimental Social Psychology*, *40*(3), 350–356. https://doi.org/10.1016/j.jesp.2003.07.001

Crawford, J. T., & Ruscio, J. (2021). Asking People to Explain Complex Policies Does Not Increase Political Moderation: Three Preregistered Failures to Closely Replicate Fernbach, Rogers, Fox, and Sloman's (2013) Findings. *Psychological Science*, *32*(4), 611-621

Creswell, J. D. (2017). Mindfulness Interventions. *Annual Review of Psychology*, *68*(1), 491–516. https://doi.org/10.1146/annurev-psych-042716-051139

Dale, R., Spivey, M. J., Brône, G., & Oben, B. (2018). Weaving oneself into others. *Eye-Tracking in Interaction. Studies on the Role of Eye Gaze in Dialogue*, 67-90.

Davis, M. H., Conklin, L., Smith, A., & Luce, C. (1996). Effect of perspective taking on the cognitive representation of persons: A merging of self and other. *Journal of Personality and Social Psychology*, *70*(4), 713–726. https://doi.org/10.1037/0022-3514.70.4.713

DeWall, C. N., Visser, P. S., & Levitan, L. C. (2006). Openness to attitude change as a function of temporal perspective. *Personality and Social Psychology Bulletin*, *32*(8), 1010-1023.

Dieffenbach, M. C., Gillespie, G. S., Burns, S. M., McCulloh, I. A., Ames, D. L., Dagher, M. M., ... & Lieberman, M. D. (2021). Neural reference groups: a synchrony-based classification approach for predicting attitudes using fNIRS. *Social Cognitive and Affective Neuroscience*, *16*(1-2), 117-128.

Dieffenbach, M. C., Burns, S., Li, J., Ames, D. L., Lieberman, M. D. (in prep). Leveraging Differences to Bring People Together: Using Neural Synchrony to Detect Polarized Thinking and Evaluate Open-Mindedness Interventions

Dovidio, J. F., Johnson, J. D., Gaertner, S. L., Pearson, A. R., Saguy, T., & Ashburn-Nardo, L. (2010). Empathy and intergroup relations. In M. Mikulincer & P. R. Shaver (Eds.), *Prosocial motives, emotions, and behavior: The better angels of our nature.* (pp. 393–408). https://doi.org/10.1037/12061-020

Dweck, C. (2012). Implicit theories. In P. A. Van Lange, A. W. Kruglanski, & E. T. Higgins *Handbook of theories of social psychology: volume 2* (Vol. 2, pp. 43-61). SAGE Publications Ltd, https://www.doi.org/10.4135/9781446249222.n28

Dweck, C. S. (2012b). Mindsets and human nature: Promoting change in the Middle East, the schoolyard, the racial divide, and willpower. *American Psychologist*, *67*(8), 614.

Epley, N., Caruso, E. M., & Bazerman, M. H. (2006). When perspective taking increases taking: reactive egoism in social interaction. *Journal of personality and social psychology*, *91*(5), 872.

Estrada, C. A., Isen, A. M., & Young, M. J. (1997). Positive Affect Facilitates Integration of Information and Decreases Anchoring in Reasoning among Physicians. *Organizational Behavior and Human Decision Processes*, *72*(1), 117–135. https://doi.org/10.1006/obhd.1997.2734

Eyal, T., Steffel, M., & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of Personality and Social Psychology*, *114*(4), 547.

Förster, J., Friedman, R. S., & Liberman, N. (2004). Temporal Construal Effects on Abstract and Concrete Thinking: Consequences for Insight and Creative Cognition. *Journal of Personality and Social Psychology*, *87*(2), 177–189. https://doi.org/10.1037/0022-3514.87.2.177

Fredrickson, B. L. (2001). The Role of Positive Emotions in Positive Psychology. *The American Psychologist*, *56*(3), 218–226.

Fredrickson, B. L. (2004). The broaden–and–build theory of positive emotions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1449), 1367-1377.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion*, *19*(3), 313–332. https://doi.org/10.1080/02699930441000238

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*(4), 708–724. https://doi.org/10.1037/0022-3514.78.4.708

Galinsky, A. D., Ku, G., & Wang, C. S. (2005). Perspective-taking and self-other overlap: Fostering social bonds and facilitating social coordination. *Group Processes & Intergroup Relations*, *8*(2), 109-124.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in Experimental Social Psychology* (Vol. 47, pp. 55-130). Academic Press.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, *96*(5), 1029.

Grossmann, I., Weststrate, N. M., Ardelt, M., Brienza, J. P., Dong, M., Ferrari, M., ... & Vervaeke, J. (2020). The science of wisdom in a polarized world: Knowns and unknowns. *Psychological Inquiry*, *31*(2), 103-133.

Halperin, E. (2014). Emotion, Emotion Regulation, and Conflict Resolution. *Emotion Review*, *6*(1), 68–76. https://doi.org/10.1177/1754073913491844

Halperin, E., Crisp, R. J., Husnu, S., Trzesniewski, K. H., Dweck, C. S., & Gross, J. J. (2012). Promoting intergroup contact by changing beliefs: Group malleability, intergroup anxiety, and contact motivation. *Emotion*, *12*(6), 1192.

Halperin, E., Cohen-Chen, S., & Goldenberg, A. (2014). Indirect emotion regulation in intractable conflicts: A new approach to conflict resolution. *European Review of Social Psychology*, *25*(1), 1-31.

Halperin, E., & Gross, J. J. (2011). Emotion regulation in violent conflict: reappraisal, hope, and support for humanitarian aid to the opponent in wartime. *Cognition & Emotion*, *25*(7), 1228–1236. https://doi.org/10.1080/02699931.2010.536081

Halperin, E., & Tagar, M. R. (2017). Emotions in conflicts: understanding emotional processes sheds light on the nature and potential resolution of intractable conflicts. *Current Opinion in Psychology*, *17*, 94–98. https://doi.org/10.1016/j.copsyc.2017.06.017

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 and Social Psychology Bulletin*, *44*(1), 122-139.

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), 266-283.

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.

Hare, William. (2001). Bertrand Russell and the ideal of critical receptiveness. *Skeptical Inquirer, 25*, 3: 40- 44.

Hare, W. (2009). What open-mindedness requires. *Skeptical Inquirer*, *33*(2), 36-39.

Hasson, Y., Amir, E., Sobol-Sarag, D., Tamir, M., & Halperin, E. (July, 2020). Believing empathy is unlimited: Using performance art to reconstruct intergroup empathy and bring people together. The annual meeting of the International Society of Political Psychology (ISPP), Warsaw, Poland.

Hernandez, I., & Preston, J. L. (2013). Disfluency disrupts the confirmation bias. *Journal of Experimental Social Psychology*, *49*(1), 178-182.

Hernández-Mogollon, R., Cepeda-Carrión, G., Cegarra-Navarro, J. G., & Leal-Millán, A. (2010). The role of cultural barriers in the relationship between open-mindedness and organizational innovation. *Journal of Organizational Change Management, 23*(4), 360-376.

Hausmann, L. R., Levine, J. M., & Tory Higgins, E. (2008). Communication and group perception: Extending the ‘saying is believing' effect. *Group Processes & Intergroup Relations*, *11*(4), 539-554.

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PloS one*, *13*(10), e0204494.

Hudley, C., & Graham, S. (1993). An attributional intervention to reduce peer-directed aggression among African-American boys. *Child Development*, *64*(1), 124-138.

Influs, M., Pratt, M., Masalha, S., Zagoory-Sharon, O., & Feldman, R. (2019). A social neuroscience approach to conflict resolution: Dialogue intervention to Israeli and Palestinian youth impacts oxytocin and empathy. *Social Neuroscience*, *14*(4), 378–389. https://doi.org/10.1080/17470919.2018.1479983

Isen, A. M., Johnson, M. M., Mertz, E., & Robinson, G. F. (1985). The influence of positive affect on the unusualness of word associations. *Journal of Personality and Social Psychology*, *48*(6), 1413–1426.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*, 129-146.

Jetten, J., Spears, R., & Postmes, T. (2004). Intergroup Distinctiveness and Differentiation: A Meta-Analytic Integration. *Journal of Personality and Social Psychology*, *86*(6), 862–879. https://doi.org/10.1037/0022-3514.86.6.862

Johnson, D. R., Murphy, M. P., & Messer, R. M. (2016). Reflecting on explanatory ability: A mechanism for detecting gaps in causal knowledge. *Journal of Experimental Psychology: General*, *145*(5), 573–588. https://doi.org/10.1037/xge0000161

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.

Kross, E., & Grossmann, I. (2012). Boosting wisdom: Distance from the self enhances wise reasoning, attitudes, and behavior. *Journal of Experimental Psychology: General*, *141*(1), 43.

Lambie, J. (2014). *How to be critically open-minded: A psychological and historical analysis.* Springer.

Lewin, K. (1946). *Resolving social conflicts and field theory in social science.* Washington, D.C.: American Psychological Association.

Levy, A., & Maaravi, Y. (2018). The boomerang effect of psychological interventions. *Social Influence*, *13*(1), 39-51.

Jia, L., Hirt, E. R., & Karpen, S. C. (2009). Lessons from a faraway land: The effect of spatial distance on creative cognition. *Journal of Experimental Social Psychology*, *45*(5), 1127-1131.

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour 4*, 279–286 https://doi.org/10.1038/s41562-019-0766-4

Light, N., & Fernbach, P. (2020). The role of knowledge calibration in intellectual humility. In *The Routledge Handbook of Philosophy of Humility* (pp. 411-424). Routledge.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare?. *Perspectives on Psychological Science*, *4*(4), 390-398.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

Lyons, B., Farhart, C., Hall, M., Kotcher, J., Levendusky, M., Miller, J., . . . Zhao, X. (2021). Self-Affirmation and Identity-Driven Political Behavior. *Journal of Experimental Political Science,* 1-16. doi:10.1017/XPS.2020.46

Maddux, W. W., Mullen, E., & Galinsky, A. D. (2008). Chameleons bake bigger pies and take bigger pieces: Strategic behavioral mimicry facilitates negotiation outcomes. *Journal of Experimental Social Psychology*, *44*(2), 461–468.

McCrae, R. R., & Costa, P. T., Jr. (1997). Conceptions and correlates of openness to experience. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of Personality Psychology* (p. 825–847). Academic Press. https://doi.org/10.1016/B978-012134645-4/50032-9

McDonald, M., Porat, R., Yarkoney, A., Reifen Tagar, M., Kimel, S., Saguy, T., & Halperin, E. (2017). Intergroup emotional similarity reduces dehumanization and promotes conciliatory attitudes in prolonged conflict. *Group Processes & Intergroup Relations*, *20*(1), 125–136. https://doi.org/10.1177/1368430215595107

Minson, J., Chen, F., & Tinsley, C. H. (September, 2018), Why Won’t You Listen to Me? Measuring Receptiveness to Opposing Views. HKS Working Paper No. RWP18-028;

Georgetown McDonough School of Business Research Paper No. 3295946. Available at SSRN: https://ssrn.com/abstract=3295946

Mitchell, R. L. C., & Phillips, L. H. (2007). The psychological, neurochemical and functional neuroanatomical mediators of the effects of positive and negative mood on executive functions. *Neuropsychologia*, *45*(4), 617–629. https://doi.org/10.1016/j.neuropsychologia.2006.06.030

Monroe, B. M., & Read, S. J. (2008). A general connectionist model of attitude structure and change: The ACS (Attitudes as Constraint Satisfaction) model. *Psychological Review, 115*(3), 733–759. https://doi.org/10.1037/0033-295X.115.3.733

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129-140.

Murray, N., Sujan, H., Hirt, E. R., & Sujan, M. (1990). The influence of mood on categorization: A cognitive flexibility interpretation. *Journal of Personality and Social Psychology*, *59*(3), 411–425. https://doi.org/10.1037/0022-3514.59.3.411

Nadler, J. (2003). Rapport in negotiation and conflict resolution. *Marq. L. Rev.*, *87*, 875.

Nasie, M., Bar-Tal, D., Pliskin, R., Nahhas, E., & Halperin, E. (2014). Overcoming the barrier of narrative adherence in conflicts through awareness of the psychological bias of naïve realism. *Personality and social psychology bulletin*, *40*(11), 1543-1556.

Nelson, D. W. (2009). Feeling good and open-minded: The impact of positive affect on cross cultural empathic responding. *The Journal of Positive Psychology*, *4*(1), 53–63. https://doi.org/10.1080/17439760802357859

Nielsen, K., & Abildgaard, J. S. (2013). Organizational interventions: A research-based framework for the evaluation of both process and effects. *Work & Stress, 27*(3), 278-297.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Prentice Hall.

Nussbaum, A. D., & Dweck, C. S. (2008). Defensiveness versus remediation: Self-theories and modes of self-esteem maintenance. *Personality and Social Psychology Bulletin*, *34*(5), 599-612.

Pew Research Center. (2014). *Political polarization in the American public* [Report]. https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/

Pew Research Center for the People and the Press. (2015). *Support for same-sex marriage at record high, but key segments remain opposed* [Report]. http://www.people-press.org/2015/06/08/support-for-same-sex-marriage-at-record-highbut-key-segments-remain-opposed/

Pew Research Center. (2019). *Partisan Antipathy: More intense, more personal* [Report]. https://www.pewresearch.org/politics/2019/10/10/how-partisans-view-each-other/

Porter, T., & Schumann, K. (2018). Intellectual humility and openness to the opposing view. *Self and Identity*, *17*(2), 139-162.

Porter, T., Schumann, K., Selmeczy, D., & Trzesniewski, K. (2020). Intellectual humility predicts mastery behaviors when learning. *Learning and Individual Differences*, *80*, 101888.

Oaksford, M., Morris, F., Grainger, B., & Williams, J. M. G. (1996). Mood, reasoning, and central executive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 476–492. https://doi.org/10.1037/0278-7393.22.2.476

Ofosu, E. K., Chambers, M. K., Chen, J. M., & Hehman, E. (2019). Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proceedings of the National Academy of Sciences, 116*(18), 8846-8851.

Okimoto, T. G., & Wenzel, M. (2011). The other side of perspective taking: Transgression ambiguity and victims' revenge against their offender. *Social Psychological and Personality Science*, *2*(4), 373-378.

Paluck, E. L. (2010). Is it better not to talk? Group polarization, extended contact, and perspective taking in Eastern Democratic Republic of Congo. *Personality & Social Psychology Bulletin*, *36*(9), 1170–1185. https://doi.org/10.1177/0146167210379868

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*, 339-367.

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2020). Prejudice reduction: Progress and challenges. *Annual Review of Psychology, 72.*

Paluck, E. L., Shepherd, H., & Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences, 113*(3), 566-571.

Park, J., & Banaji, M. R. (2000). Mood and heuristics: The influence of happy and sad states on sensitivity and bias in stereotyping. *Journal of Personality and Social Psychology*, *78*(6), 1005.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*(1), 65-85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Pettigrew, T. F., Tropp, L. R., Wagner, U., & Christ, O. (2011). Recent advances in intergroup contact theory. *International Journal of Intercultural Relations*, *35*(3), 271–280.

Pierce, J. R., Kilduff, G. J., Galinsky, A. D., & Sivanathan, N. (2013). From glue to gasoline: how competition turns perspective takers unethical. *Psychological Science*, *24*(10), 1986–1994. https://doi.org/10.1177/0956797613482144

Regan, D. T., & Totten, J. (1975). Empathy and attribution: Turning observers into actors. *Journal of Personality and Social Psychology*, *32*(5), 850–856. https://doi.org/10.1037/0022-3514.32.5.850

Rogers, S. L., Howieson, J., & Neame, C. (2018). I understand you feel that way, but I feel this way: the benefits of I-language and communicating perspective during conflict. *PeerJ*, *6*, e4831. https://doi.org/10.7717/peerj.4831

Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive failure as a feature of those holding radical beliefs. *Current Biology*, *28*(24), 4014-4021.

Rollwage, M., & Fleming, S. M. (2021). Confirmation bias is adaptive when coupled with efficient metacognition. *Philosophical Transactions of the Royal Society B*, *376*(1822), 20200131.

Routledge, C., Arndt, J., & Sheldon, K. M. (2004). Task engagement after mortality salience: The effects of creativity, conformity and connectedness on worldview defence. *European Journal of Social Psychology*, *34*(4), 477–487. https://doi.org/10.1002/ejsp.209

Routledge, C. D., & Arndt, J. (2009). Creative Terror Management: Creativity as a Facilitator of Cultural Exploration After Mortality Salience. *Personality and Social Psychology Bulletin*, *35*(4), 493–505. https://doi.org/10.1177/0146167208329629

Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., ... & Folke, T. (2021). The general fault in our fault lines. *Nature Human Behaviour*, 1-11.

Russell, B. (1927). *An Outline of Philosophy.* Taylor & Francis.

Russell, B. (1928). *Sceptical Essays.* New York: W.W. Norton.

Sassenberg, K., & Moskowitz, G. B. (2005). Don't stereotype, think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology*, *41*(5), 506–514. https://doi.org/10.1016/j.jesp.2004.10.002

Sassenrath, C., Hodges, S. D., & Pfattheicher, S. (2016). It's all about the self: When perspective taking backfires. *Current Directions in Psychological Science*, *25*(6), 405-410.

Schein, E. H. (1990). Organizational culture. *American Psychologist, 45*(2), 109–119. https://doi.org/10.1037/0003-066X.45.2.109

Schumann, K., Zaki, J., & Dweck, C. S. (2014). Addressing the empathy deficit: Beliefs about the malleability of empathy predict effortful responses when empathy is challenging. *Journal of Personality and Social Psychology*, *107*(3), 475.

Schwarz, N. (2000). Agenda 2000—Social judgment and attitudes: warmer, more social, and less conscious. *European Journal of Social Psychology*, *30*(2), 149-176.

Shapiro, I. (2006). Extending the framework of inquiry: Theories of change in conflict interventions. *Berghof Handbook*, (5).

Sherif, M., & Hovland, C. I. (1961). Social judgment: *Assimilation and contrast effects in communication and attitude change.* Yale University Press.

Skorinko, J. L., & Sinclair, S. A. (2013). Perspective taking can increase stereotyping: The role of apparent stereotype confirmation. *Journal of Experimental Social Psychology*, *49*(1), 10–18. https://doi.org/10.1016/j.jesp.2012.07.009

Spies, K., Hesse, F., & Hummitzsch, C. (1996). Mood and capacity in Baddeley's model of human memory. *Zeitschrift Für Psychologie Mit Zeitschrift Für Angewandte Psychologie*, *204*(4), 367–381.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate?. *Behavioral and Brain Sciences*, *23*(5), 645-665.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In *Advances in experimental social psychology* (Vol. 21, pp. 261–302). Elsevier.

Steward, J. H. (1972). *Theory of culture change: The methodology of multilinear evolution.* University of Illinois Press.

Symborski, C., Barton, M., Quinn, M., Morewedge, C., Kassam, K., Korris, J. H., & Hollywood, C. A. (2014). Missing: A serious game for the mitigation of cognitive biases. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)* (No. 14295, pp. 1-13).

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Yale University Press.

Tankard, M. E., & Paluck, E. L. (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review, 10*(1), 181-211.

Tankard, M. E., & Paluck, E. L. (2017). The effect of a Supreme Court decision regarding gay marriage on social norms and personal attitudes. *Psychological Science, 28*(9), 1334-1344.

Tarrant, M., Calitri, R., & Weston, D. (2012). Social Identification Structures the Effects of Perspective Taking. *Psychological Science*, *23*(9), 973–978. https://doi.org/10.1177/0956797612441221

Thompson, L. L. (1991). Information exchange in negotiation. *Journal of Experimental Social Psychology*, *27*(2), 161–179. https://doi.org/10.1016/0022-1031(91)90020-7

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, *1*(4), 285–293.

Todd, A. R., Bodenhausen, G. V., & Galinsky, A. D. (2012). Perspective taking combats the denial of intergroup discrimination. *Journal of Experimental Social Psychology*, *48*(3), 738–745. https://doi.org/10.1016/j.jesp.2011.12.011

Todd, A. R., & Burgmer, P. (2013). Perspective taking and automatic intergroup evaluation change: Testing an associative self-anchoring account. *Journal of Personality and Social Psychology*, *104*(5), 786–802. https://doi.org/10.1037/a0031999

Todd, A. R., & Galinsky, A. D. (2014). Perspective-Taking as a Strategy for Improving Intergroup Relations: Evidence, Mechanisms, and Qualifications. *Social and Personality Psychology Compass*, *8*(7), 374–387. https://doi.org/10.1111/spc3.12116

Trope, Y., & Liberman, N. (2010). Construal-Level Theory of Psychological Distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Tuller, H. M., Bryan, C. J., Heyman, G. D., & Christenfeld, N. J. (2015). Seeing the other side: Perspective taking and the moderation of extremity. *Journal of Experimental Social Psychology*, *59*, 18-23.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Valente, T. W. (2012). Network interventions. *Science, 337*(6090), 49-53.

Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, *49*(3), 577–585. https://doi.org/10.1037/0022-3514.49.3.577

Van Loon, A., Bailenson, J., Zaki, J., Bostick, J., & Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PloS one*, *13*(8), e0202442.

Vescio, T. K., Sechrist, G. B., & Paolucci, M. P. (2003). Perspective taking and prejudice reduction: The mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology*, *33*(4), 455-472.

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, *125*(5), 617.

Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, *23*(1), 73-82.

Ward, A., Atkins, D. C., Lepper, M. R., & Ross, L. (2011). Affirming the self to promote agreement with another: lowering a psychological barrier to conflict resolution. *Personality & Social Psychology Bulletin*, *37*(9), 1216–1228. https://doi.org/10.1177/0146167211409439

Wegener, D. T., Petty, R. E., & Smith, S. M. (1995). Positive mood can increase or decrease message scrutiny: The hedonic contingency view of mood and message processing. *Journal of Personality and Social Psychology*, *69*(1), 5–15. https://doi.org/10.1037/0022-3514.69.1.5

Wegener, D. T., Petty, R. E., & Dunn, M. (1998). The metacognition of bias correction: Naive theories of bias and the flexible correction model. In V. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions.* New York: Sage.

Weisz, E., Ong, D. C., Carlson, R. W., & Zaki, J. (2020). Building empathy through motivation-based interventions. *Emotion*. https://doi.org/10.1037/emo0000929

Weisz, E., & Cikara, M. (2021). Strategic Regulation of Empathy. *Trends in Cognitive Sciences*, *25*(3), 213–227. https://doi.org/10.1016/j.tics.2020.12.002

Weisz, E., & Zaki, J. (2018). Motivated empathy: A social neuroscience perspective. *Current Opinion in Psychology*, *24*, 67–71. https://doi.org/10.1016/j.copsyc.2018.05.005

Weisz, E., Chen, P., Ong, D. C., Carlson, R. W., Clark, M. D., & Zaki, J. (under review). A Brief Intervention to Motivate Empathy in Middle School.

Yang, D. Y. J., Preston, J. L., & Hernandez, I. (2013). Polarized attitudes toward the Ground Zero mosque are reduced by high-level construal. *Social Psychological and Personality Science*, *4*(2), 244-250.

Yeager, D. S., Trzesniewski, K. H., & Dweck, C. S. (2013). An implicit theories of personality intervention reduces adolescent aggression in response to victimization and exclusion. *Child Development*, *84*(3), 970-988.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They’re not magic. *Review of Educational Research*, *81*(2), 267-301.

Yee, N., Bailenson, J. N., & Ducheneaut, N. (2009). The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research*, *36*(2), 285-312.

SUPPLEMENT: MEASURING OPEN-MINDEDNESS

Self-report measures

Researchers have used a range of self-report scales to measure different components of open-mindedness. In Supplemental Table 1, we provide an overview of the scales that we perceive to be most related to open-mindedness. This table includes measures that assess open-mindedness directly, as well as some of its hypothesized underlying cognitive and motivational mechanisms.

Term related to open-mindedness	Term related to closed-mindedness	Definition(s)	Relationship to open-mindedness	Process	Discipline(s)	Scale(s)
Receptiveness (to opposing views)	—	"a willingness to expose oneself to, and to thoughtfully and fairly consider, the opposing views of others" (Minson, Chen & Tisley, 2018)	Core - general	Cognitive and affective sub-components	Social psychology	Receptiveness to opposing views scale (Minson, Chen & Tisley, 2018)
Open-Minded Cognition	Closed-minded cognition	"a cognitive style marked by a willingness to consider a variety of intellectual perspectives, values, attitudes, opinions, or beliefs, even those that contradict the individual's prior opinion" (Price et al., 2015)	Core - general	Cognitive	Social psychology, political psychology; judgment & decision-making; education; cognitive psychology	Open-Minded Cognition-General, Political, and Religious Versions (Price et al., 2015)
Actively open-minded thinking	—	"a multifaceted construct encompassing the cultivation of reflectiveness rather than impulsivity, the seeking and processing of information that disconfirms one's belief (as opposed to confirmation bias in evidence seeking), and the willingness to change one's beliefs in the face of contradictory evidence" (Stanovich & West, 1997)	Core - general	Cognitive	Social psychology, political psychology; judgment & decision-making; education; cognitive psychology	Actively Open-Minded Thinking Scale (Stanovich & West, 2007)
—	Dogmatism	"relatively unchangeable, unjustified certainty" (Altemeyer, 2002)	Core - general	Cognitive	Social psychology, philosophy; political psychology	Altemeyer DOG scale (1996); D Scale (Rokeach, 1960)

Intellectual humility	Intellectual arrogance	“recognizing that a particular personal belief may be fallible, accompanied by an appropriate attentiveness to limitations in the evidentiary basis of that belief and to one's own limitations in obtaining and evaluating relevant information” (Leary et al., 2017)	Core - general and specific	Cognitive	Positive psychology, social psychology, philosophy, education	Comprehensive Intellectual Humility Scale (Krumrei-Mancuso & Rouse, 2016); General Intellectual Humility Scale (Leary et al., 2017); Sociopolitical Comprehensive Intellectual Humility Scale (Krumrei-Mancuso & Newman, 2020)
—	Belief superiority	“the belief that one’s own views are superior to alternatives” (Raimi & Jongman-Sereno, 2020)	Core - specific	Cognitive	Social psychology	General Belief Superiority Scale, Domain-Specific Belief Superiority (Raimi & Jongman-Sereno, 2020)
Latitude of acceptance	Latitude of rejection	“the range of positions that an individual accepts,” (Eagly & Telaak, 1972)	Core - specific	Cognitive	Social psychology	Latitude of acceptance measure
–	Social vigilantism	“an enduring individual difference that captures the tendency some individuals have to assert their ‘superior’ beliefs onto others to correct others’ more ‘ignorant’ opinions for the ‘greater good.’” (Saucier & Webster, 2010)	Related concept	Behavioral	Social psychology	Social Vigilantism Scale (Saucier & Webster, 2010)
—	Myside bias	“the tendency to evaluate propositions from within one’s own perspective when given no instructions or cues (such as within-participants conditions) to avoid doing so”	Related concept	Cognitive	Social psychology	Scale items (Stanovich & West, 2007)
Openness [to experience]	—	“… seen in the breadth, depth, and permeability of consciousness and in the recurrent need to enlarge and examine experience” (McCrae & Costa, 1997)	Related concept	Cognitive	Personality psychology	NEO Personality Inventory (Costa & McCrae, 1992); Big Five Inventory (Goldberg, 1993); Ten-Item Personality Inventory (Gosling et al., 2003)
–	Motive attribution asymmetry	“a tendency to attribute love vs. hate to one’s ingroup to a greater degree than to one’s outgroup and to attribute hate vs. love to one’s outgroup to a greater degree than to one’s	Related concept	Cognitive	Social psychology	Scale items (Waitz, Young & Jingest, 2014)

		ingroup” (Waitz, Young & Jingest, 2014)				
Wise reasoning	—	“A set of meta-cognitive strategies that guide people toward managing complexity and balancing different interests.” (Santos, Huynh, & Grossmann, 2017)	Related concept	Cognitive	Philosophy; positive psychology	Situated Wise Reasoning Scale (Brienza et al., 2018)
–	Self-righteousness	“the conviction that one’s beliefs and actions are correct, especially in contrast to the alternative beliefs and actions of others”	Related concept	Cognitive		Falbo & Shepperd, 1986, p. 145
Cognitive empathy/Perspective-taking	—	“the tendency to spontaneously adopt the psychological point of view of others” (Davis, 1983)	Related concept	Cognitive	Social psychology	Interpersonal Reactivity Index-Perspective -Taking Subscale (Davis, 1983)
Cognitive flexibility	Cognitive rigidity, cognitive inflexibility	“the human ability to adapt the cognitive processing strategies to face new and unexpected conditions in the environment” (Cañas et al., 2003)	Underlying mechanism	Cognitive	Cognitive psychology; Communication studies	Cognitive Flexibility Scale (Martin & Ruben, 1995), Alternative Uses Task (Guilford, 1967), Verbal Fluency Task (Tombaugh, Kozak, & Rees, 1999; Troyer, Moscovitch, & Winocur, 1997)
Integrative complexity	—	“a concept and measure of the degree to which cognitive processing involves recognizing multiple perspectives and possibilities and integrating them into a coherent view”	Underlying mechanism	Cognitive	Conflict resolution; Personality psychology	Integrative complexity coding scheme (Baker-Brown, Ballard, Bluck, de Vries, Suedfeld, & Tetlock, P. E., 1992).
Metacognitive awareness	—	“the ability to reflect upon, understand, and control one's learning.” (Schraw & Sperling, 1994)	Underlying mechanism	Cognitive	Education; Cognitive psychology	Metacognitive awareness inventory (Schraw & Sperling, 1994)
Epistemic curiosity	—	a “drive to know” (Berlyne, 1954)	Underlying mechanism	Motivational	Personality psychology	Epistemic Curiosity scale (Litman & Spielberg, 2003)
—	Need for closure	“individuals’ desire for a firm answer to a question and an aversion toward ambiguity” (Kruglanski & Webster, 1996)	Underlying mechanism	Motivational	Social psychology	Need for Closure Scale (Kruglanski, Atash, De Grada, Mannetti, & Pierro, 1997)
Tolerance for ambiguity	Intolerance for ambiguity	“the way an individual (or group) perceives and processes information about ambiguous situations or stimuli when confronted by an array of unfamiliar, complex, or	Underlying mechanism	Motivational	Social psychology	Intolerance for ambiguity scale (Martin & Parker, 1995; Martin & Westie, 1959; Uemura, 2001)

		incongruent clues” (Furnham & Ribchester, 1995)				
Need for cognition	—	“the tendency for an individual to engage in and enjoy thinking” (Cacioppo & Petty, 1982)	Underlying mechanism	Motivational	Social psychology	Need for Cognition Scale (Cacioppo & Petty, 1982)

Supplemental Table 1. Open-mindedness measures. This table displays the terminology, definitions, theoretical background, and scales for constructs that describe, underlie, or relate to open-mindedness.

The scales that measure what we consider to represent trait-based, “core open-mindedness” include the Receptiveness to Opposing Views scale (Minson, Chen, & Tinsley, 2018), Open-Minded Cognition scale (Price et al., 2015), Actively Open-Minded Thinking Scale (Haran, Ritov, & Mellers, 2013), Comprehensive Intellectual Humility Scale (Krumrei-Mancuso & Rouse, 2016), General Intellectual Humility Scale (Leary et al., 2017), Altemeyer’s Dogmatism Scale (Altemeyer, 2002), and Haiman’s revised dogmatism scale (Haiman, 1964). Further information on the differences between the scales shown in the table can be found in the papers associated with their development. These scales are useful for studying open-mindedness interventions insofar as they may help identify individuals who are more or less susceptible to interventions to begin with. Furthermore, they might be used to assess changes in dispositional open-mindedness as a result of interventions that can target long-term change.

Although these general trait-based scales can measure dispositional open-mindedness, they do not reflect the fact that individuals can be more or less open-minded on specific issues, or that their open-mindedness can be state-dependent. Thus, researchers have also developed measures of receptivity to alternative views on specific issues. For instance, one issue-specific measure is belief superiority, which uses one item that asks an individual “how much more correct their belief about that issue is compared with other people’s beliefs.” (Raimi & Jongman-Sereno, 2020). Whereas an individual might be flexible on certain issues, they can be quite dogmatic about others, and this measure allows for those distinctions. More specific intellectual humility measures have also been developed, including versions that focus on sociopolitical issues (Krumrei-Mancuso & Newman, 2020), religion (Hopkin et al., 2014) and specific topics (Hoyle et al., 2016). Another measure that can be used to study more specific open-mindedness is latitude of acceptance, which represents all attitudes on a scale that an individual deems acceptable for another person to hold. Whereas, traditionally, latitudes of acceptance have been used as indicators of susceptibility to attitude change, Dieffenbach & Lieberman (in prep, 2021, Chapter 4) used them as indicators of receptivity to opposing views on a particular issue. In summary, measures that assess open-mindedness about specific issues can be useful in measuring the efficacy of interventions; they can be administered at different time points relative to an intervention, including before, after, and at follow-up.

Finally, some measures of open-mindedness frame questions within a particular social context (e.g., the Situated Wise Reasoning Scale from Brienza et al., 2018, along with "responses to disagreement" measures developed by Porter & Schumann, 2018). Given that these measures are contextualized, they can assess participants' intended or past open-minded behavior in addition to their thought processes. In these measures, participants are asked to think about a past or hypothetical situation in which they had a conversation with someone that they disagreed with. Then, they answer questions related to their thoughts and actions in that situation. For instance, they might indicate whether they tried to take the other person's perspective, listened to the other person's reasoning for their opinions, and made positive or negative attributions about what the other person was saying.

Moving beyond self-report

Although self-report scales are easy to administer, they have their limitations. Participants may not have accurate insight into the extent to which they are open-minded. Further, given that open-mindedness is often perceived to be a valued and desirable trait, people may engage in self-enhancement when reporting on it. For instance, previous studies have found a significant correlation between intellectual humility and socially desirable responding (Haggard et al., 2018; Krumrei-Mancuso et al., 2020; Krumrei-Mancuso & Rouse, 2016). Also, most of these scales tend to focus only on the cognitive and motivational components of open-mindedness. Most do not measure the extent to which people engage in open-minded behavior and experience positive or negative affect within real social interactions. Therefore, researchers have begun to use other types of measures, including behavioral, physiological, neural, and cognitive to complement the traditional self-report measures.

Behavioral measures

One simple method for measuring open-minded behavior is through obtaining reports from others. For instance, Meagher et al. (2021) collected both self and peer ratings of intellectual humility after participants had conversations about controversial sociopolitical issues. This work found that people tend to use different criteria to evaluate intellectual humility within themselves versus another person, providing evidence for the added value from using both types of reports together. Furthermore, researchers have also relied on informant reports, in which a participant’s acquaintances rates the participant’s behavioral tendencies. When multiple informants report on the same participant, this is referred to as using a ‘360’ approach. For example, the Conflict Competencies 360 (MD-ICCCR, 2021) uses multi-rater feedback to assess how people feel, think, and behave in conflict situations.

In addition, researchers can investigate the presence of open-mindedness in written language. Traditional methods have involved hand-coding language in terms of its overall content and also whether or not it contains certain features (e.g., number of questions, number of definitive statements, etc.). For example, Kugler & Coleman (2020) operationalize behavioral complexity during a conversation based on the ratio of participants' use of language related to 'inquiry' versus 'advocacy.' Recently, researchers have also taken a novel approach to examining language for open-mindedness using natural language processing (NLP). For instance, Yeomans et al. (2020) developed an algorithm that can detect levels of 'conversational receptiveness' in text from online conversations. This algorithm was useful in that it allowed researchers to identify the level of receptiveness in new pieces of text, but it also helped to identify a 'receptiveness formula' that could be taught as part of an intervention. Morteza et al. (in prep) have developed an NLP model that can assess whether or not text contains expressions of intellectual humility. Google Jigsaw's Perspective API classifier can assess how likely it is that text is toxic. And relatedly, Zhang et al. (2018) developed a model that can predict whether an online conversation is likely to become uncivil based on the features of the beginning of the conversation.

Other useful behavioral indicators include nonverbal cues, such as body language, facial expressions, and intonation. These can be measured at the individual level, and also at the dyad or group level by examining interpersonal synchrony on a moment-by-moment basis. For instance, research has found that bodily synchrony decreases during arguments (Paxton & Dale, 2013). See Bousmalis et al. (2013) for an overview of non-verbal cues that can be measured during disagreement.

Researchers might also be interested in measuring the real-world behavioral consequences of people being open- or closed-minded. However, further research is needed to determine what real-world outcomes are most relevant, which will depend on the context. For instance, workplaces might be interested in assessing whether open-mindedness interventions improve employee engagement, reported belonging, retention, and even performance reviews. It may be desirable to capture the ideological diversity within local communities or even the number of bipartisan bills that are passed at local or national levels. Although it can be difficult to isolate the relationship between interventions and these measures, if possible, they could provide a richer understanding of the interventions' real-world impact.

Neural and physiological measures

In addition to behavioral measures, recent research has begun to use neuroimaging to detect concepts related to open-mindedness using univariate, multivariate, and neural synchrony-based approaches. For instance, Kaplan, Gimbel, & Harris (2016) found that information that challenged individuals' political beliefs produced increased activity in the brain's default mode network, a set of structures involved in thinking about the self and others. Individuals who were more open to changing their minds showed less activity in insula and amygdala, regions associated with affective processing, when evaluating this information. Studies have found that the brain's dorsolateral prefrontal cortex becomes more active when people are counterarguing against challenging information (O'Donnell et al., 2020; Lieberman et al., in prep). Falk, Spunt, and Lieberman (2013) found that perspective-taking ability and motivation modulated neural activity in the default mode network and in affective regions in response to political candidates who held similar or opposing views. Although a 'neural "signature" of open-mindedness' has yet to be discovered, research in related areas suggests that this may be possible. For instance, neuroscientists have successfully used machine learning to classify affective experiences such as physical pain (Wager et al., 2013) and discrete emotion categories in the brain (Kragal & Labar, 2015) from brain data.

Researchers have also examined the extent to which people show similar brain responses to other people who have a similar viewpoint to them. For example, studies have examined the extent to which people show similar neural responses to each other when they are engaged in different types of conversations, finding that neural synchrony is greater during agreement in comparison to disagreement (Hirsch et al., 2021; Kealoha et al., in prep). Recent studies have also found that ‘neural polarization’ exists between groups that hold different political opinions, such that they show greater neural synchrony with other members of their in-group when watching political videos (Dieffenbach et al., 2021; Leong et al., 2020; Moore-Berg et al., 2020). Dieffenbach et al. (2021) showed that a classification algorithm could categorize people into their ideological group at above-chance levels based on neural synchrony patterns. In other words, they were able to predict people’s ideological viewpoints using a method which they refer to as the ‘neural reference groups’ approach. Furthermore, Dieffenbach et al. (in prep) built upon this idea, reasoning that if individuals who are in different mindsets show differentiable neural responses, then this approach might be able to reveal whether or not a mindset intervention had shifted participants’ information processing. They found that after participants went through a self-affirmation intervention, there was significant differentiation in neural synchrony between the intervention and control groups, suggesting that the intervention caused a meaningful shift in participants’ information processing of speech that ran counter to their own views.

In addition to neural indicators, researchers can also examine physiological markers that reflect stress responding and arousal during or after moments of disagreement or interacting with out-group members. Previous studies have looked at changes in cortisol reactivity following empathy-related interventions (Influs et al., 2019; Page-Gould et al., 2008). Other studies have examined the role of oxytocin (Arueti et al., 2013), heart rate (Martinez-Tur et al., 2014), galvanic skin response (Porier & Lott, 1967) and levels of proinflammatory cytokines (Kiecolt-Glaser et al., 2015; Kiecolt-Glaser et al., 2005) during or after interpersonal interactions.

Cognitive and perceptual tasks

Some researchers have also developed psychological signatures of ideological thinking based off of individuals' performance on cognitive tasks. Using these tasks, studies have found that cognitive flexibility, metacognitive awareness, and speed of evidence accumulation are highly predictive of constructs related to open-mindedness (Zmigrod, 2020). Zmigrod et al. (2020a) used a data-driven approach to identify signatures of dogmatism and intellectual humility based on participants' performance on 37 cognitive tasks and their completion of 27 personality surveys. In particular, they found that dogmatic individuals tended to show slower rates of evidence accumulation during the cognitive tasks. In other studies, Zmigrod and colleagues have also found that individuals who show greater cognitive flexibility are less likely to have strong partisan identities (Zmigrod et al., 2020b), less likely to have extreme attitudes (Zmigrod et al., 2019a), and more likely to be intellectually humble (Zmigrod et al., 2019b). In addition, Rollwage et al. (2018) found that dogmatic individuals show impaired metacognition (the ability to monitor and regulate one's thoughts). As part of a simulation study, Rollwage et al. (2021) suggest that cognitive training that attempts to boost metacognitive awareness might help ameliorate the negative impact of confirmation bias.

Chapter 5 - General Discussion

Overview of Findings

When people *see* things differently, the divide between them is evident in their brain activity. Synchrony-based analyses can detect *neural polarization* between people who hold different viewpoints. Building upon this idea, this dissertation presents work from two studies that demonstrate how neural synchrony analyses can be leveraged to help define the divide between political partisans, predict its presence, and measure the impact of interventions that aim to reduce it. This dissertation introduces a synchrony-based technique called the *neural reference groups approach*, demonstrating how it can be used to make different kinds of predictions about individuals' subjective construals. In the first study, I showed that it is possible to use this technique to predict individuals' political views. In the second study, I demonstrated that this technique can predict whether or not participants have gone through an intervention to shift their mindset.

The first study in this dissertation, which was presented in Chapter 2, was conducted as a proof of concept with two novel contributions. First, it demonstrated that neural polarization work can be conducted effectively using functional near infrared spectroscopy (fNIRS). Prior work in this domain had only been conducted in functional magnetic resonance imaging (fMRI), which provides a high spatial resolution, but has a high cost and is immobile. In contrast, fNIRS can be packed into carry-on luggage and set up in a pop-up neuroscience lab anywhere in the world, as we demonstrated in this study. Second, this study demonstrated that it is possible to predict individuals' political views from their brain data in DMPFC at above-chance accuracy levels. The neural reference groups approach is a fairly simple classifier that is easy to implement. It requires collecting neural data from two separate 'reference groups,' computing an average brain response for each group, and then comparing future participants to those references to identify which reference group they more closely resemble. We propose that this approach can have many potential applications. For instance, it could be applied to identify the extent to which public health messages are resonating with viewers, or it could be used to identify employees who are in a mental state of burnout.

The second study, presented in Chapter 3, built upon and extended our prior work. With this study, we had two primary aims. First, our goal was to replicate our findings from the previous study in a new sample, using a different political issue, and recording signal from all more cortical regions. We focused our analyses on the mentalizing network, which included VMPFC, DMPFC, and bilateral TPJ/IPL. We found that liberals had significantly greater synchrony in the mentalizing network with other liberals when they were watching videos that they disagreed with. However, liberals did not show greater within-group synchrony for videos that they agreed with. In addition, conservatives did not show the neural polarization effect when watching either video type. Our second and primary goal of this study was to demonstrate that neural synchrony analyses can be used to identify whether an intervention has made an impact on a person’s subjective construal. Using synchrony analyses, we found that the intervention and control liberals showed greater within-group than between-group synchrony, which suggested that the two groups had distinctive patterns of neural responding. Furthermore, applying the neural reference groups approach to neural data from the mentalizing network, we were able to predict whether or not individuals had gone through a mindset shifting intervention at a significantly above-chance accuracy level.

Given that this work demonstrates the power of the neural reference groups in being able to detect the impact of mindset-shifting interventions, in Chapter 4, I provided a comprehensive narrative review of interventions that have been developed to improve open-mindedness. I hope that this chapter can provide researchers and practitioners with a useful resource to help facilitate the development of impactful interventions that can help reduce political polarization. Social neuroscientists who are interested in using the neural reference groups approach to test the impact of mindset interventions may look to this review as a reference. In addition to reviewing the prior literature, I provided a conceptual model for understanding open-mindedness as being situated within a dynamical system. As part of this model, I proposed that interventions can induce open-minded thinking by operating on cognitive and/or motivational pathways. Then, in order to sustain open-minded thinking during social interactions, individuals must also have appropriate emotion regulation and social skills. Furthermore, social norms and social structures facilitate or impede open-mindedness by influencing the extent to which people are willing and able to engage in open-minded thinking and behave in an open-minded manner.

Future Directions

There are many potential directions that can emerge from the present work. This dissertation's empirical studies are the first of their kind to use synchrony-based classification to predict political viewpoints and to measure the impact of a mindset intervention. However, these studies did not conduct classification on an out-of-sample dataset, which would be a natural next step. If future research could identify that the neural reference groups approach can successfully predict the mindsets of new individuals who have not been a part of the development of the reference groups, then it would demonstrate its ability to be used in more real-world applications.

Another way to build upon this work would be to build out a diversity of neural reference groups classification models to accommodate different patterns of neural synchrony clustering. For instance, as we found in our second study, if people are in different psychological states when they process a stimulus, those states could be marked by different clustering patterns. One group might cluster together tightly while the other one is more diffuse. Given that the current neural reference groups approach relies on an assumption that two groups have become tightly clustered, this approach may be limited in its ability to handle data that defies this assumption. Therefore, testing additional models might help to boost the accuracy of these classifiers. Another way to extend the utility of the neural reference groups approach would be to build models that can accommodate patterns of brain activity across multiple regions, rather than pulling timecourse data from single brain regions at a time, which is the current approach. In this way, the neural reference groups approach might be able to identify momentary instantiations of certain mindsets (e.g. a 'counterarguing' brain state) that occur in more dynamic, social interactions.

In conclusion, this dissertation has demonstrated ways in which novel neuroimaging approaches can be applied to understand political polarization and to measure the impact of open-mindedness interventions. It is my hope that this work can help facilitate new insights in order to help to heal the partisan divide and facilitate better understanding.