THE ROLE OF THE MIRRORING AND MENTALIZING SYSTEMS IN ACTION UNDERSTANDING

by

Amna Hyder

B.Sc., McMaster University, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF NEUROSCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

Department of Medicine

THE UNIVERSITY OF BRITISH COLUMBIA

Vancouver

January 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis/dissertation entitled:

The role of the mirroring and mentalizing systems in action understanding

submitted by Amna Hyder in partial fulfillment of the requirements for

the degree of Masters of Neuroscience

in The Faculty of Medicine

Examining Committee:

Dr. Christine Tipper
Co-supervisor

Dr. Liisa Galea
Co-supervisor

Dr. Todd Handy
Supervisory Committee Member

Dr. Anthony Herdman
Additional Examiner

Additional Supervisory Committee Members:

Supervisory Committee Member

Supervisory Committee Member

Abstract

Mirror neurons, which are critical for social learning, respond to both the observation and execution of an action[1–3]. They do so by forming a complex network known as the Mirrors Neuron System (MNS) in which visual representations of motion fashion a corresponding motor representation. Another essential feature of social learning is the ability to reason about the minds of others by inferring not just *how* they do something but *why* they are doing it[4]. The mentalizing system (MZN) is implicated in the process of inferring other's mental states[5]. Previous studies three gaps that we explore: 1) How the two systems differ within the same subject 2) Whether spatiotemporal activation patterns can uniquely identify them and 3) How electrical brain activity measured with EEGs can be used to investigate these questions as opposed to fMRI which is frequently used to study the MZN and MNS. This study utilizes high-density EEG to quantify dynamics in functional brain networks supporting mirroring and mentalizing processes in neurotypical adults. We use the photo judgement task[6] which has been shown to differentiate the MNS and MZN. Participants are shown pictures of faces or hands and asked about *how* (mirroring) or *why* (mentalizing) the actions are being performed. We see clear differences between mirroring and mentalizing tasks that can be detected by EEG. Brain activity appears to diverge around 300 ms after stimulus onset, and several EEG ERP components uniquely identify mirroring and mentalizing activity. Using mu and beta suppression as markers for mirroring and mentalizing activity respectively, the data suggests that the MNS is more active in processing action means associated with facial expressions, whereas the MZN is more active in processing intent associated with hand movements. Using microstate analysis, we show that ~300 ms after stimulus presentation, the brain undergoes several state transitions while processing intent, whereas while processing action means (mirroring), the brain appears to stay in one stable state. Investigating source space results shows that indeed the *how* and *why* conditions more strongly activate regions associated with the MNS (occipital and left superior temporal gyrus) and MZN (medial prefrontal cortex) respectively.

Lay Summary

Being able to successfully interpret the actions of other people is critical both for motor learning and social cognition. There are two major systems implicated in this process for humans – the mirror neuron system (MNS) and the mentalizing network (MZN), processing the *how* and *why* respectively. In this study we explore: 1) How the MNS and MZN differ within the same subjects; 2) How the timing of activation patterns differs between them; and 3) How electrical brain activity measured with EEG can be used to investigate these questions. We collect high density EEG signals as individuals take part in an experiment that can preferentially activate the MNS and MZN. We find that the two systems are distinguishable using EEG. Moreover, we find that the MNS precedes the MZN and that the activity of each is dependent on the type of stimuli presented.

Preface

Data was collected by Ryan Kopstick at BC Children's Hospital. The design of the experiment was done by Ryan Kopstick and my supervisor Christine Tipper, and was based on an experiment previously done Spunt and Adolfs (2014)[6] at the California Institute of Technology. The research project, of which this thesis is a part, received research ethics approval from the University of British Columbia Behavioural Research Ethics Board, Project Name "Social Brain", No. H16-00915, 26/04/2016. This thesis is an original work of me, Amna Hyder and all the analysis and research questions relating to the data were developed and conducted by me.

List of Tables

Table	Description
1	Roles of the MNS and MZN
2	Properties of the Photo Judgement Task
3	ERP components of interest
4	T-Test results from Reaction Times across conditions
5	T-Test results from Task Accuracy across conditions
6	ERP peaks with P-Values that are significantly different between *how* and *why* conditions

List of Figures

Figure	Description
1	Modality specific subnetworks of the MNS
2	Modality specific subnetworks of the MZN
3	Stimuli for a study comparing MNS and MZN
4	Brain regions involved in the MNS and MZN
5	Image explaining the false belief task
6	Factorial Design for Photo Judgement Task
7	Examples of image blocks for Photo Judgement Task
8	Instructions for Photo Judgement Task
9	A list of all questions in the Photo Judgement Task
10	Pre-processing steps
11	Reaction Times across conditions
12	Task Accuracy across conditions
13	Sex based differences in Reaction Time
14	Sex based differences in Task Accuracy
15	Expectation violations or ratio of correct "yes" answers across conditions
16	ERPs in central electrodes for all conditions
17	Averaged ERPs for *how* and *why* conditions in central electrodes
18	Averaged ERPs for *face* and *hand* conditions in central electrodes
19	Averaged ERPs for *face* and *hand* conditions in occipital electrodes
20	Bar graph of number of p-values under 0.05 for all conditions
21	Face specific component for *how* vs *why* conditions
22	Sex based differences in ERPs for central electrodes
23	Sex based differences in ERPs for left parietal electrodes
24	Scatterplot of N400 amplitude and task accuracy across conditions
25	Time frequency plots for averaged *how* and *why* conditions at central electrodes

26 Time frequency plots for all *how* and *why* conditions at central electrodes
27 Time frequency plots in right central electrodes
28 Time frequency plots in left frontal electrodes
29 Root Mean Squared Error for microstates
30 Blocks representing length of each microstate
31 Chart showing the mid-point of each microstate used for source localization
32 Source plots of the *how face* condition
33 Source plots of the *why face* condition
34 Source plots of the *How minus Why* contrast at microstate mid-points determined from the entire population
35 Source plots of the *How minus Why* contrast at microstate mid-points of just the *how* condition
36 Source analysis for the first three microstates in multiple conditions.

List of Abbreviations

MNS	Mirror Neuron System
MZN	Mentalizing System
ERP	Event-Related Potentials
fMRI	Functional Magnetic Resonance Imaging
IPL	Inferior Parietal Lobule
STS	Superior Temporal Sulcus
MTG	Middle Temporal Gyrus
IFG	Inferior Frontal Gyrus
aIPS	Anterior Intraparietal sulcus
TPJ	Temporoparietal junction
mPFC	Medial prefrontal cortex
ASD	Autism spectrum disorder
EEG	Electroencephalography
hdEEG	High density Electroencephalography
ICA	Independent Component Analysis
PCA	Principle component analysis
EKG	Electrocardiography
RMSE	Root mean squared error
MNE	Minimum norm estimate
ANOVA	Analysis of variance
MEG	Magnetoencephalography
LPP	Late positive potential
N400	Negative potential around 400s
N170	Negative potential around 170s
TFA	Time frequency analysis
JA	Joint attention
GFP	Global field potential

Acknowledgements

I would like to acknowledge first and foremost my committee for going above and beyond in getting me into this lab to complete this project. My co-supervisors, Christine Tipper and Liisa Galea as well as Todd Handy have been immensely supportive of me and my work in this journey. Their advice has been so helpful in getting me to create a comprehensive and exciting project. I would also like to acknowledge Ryan Kopstick for the great data he collected – the study would not have been possible without it. Moreover, I would like to thank my dear friend and mentor Ella Weik whose guidance has been wonderful in giving me a better idea of how proper research should be conducted. Lastly, I would like to thank my husband, brothers and parents for supporting me throughout the process – this thesis is for them.

Table of Contents

ABSTRACT 2

LAY SUMMARY 4

PREFACE 5

LIST OF TABLES 6

LIST OF FIGURES 7

LIST OF ABBREVIATIONS 9

ACKNOWLEDGEMENTS 10

1 BACKGROUND 13

- 1.1 Mirroring System 13
- 1.2 Mentalizing System 16
- 1.3 Interplay Between Two Systems 18
- 1.4 Implications 25
- 1.5 Limitations of Previous Studies 28
- 1.6 Neuronal Basis of EEG 31
- 1.7 EEG markers of the MNS and MZN 33

2 AIMS AND HYPOTHESES 36

3 METHODS 38

- 3.1 Data collection 38
- 3.2 Analysis 42
  - *3.2.1 Preprocessing* 42
  - *3.2.2 ERP analysis* 46
  - *3.2.3 Cluster Permutation Analysis* 46
  - *3.2.4 Time Frequency Analysis* 48
  - *3.2.5 Microstates* 49
  - *3.2.6 Source analysis* 50

4 RESULTS 51

- 4.1 Cleaning Data 51
- 4.2 Behavioural 51
  - *4.2.1 Do we see sex-based differences?* 53
  - *4.2.2 How many times is the correct answer "yes" in each condition?* 55
- 4.3 ERP Analysis 57
  - *4.3.1 Mentalizing vs. Mirroring* 58
  - *4.3.2 Stimulus specific effects* 60
  - *4.3.3 Is there a face specific effect that relates to mirroring or intent?* 63
  - *4.3.4 Sex-Based Differences in ERP* 65

4.3.5 Is N400 amplitude correlated to behavioural results in any condition? *66*
4.4 Do we see differences in alpha and beta suppression? 67
4.5 What is the sequence of state transitions in each condition? 71
4.6 What sources are activated? 73
5 DISCUSSION 79
5.1 Behavioural Analysis 79
5.2 ERP analysis 81
5.2.1 Late ERP components *81*
5.2.2 Middle ERP components *85*
5.2.3 Early ERPs *86*
5.2.4 Importance of keeping the stimuli the same *87*
5.2.5 Face-How Interaction effects. *88*
5.2.6 Sex based differences *89*
5.3 Task based oscillatory frequencies 90
5.4 Microstate Analysis 94
5.5 Source Analysis 97
5.6 Limitations 100
6 CONCLUSION 102
6.1 Future Work 105
OVERALL SUMMARY 106
APPENDIX A 110
APPENDIX B 111
APPENDIX C 116
APPENDIX D 122
APPENDIX E 125
APPENDIX F 127
APPENDIX G 129
APPENDIX H 132
APPENDIX I 134
WORKS CITED 135

1 Background

"It is difficult to overstate the importance of understanding mirror neurons and their function. They may well be central to social learning, imitation, and the cultural transmission of skills and attitudes—perhaps even of the pressed-together sound clusters we call words. By hyper developing the mirror-neuron system, evolution in effect turned culture into the new genome. Armed with culture, humans could adapt to hostile new environments and figure out how to exploit formerly inaccessible or poisonous food sources in just one or two generations—instead of the hundreds or thousands of generations such adaptations would have taken to accomplish through genetic evolution. Thus, culture became a significant new source of evolutionary pressure, which helped select brains that had even better mirror-neuron systems and the imitative learning associated with them. The result was one of the many self-amplifying snowball effects that culminated in Homo sapiens, the ape that looked into its own mind and saw the whole cosmos reflected inside."

— V.S. Ramachandran[7]

A path leads from identification by way of imitation to empathy, that is to the comprehension of the mechanism by which we are enabled to take up any attitude at all towards another mental life.

— Freud[8]

1.1 Mirroring System

Mirror neurons were first discovered in neural recordings of the macaque monkey, and are described as neurons that fire both when an animal performs an action and passively observes the same action[9]. Since this discovery, numerous studies sought to identify a mirror neuron system (MNS) in the human brain because of its potential roles in imitation, action understanding and social functioning[8,10]. As the MNS allows for individuals to map an observed action onto their own motor system, actions can be understood "from the inside"[11]. This leads to the hypothesis that the mirror system allows us to understand the means and goals of an action, and may underpin both gestural communication and empathy[11–14]. There is plenty of evidence that individuals spontaneously and rapidly mimic observed facial expressions, and that this mimicry has been causally related to emotion identification[15]. A meta-analysis of 125 fMRI studies that met a strict criteria[1] for the MNS identified 14 separate clusters that were significantly activated during both action-observation and action-execution tasks (see Appendix H for specific regions)[9]. These clusters included human analogues of regions originally found to show mirroring properties in macaque monkeys, such as the inferior parietal lobule (IPL), inferior frontal gyrus (IFG) and the ventral premotor cortex. Breaking the meta-analysis down into separate domains showed that distinct subregions of the MNS were activated depending on the modality of the task (eg. emotional, somatosensory, auditory). The meta-analysis also showed that additional areas are recruited during tasks that engage non-motor functions to create modality specific subnetworks of the MNS (see Figure 1)[9].

---

[1] Studies were excluded if they: did not attribute results to specifically the "mirror system" (ie. the action observation system, or mirrored movements); employed techniques other than fMRI; and did not report coordinates of the activation clusters[9]

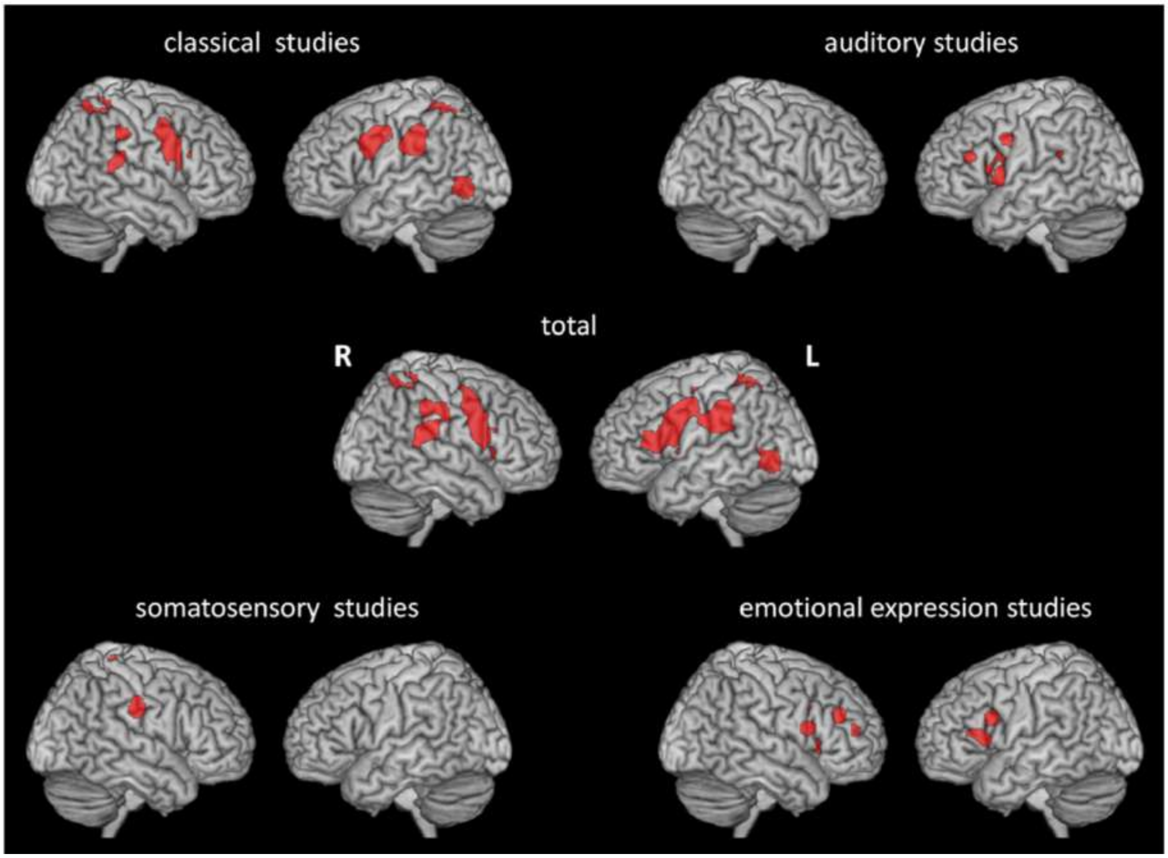


Figure 1: Precise regions activated during mirror neuron activity are dependent on the modality of the task[9]. Activated regions are shown in red on the cortical maps.

Studies investigating the MNS also show a distributed function whereby frontal regions tend to code for goals and parietal regions code for understanding action intent[12]. The inferior frontal gyrus (IFG) is recruited in understanding motor kinematic features of a movement alongside the STS and MTG[16]. The inferior parietal lobule (IPL) is thought to encode action outcome[16] and the anterior intraparietal sulcus (aIPS) is sensitive to action goals or intent[11,17]. Both the IPL and aIPS are independent of the kinematics used to arrive at that goal[11,15,17]. However, the idea that the mirror system provides a basis for emotion understanding and other domains of social cognition is undermined by studies which show that when participants are explicitly asked to make judgments regarding the internal states of others (such as their beliefs, preferences or emotional states) a reliably different set of cortical brain regions, known as the mentalizing system, is recruited[15].

1.2 Mentalizing System

While the MNS activated primarily in the presence of biological motion, mentalizing processes can be recruited during more abstract processing of intent in the absence of any biological motion[18]. Mentalizing is defined as the process of attributing mental states (such as beliefs, desires, and intentions) to another person[11]. The mentalizing network (MZN) concerns regions that are reliably activated when inferring intentions: the temporoparietal junction (TPJ), medial prefrontal cortex (mPFC) and Posterior Cingulate Cortex (PCC) [12]. The TPJ is plays an important role in attributing external agency[12], and is thought to identify when internal cognitive processes are concerned with reasoning about another person as opposed to oneself[5]. There are also studies supporting an induced "out-of-body experience" when the TPJ is stimulated, and it is the only region in the MZN that is more strongly activated during other processing than self-referential processing[5]. The mPFC however, is thought to be involved in interpreting the rationality and intention behind actions[19]. MZN regions also correspond with the default mode network[12], which is implicated in self-referential processes[20]. This indicates that the mentalizing system plays a role in thinking about the self and other.

One study found that different mentalizing tasks recruit different subregions of the MZN[5]. Inferring someone's emotions activated different subregions of the MZN than inferring intention[5]. For example, tasks that were emotions based vs. intention based activated different subregions of the left and right TPJ. In addition, these subregions in showed task dependent patterns of connectivity to other mentalizing areas. For example, an anterior region of the TPJ showed a stronger functional connectivity with the ventromedial PFC for 'emotion' mentalizing[5]. In contrast, the TPJ did not show a strong functional connectivity with the ventromedial PFC in the 'intent' mentalizing condition[5] (see Figure 2).

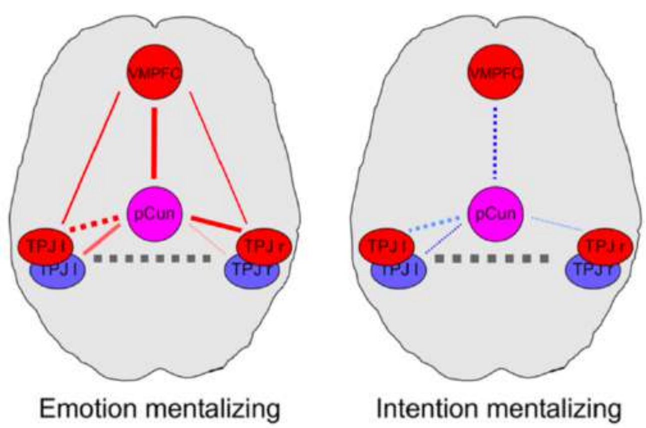


Figure 2: Results from a study comparing different mentalizing tasks (emotion vs intention). The lines show the average correlation of activity between regions over time. Red lines represent correlation during emotion mentalizing; blue lines represent correlation during intention mentalizing. Line thickness indicates the strength of the correlations[5].

To date, the task used most often to study theory of mind or mentalizing is the "false-belief localizer". This task requires participants to make predictions about a character's future behavior based on narratives provided about them. Studies that explore mentalizing through other means[2] have shown that these tasks modulate distinct neural systems, thought to be subregions of the mentalizing system[6]. It is not surprising that a process as complex as mentalizing would rely on multiple distinct processes, with unique methodologies required to investigate them. However, there is a need to add more standardized tasks to the literature in order to help define anatomical delineations of the MZN.

1.3 Interplay Between Two Systems

The MNS plays a strong role in interpreting motion, actions and immediate goals. It is selective to immediate goals but is limited to familiar or frequently executed actions. The MZN system is a higher level cognitive process that enables one to interpret task goals/intent by drawing on previous knowledge, or "social intelligence"[17]. Previously, it was thought that the MNS both informs and supports the MZN, implying that it plays a foundational role in mentalizing functions[17].

Although both the MNS and MZN serve similar functions in making sense of social interactions, a meta-analysis of more than 200 fMRI studies confirmed that both functionally and anatomically they are in fact two distinct systems[18]. This calls into question the idea that the mirror system is the primary basis for understanding mental states of others.

---

[2] For example, the task we use the "how/why yes/no" was tested with a false belief task and shown to activate different regions of the MZN[6].

To distinguish between the functions of the two systems, we differentiate between roles they play according to Table 1. The MNS is thus involved in understand the physical properties of an action as well as immediate goals, whereas the MZN is activated when interpreting higher level goals and intent.

Table 1: Roles of the MNS and MZN.

Role	Definition	Example
Motion (MNS)	Description of the motion sequence.	Moving the leg up.
Action (MNS)	Combination of the motion with the object it interacts with.	Kicking a soccer ball.
Immediate Goals (MNS)	Lower level goals reflect an immediate understanding of the action.	Pass the ball to another player.
Task Goals/Intention (MZN)	Higher level goals that involve longer perspectives and consider the "why" of the action.	Win the game by successfully getting the ball in the net as a team.

How do we access the mental states of others?

Successfully understanding another's actions and intentions is at the heart of almost any human endeavor. Two highly developed social cognitive accounts to explore how this is done are known as embodied simulation and theory of mind[21,22]. Embodied simulation is based on the idea that emotional states are associated with motor behaviours (for example, sadness often leads to frowning or crying)[15]. The MNS allows one to map another's actions onto an internal neural system, so it is often linked to theoretical investigations of embodied simulation[23]. "Theory of mind" on the other hand describes the ability to understand the internal mental states of another individual, and maps biologically to the MZN[11]. In contrast to the MNS, the MZN is also recruited in the absence of detailed information of another's actions[17]. Although there are various differences between the two, the MNS and MZN may work together to generate internal simulations that allow us to experience another person's state in on our own bodies. In order to explore how we can understand the mental states of others, we explore a few different questions.

1) Do intentions modulate action kinematics? Actions typically have unique kinematic profiles corresponding to the underlying intent[24]. Moreover, Several studies where individuals were asked to perform a variety of actions on the same object with different intents have shown that the kinematic profiles differ significantly[24–26]. For example, participant's actions are significantly different when they are asked to place an object in a small container vs. a large box[27].

2) How do we acquire intentions from actions? Embodied simulation accounts suggest that action kinematics result in a 'direct' perception that allows us to immediately attribute intent (MNS) whereas "Theory of mind" (MZN) view posits that the intentions are understood from actions via inferential processes.

3) Do individuals rely on action kinematics to assess intention? Whether or not individuals rely on subtle differences in kinematic profiles to make predictions about intent is not clear cut, as there are conflicting results on the subject. Most studies show that observers can use kinematic differences to predict intent[24]. However, one study shows that although there were clear kinematic differences between intent, participants were unable to reliably guess the intent[28]. It is not clear whether or not failures to predict intent are based on an inability to perceive kinematic differences, or on correctly associating intentions with kinematic profiles. Moreover, it is important to note that there is a need for more ecologically valid stimuli to investigate this question[24].

4) Are the MNS and MZN activated spontaneously?

Living in a social world requires that we process countless complex stimuli at a given instant. We need to recognize people, interpret behaviours, guess emotions, recall memories and interact, all while constantly updating our understanding with incoming information. Much of these processes happen automatically in the presence of social stimuli, however some require conscious control. The MNS is thought to allow for automatic comprehension of actions based on sensory information alone. However, reports are more conflicted about whether or not mentalizing relies on controlled or automatic processes[29]. Many studies show that it is also active in the absence of any instructions to make mental state attributions. For example, a few studies have shown that the MZN is activated more strongly in irrational actions relative to rational ones[19]. This shows that actions that are difficult to interpret can engage the MZN more strongly, even without any instructions to mentalize. Additional studies have shown that individuals make spontaneous inferences about mental states without instruction, and that MZN activity is strongly associated with those inferences[19].

On the other hand, many studies have shown that MZN activity is most often seen when participants are explicitly asked to infer mental states[29]. One study explored the automaticity of these processes by measuring the extent to which the MNS or MZN demand attentional resources. The authors found that increasing cognitive load interfered with mentalizing abilities[30]. However, there was no modulation of mirror function with cognitive load[30]. This lends evidence to the idea that the mirroring system supports automatic action understanding whereas the mentalizing system supports controlled causal attribution. It is also possible that activity in mentalizing areas is disrupted with the introduction of a second task and is deactivated with increasing task difficulty. There are also individual differences in the level of conscious control required for mentalizing abilities. For example, people who are more inclined to adopt inferences about mental states show increased spontaneous activation of the mentalizing system[31,32].

Studies have also shown that individuals with Asperger syndrome, or high function autism, can understand peoples mental states when explicitly prompted to do so, but fail to do so spontaneously[33]. Understanding whether MZN and MNS are automatic or controlled may help elucidate potential mechanisms at play in autism spectrum disorder (ASD) and leave room for another potential area of study.

What is the time course for understanding intent?

Perceptual discrimination studies where individuals are asked to differentiate between sensory features of stimulus typically find that response times range from 250-300 ms[24,34]. However, studies that ask subjects to guess the intent of an action typically show significantly longer response times and a wide degree of variability, ranging from 600 ms[35] to 1500 ms[36]. One study showed that repeating the stimuli resulted in shorter response times (330-400 ms), however this may be related to perceptual processing instead of intent inferences[35].

5) How do the MNS and MZN inform each other?

A bottom-up visuomotor processing of intent would indicate feed forward information from the MNS. A top-down attribution of mental states could involve internal models from the MZN being communicated to the MNS to help minimize prediction errors. One fMRI study used dynamic causal modelling to determine the direction of connectivity patterns underlying tasks that either had shared action goals or shared mental states – see Figure 3[37]. They found that cooperative tasks had more bottom-up processing with higher forward connectivity from key MNS to MZN regions. In comparison, affective tasks showed stronger top-down processing from the vmPFC to the pSTS[37].

Figure 3: Stimuli used in an fMRI study to investigate an interaction between the MNS and MZN. Stimuli on the left represents cooperative interactions (shared action goals) thought to rely on the MNS, and stimuli on the right represents affective interactions (shared mental states) thought to depend on mentalizing functions[37].

Although little is known about the interaction of the MNS and MZN in reciprocal interactions, one study explored this in the context of imitative exchanges with fMRI[12]. They found a strong coupling between the two systems during the exchanges. They showed that the MNS was recruited in the preparation of ones' own actions and simulation of another's, while the MZN was engaged in the anticipation of the other's intentions[12]. In addition, numerous studies have shown that, MNS activity is significantly higher during mentalizing tasks[38], implying that the two systems (MNS and MZN) may work concurrently and inform one another. However some studies have failed to see this effect, leading to inconsistencies in understanding the interaction between the two systems, if there is one[38].

In summary, both the mentalizing and mirroring systems participate in the anticipation and understanding of other's behaviors, but are likely involved in different levels of representing mental states[12] (see Figure 4). The first stage of intent perception is likely implicated in the perception of action kinematics by the MNS[39]. The

MNS maps visual or auditory information onto the motor representation of our own actions. This can provide a low-level awareness of the goals. The MZN is recruited in the second stage and relies on inferential processes to fill the missing information.

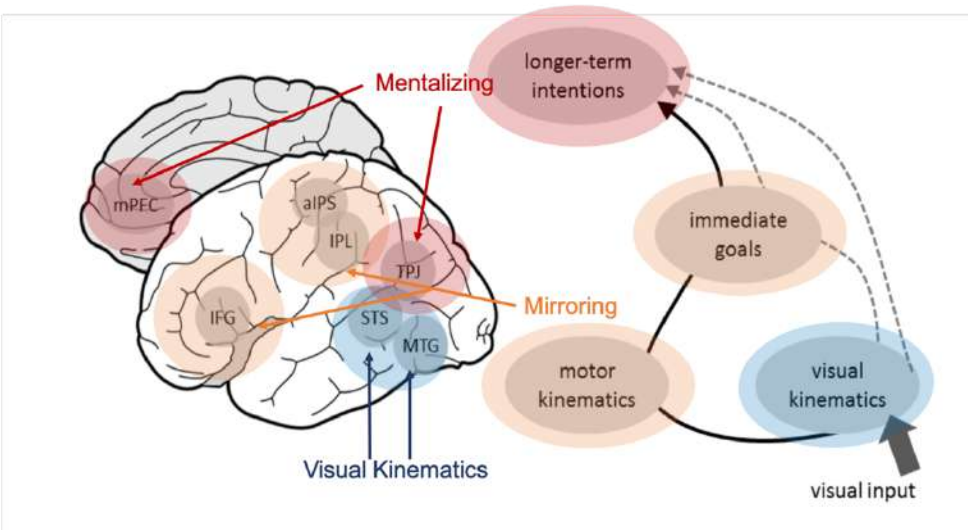


Figure 4: Comparison between the MNS and MZN (adapted from Hamilton and Marsh, 2013)[11].

1.4 Implications

Autism Spectrum Disorder and MNS

Autism Spectrum Disorder (ASD) is a term used to describe a wide range of neurological disorders that are primarily characterized by impairments in verbal and non-verbal communication[38,40]. Evidence shows that the ability to imitate, largely controlled by the MNS, is compromised in individuals with ASD[18]. This has given rise to the 'broken mirror' theory of ASD[11], that atypical MNS functioning underlies many of the social difficulties experienced by individuals with ASD[38]. In an attempt to investigate this theory, fMRI studies have shown that there is reduced MNS activity in individuals with autism relative to controls while watching a set of goal directed actions[38]. One EEG study measured mu suppression in different areas to show that individuals with autism have lower levels of MNS activity in the right hemisphere when inferring action intent[38]. However, this activity was not correlated with mentalizing performances in the task. On the other hand, left hemispheric MNS activity was correlated with mentalizing performance in the task, but did not appear to be different in individuals with autism[38]. These results show that the mirror system may be implicated in symptoms of ASD, but the results are not clear cut enough to make a direct causal statement.

Autism Spectrum Disorder and MZN

Theory of mind abilities are developed around the second year of life and account for the emergence of pretend play[40]. It is well established that children with ASD show a marked lack of pretend play[40]. Several studies have attempted to investigate if deficiencies in the development of theory of mind are related to possible mental disability or general intellectual level. In one study, children with autism, Down's syndrome, and neurotypical children were tested on a false belief task[40]. A description of the task is demonstrated in Figure 5. Results show that regardless of verbal and non-verbal age, autistic individuals tended to struggle with the task (80% failure rate). On the other hand, neurotypical pre-schoolers and individuals with Down's syndrome, who had a much lower verbal and non-verbal age, did not. This was one of the first explorations into localizing the specific deficit manifested by ASD. The "mind blindness theory of autism" posits that the primary deficits in ASD are a result of impairments in 'high level' reasoning about mental states, which the MZN is thought to be responsible for[11,18].

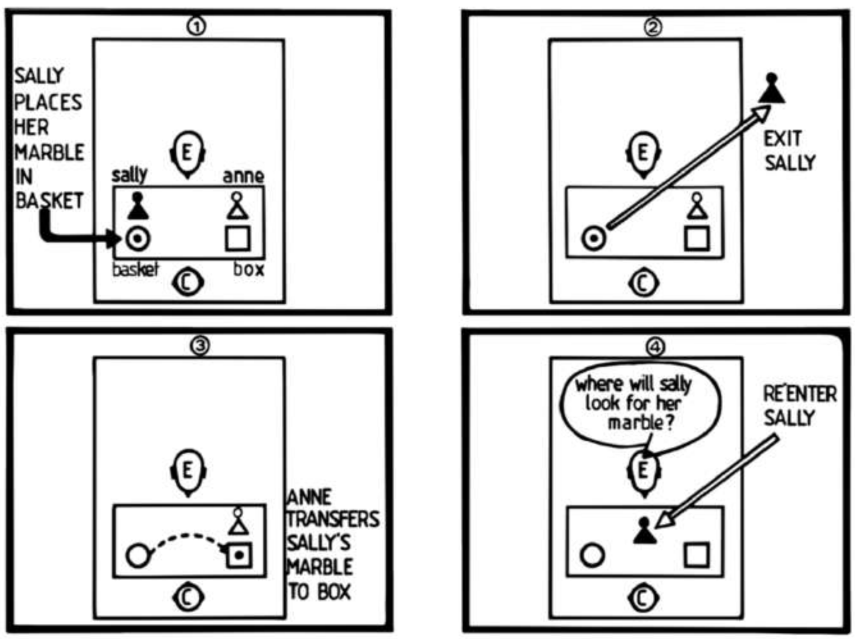


Figure 5: Each image from 1-4 above outlines the false belief task used to study the mentalizing system: (1) Participants are told a story in which 2 protagonists have a container each. One of the protagonists, Sally places her marbles in her basket. (1) This is followed by her exiting the room, during which time – unbeknownst to her – the other protagonist, (3) Anne, transfers Sally's marbles into her own box. (4) Sally returns to the room and the participants are then asked 'where will Sally look for her marble?' [40].

Impaired cross talk between the MNS and MZN in Autism:

More recent studies have indicated that atypical network connectivity may underly problems in social cognition[18]. One study using resting state fMRI in children and youth between 11 to18 years found that increased connectivity between the MNS and MZN areas during resting state[18] was positively correlated with social impairment[18]. Other reviews have shown that the relationship between functional connectivity and ASD is age dependent[41]. For example, while children with ASD typically have higher levels of overall functional connectivity than neurotypicals, this relationship is switched in adolescence and adulthood[41]. Cole et al., (2018) showed that there is reduced functional connectivity between mirroring and mentalizing areas when inferring intent for adults with autism relative to neurotypical adults[42]. They also found that the degree of connectivity was correlated with the level of autistic traits[42]. Another fMRI study corroborates these findings and shows that connectivity patterns specifically between the MNS and MZN are altered in individuals with ASD[43]. These results, along with genetic models and other clinical observations have contributed to a theory that ASD symptoms may arise from an increase in excitatory and decrease in inhibitory functional brain activity[44].

1.5 Limitations of Previous Studies

Both mirroring and mentalizing processes have been studied extensively in humans. However, there are two major limitations of previous studies that we aim to address. Firstly, there is a dearth of standardized tasks used to study the mentalizing system. Secondly, studies that aim to compare the mentalizing and mirroring network often fail to keep stimuli or other confounding task details consistent between conditions. We explore each limitation below.

Limitation 1: There is an over reliance on the "false belief task" to study theory of mind in literature. Other than this task, there is enormous variability in how theory of mind is operationally defined, and what regions are involved. When large meta-analyses are conducted, they often fail to find a precise anatomical definition of the MZN, and when they do it is often explained by the fact that the regions within the networks are themselves anatomically imprecise. For example, the medial prefrontal cortex and temporoparietal junction can be used to refer to large areas of the cortex that have both functional and structural heterogeneity. When the false belief task was compared to the photo-judgement task that we use, a completely different underlying network was found to be elicited[6]. Given that mentalizing is a broad ability which requires the use of a wide range of mental representations to understand diverse stimuli in the context of a variety of goals, it is not surprising that there would be significant differences between tasks. However, if we accept that there are subnetworks in the MZN, it is possible that only one of those may be implicated in disordered states, meaning that it is critical to explore and standardize a variety of tasks. The photo-judgement task we use has been tested in fMRI and validated in two different studies. However, there is a need to explore task responses using different neuroimaging methods and we choose to do this with EEG for reasons outlined in the next section.

Limitation 2: Most studies that investigate differences between the mirroring or mentalizing systems utilize independent and distinct tasks, subjects or separate studies altogether. A feature of the task we use that addresses this limitation is that the stimuli and participants are consistent between the mentalizing and mirroring conditions, which increases the ability to make definitive statements about the differences between the two. Additional properties of the photo judgement task that we use, and their implications are discussed in Table 2 below:

Table 2: Photo Judgement Task Properties

Photo Judgement Task	
Property	Description
Is flexible	Questions are varied in terms of the stimuli being tested (facial expressions or hand movements) and in the questions that are asked (ie. intent vs means).
Constrains responses	Having Yes/No questions allows for accuracy and response-time measures that can be compared across subjects and different studies.
Permits use of diverse naturalistic stimuli	The task makes use of a variety of photos that represent actions or emotions that individuals may be exposed to in their daily lives. Increasing the variety of photos limits concerns about measuring stimuli specific responses such as neural responses to low-level visual properties, proportion of particular objects shown, or emotional meaning.

Has discriminant validity[6]	Previous research shows that the task has little overlap with other tasks used to study the MZN. It has spatially distributed activity patterns from the False-Belief Localizer[6]. This shows that the two tasks provide complementary methods to study different uses of the MZN with different behavioural outcomes and distinct subregions involved.
Has convergent validity[6]	The Yes/No version of the photo-judgement task has been shown to have significant overlap with another task that was more open ended and activates the same network[6]. This provides evidence in support of the reproducibility of the task.
Allows for Why-How comparisons	This task keeps the stimulus and instructions consistent between conditions, allowing for a more accurate comparison between the how and why conditions.

1.6 Neuronal Basis of EEG

When several neurons receive or generate the same repeated sequence of activity, a synchronized electric field potential is induced that propagates through the brain and skull[45]. Electroencephalography (EEG) measures the difference in voltage between two different cerebral locations through time. The electric field potential drops off with increasing distance from the source, so most signals picked up by EEGs at the scalp represent both strong and synchronized activity inside the brain[45]. For this study we selected high density EEG (hdEEG) to record brain activity for several reasons that are outlined below.

Why use hdEEG?

1. Captures cognitive dynamics in the time frame in which cognition occurs: Cognitive, perceptual, linguistic, emotional, and motor processes are fast. Most cognitive processes occur within tens to hundreds of milliseconds, and studies have demonstrated that mentalizing and mirroring processes also occur at these time scales. EEGs are well suited to capture these fast, and temporally sequenced cognitive events. The temporal precision of the hemodynamic response is 2-3 orders of magnitude slower than that of the electrophysiological response[46].
2. Directly measures neural activity: Oscillations observed in EEG are direct reflections (plus noise) of neural oscillations in the cortex as opposed to an indirect method such as hemodynamic response[46].
3. EEG signals have multidimensional information: EEGs provide information on power, time, frequency, space and phase. Power and phase provide largely independent information[46].
4. High Density EEGs (hdEEGs) increases specificity and sensitivity of results: hdEEGs (we use 256 electrodes) can provide additional information that is useful in measuring effects. A study on epileptic patients found that high hdEEGs yielded much higher specificity and sensitivity relative to low-density EEG recordings, structural MRI, or PET exams[47,48].

5. Affordability: EEGs are significantly more affordable than fMRI, PET scans, MEG and other neuroimaging methods[49]. This allows for an increased number of participants to be scheduled.
6. Applicability and Practicality: Because EEGs require less constraints than fMRI, it allows for participants to be engaged in more naturalistic tasks during data collection. This is important because the it increases ability of the study to have real world implications.

1.7 EEG markers of the MNS and MZN

Event related potentials (ERPs) represent electrical activity in response to specific events and are thought to reflect information processing associated with the stimuli[61]. We divide ERPs into 3 categories of interest. The early components peak within the first 100-170 milliseconds after stimulus onset and are typically more sensory or 'exogenous' components that largely depend on physical parameters of the stimuli[61]. The middle components (200-300 ms) are typically involved in processing lower level goal inferences and semantic incongruity. The last components (>350 ms) are more cognitive or 'endogenous' ERPs as they reflect information or intent processing. We expected that mentalizing activity would be preferentially activated during the late ERP components. Some early to late ERP components that may be implicated in the photo judgement task along with their associated functions are outlined in table 3.

Table 3: ERP components of interest for our study alongside their function and typical location that they are associated with.

Component	Deflection	Time (ms)	Locations	Function
Early: **N**170	**N**egative	150–170	Lateral occipito-temporal electrodes	Stage of visual processing at which objects are categorized, specific to face stimuli[92,93]
Middle: **P**3a	**P**ositive	251-350	Frontal	Related to expectancy [77,94–96]
Middle: **P**3b	**P**ositive	350-450	Centro-parietal area	Related to memory and expectency[94,95]
Middle: **N**300	**N**egative	250-350		In the context of semantic congruity and expectancy.[61,97–99]
Late: **P**400 (LPP)	**P**ositive	350-600	Frontal, or central & parietal	Activity is different in autism[14,100–102]. Involved in attention and "evaluative incongruence"
Late: **N**400	**N**egative	350-550	Central	Involved in semantic incongruity and processing others mental states[98–100,102–104].

Previous studies have shown that high-arousal photos evoke larger N170 amplitudes than low-arousal ones[85]. These early components may be indicative of mirroring activity and reflect stimulus processing, with different strengths depending on the emotional valence of the image. A couple studies on trait inferences showed that information about personality traits and intent occur in late ERP components, around 400+ ms[105,106]. As mirroring activity is thought to be reflected by early to mid-ERP components, this implies that mentalizing begins later on.

Another EEG measure that has been previously used to investigate MNS and MZN activity is time frequency analysis, which shows the power contribution of different frequencies across time. Although there is a plethora of research on mirroring activity marked by mu (alpha) suppression, there are very few linking mentalizing function to time frequency analysis. The few mentalizing studies that do report a modulation of time-frequency have found changes in beta power following a mentalizing task[50–52]. Moreover, ASD, schizophrenia and frontotemporal dementia, which are all characterized by significant social deficits, show alterations in alpha and beta oscillations[50,121,122].

Although EEGs are limited in the spatial domain, with the use of high density EEGs and an appropriate selection of source localization methods, it is possible to estimate the source locations and strengths inside the brain[78]. There is, however, very limited research done to compare MNS to MZN sources using EEG, so our study aims to corroborate previous fMRI studies that have identified the IFG, IPL & aIPS with the MNS; and the TPJ & mPFC with the MZN.

2 Aims and Hypotheses

Aim 1: Do we detect reliable differences in event related potentials (ERPs) between mirroring and mentalizing conditions that align with previous results?

Hypothesis 1 We expect that the why and how conditions will diverge in late ERP components (400 ms onwards).

Aim 2: How are mirroring and mentalizing systems modulated by photos of faces vs. hands?

Hypothesis 2: We expect to see a stimuli specific difference in early ERP components and time frequency analysis such that we can differentiate between faces and hands at this stage. We also expect that there may be an interaction between MNS and MZN activity with the type of stimuli used.

Aim 3: What particular frequencies may be implicated in the differences we see in ERPs? Can mu and beta suppression provide some insight in the contributions of the MNS and MZN respectively for the *how* and *why* conditions?

Hypothesis 3: We expect that mu suppression will be stronger in the MNS task while beta suppression will be higher in the MZN task.

Aim 4: How do brain-states transition in mentalizing vs. mirroring conditions? How many stable states are achieved in each task and what are the cortical sources of each stable state?

Hypothesis 4: We expect that there will be more state transitions in the MZN task within the late ERP components, reflecting the involvement of different brain areas involved in interpreting intent. We expect that mentalizing tasks will be associated with greater mPFC and TPJ activity, while the mirroring task IFG, IPL & aIPS activity.

3 Methods

3.1 Data collection

Neurotypical adults (age 19 - 35) with no history of mental illness were recruited to participate in the study from UBC and neighboring areas. Ethics approval was obtained from the Behavioural Research Ethics Board at UBC. Participants were excluded from participating if they had a history of neurological disorders, such as mild traumatic brain injury with concussions, leading to unconsciousness for more than five minutes or been diagnosed with neurodevelopmental disorders (i.e. autism spectrum disorder, attention deficit hyperactivity disorder, etc.). After participants signed consent forms, high density HydroCel EEG caps from Electrical Geodesic were used to measure EEG patterns in response to two different tasks, both of which were designed to induce mirroring and mentalizing activity. The participants then completed two different tasks; a photo judgement task and a naturalistic video task. Only the photo judgement task is analysed in this study.

Photo Judgment Task: For this task, participants were asked to attend to and make judgements about the means of or intent behind a variety of photos depicting some action or emotion. Intermittent breaks approximately once every ten minutes were built into the program to give participants some time to rest and to check electrode impedances. The task was adapted from Adolf and Spunt's theory of mind Why/How protocol, and we refer to it as the Photo Judgement Task[6]. For *how* trials, participants were asked to respond to yes/no questions about how an action was being performed, given pictures of a person's facial expression or hands. Using the same pictures, for *why* trials, participants were asked to respond to yes/no questions about their thoughts or intentions (see Figure 6).

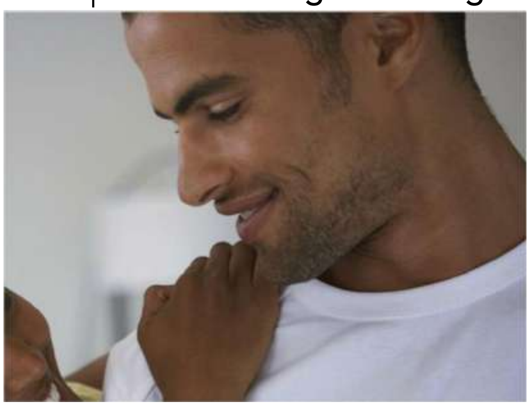
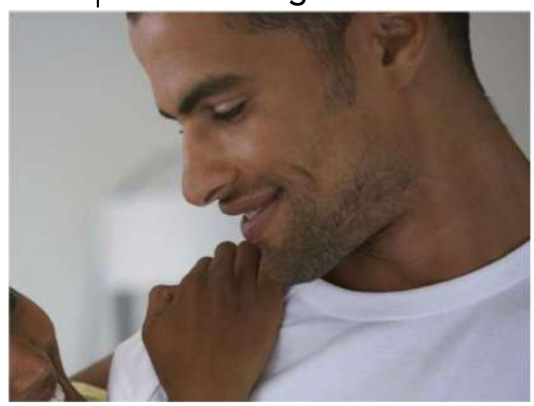
	WHY (Mentalizing)	HOW (Mirroring)
FACE	Is the person **admiring something?** 	Is the person **looking to their side?** 
HAND	Is the person **helping someone?** 	Is the person **using both hands?** 

Figure 6: The factorial design for the photo-judgement task that we use. The How/Why contrast can be used to identify mirroring/mentalizing areas, while the face/hands contrast can be used to identify stimulus specific effects.

For all trials, participants were asked to silently think of a response and use a button press to indicate their selection (see Figure 7 and Figure 8). The order of the trials and pictures was randomized between participants. The MATLAB Psychophysics Toolbox was used to present the stimuli to participants and to record their responses. Each block begins with question presentation and is followed by a set of photographs paired with that question (See Figure 8). Between each photograph is a brief reminder of the question for that block. For each photograph, participants have 1700 ms to respond. If they fail to respond by that time, the task advances. Responding before the end of the 1700 ms ends the trial and advances to the next trial. Hence, block durations were contingent on response times. Figure 9 presents all the questions presented for each condition.

Figure 7: Examples of four blocks for the photo judgement task created by pairing either a question about intention (why) or means (how) with a set of photographs featuring either hands or faces.

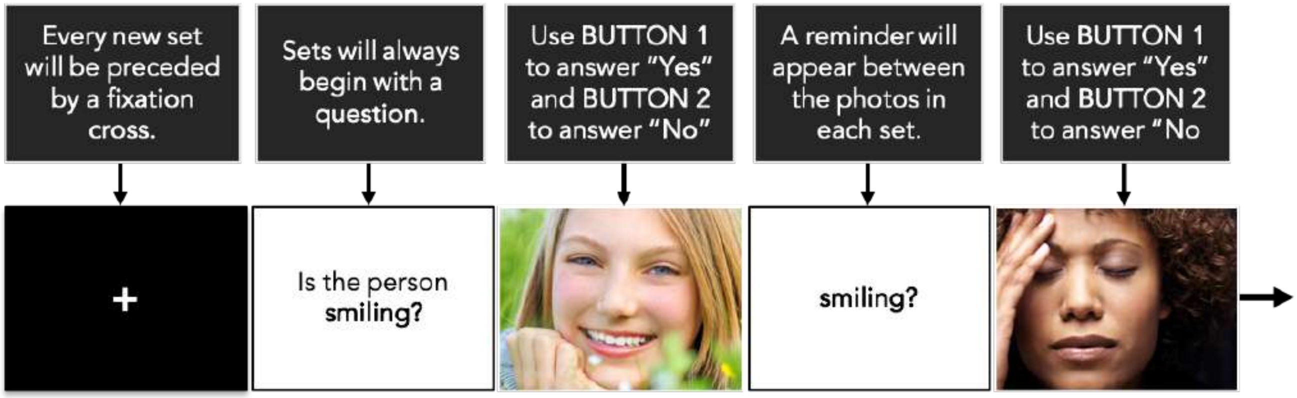


Figure 8: Instructions that explain the steps involved in the photo judgement task.

Question	Reminder Cue	Stimulus	Question
Is the person admiring someone?	admiring?	Face	Why
Is the person expressing self-doubt?	self-doubt?	Face	Why
Is the person in an argument?	argument?	Face	Why
Is the person proud of themselves?	proud?	Face	Why
Is the person competing against others?	competing?	Hands	Why
Is the person concerned with their health?	healthy?	Hands	Why
Is the person helping someone?	helping?	Hands	Why
Is the person protecting themselves?	self-protection?	Hands	Why
Is the person looking at the camera?	looking at camera?	Face	How
Is the person looking to their side?	looking to side?	Face	How
Is the person opening their mouth?	open mouth?	Face	How
Is the person smiling?	smiling?	Face	How
Is the person lifting something?	lifting?	Hands	How
Is the person pressing a button?	pressing button?	Hands	How
Is the person reaching for something?	reaching?	Hands	How
Is the person using both hands?	both hands?	Hands	How

Figure 9: A list of all the questions asked in the photo judgement task. Each question appears at the start of a block and corresponding reminder cues appear between each image presentation.

3.2 Analysis

3.2.1 Preprocessing

EEG data is collected in NetStation and imported into the matlab toolbox, *EEGlab*, for preprocessing. The steps followed are outlined in Figure and discussed in more detail below.

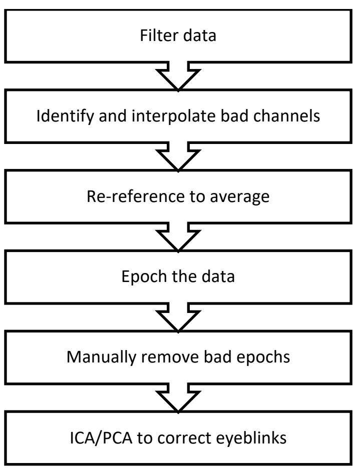


Figure 10: Preprocessing Steps

Filtering

EEG signals are composed of multiple frequencies. Some frequencies are not physiologically meaningful and removing those can greatly increase the signal to noise ratio. Based on recommendations in Luck (2014)[53], the band pass filter range selected was between 0.05 Hz and 30 Hz. In addition, because the alternating currents from electrical outlets in North America oscillate at 60 Hz, a notch filter was applied to further attenuate this frequency from the data. Further specifications of the filters selected are specified in Appendix A

.

Identify and Interpolate Bad Channels

With high density EEG, there are often several electrodes that are either unable to connect strongly to the scalp or are contaminated by artifactual noise during data collection. We identify these electrodes by using an estimate implemented in the "FASTER" algorithm. FASTER stands for Fully Automated Statistical Thresholding for EEG artifact Rejection and uses the variance, amplitude range, median gradient and channel deviation to threshold the data[54]. We selected a Z-score threshold of greater than 3 for these properties in order to select noisy channels. The selected 'bad' channels were removed with spherical interpolation[55]. This algorithm was validated by comparing it to bad electrodes identified manually on 5 subjects. All the electrodes identified by FASTER were also removed in the manual attempt, however, there were additional noisy channels that were not picked up by FASTER. This is expected because the threshold we set is quite conservative, and because automatic thresholding uses correlations to get rid of the data (ie. if 'bad' electrodes are correlated with each other they will not be picked up). After channels are interpolated using automatic thresholding, each dataset is inspected visually for additional noisy channels that may have been missed. These channels are then also interpolated using spherical interpolation. In cases where more than 10% of channels were removed (<25 channels), the subject was removed from the data.

Re-referencing to average

It is critical to set a reference in EEG, as in nature only the differences between two potentials can be measured [56,57]. Selecting an appropriate reference against which to compare all the electrodes depends on 3 critical considerations:

1) The choice of reference should not affect the source reconstruction
2) The position of the reference electrode should not be close to an area where we expect our main effects to be (We expected task related activity in Cz for ERPs and in the prefrontal cortex for source localization).
3) Using a reference in either hemisphere would introduce an undesired laterality bias in the data.

For these reasons we chose to re-reference the data to the average of all electrodes [56,58].

Epoch the data

Events were added in the EEG data using behavioral output files. The EEG data was then clipped around the events of interest (500 ms before and 1000 ms after). Epoching was done using the ERPlab and EEGlab toolbox in matlab.

Manually Remove bad epochs

The clipped data was manually inspected for bad epochs, which were marked and removed. If more than half of the epochs were removed in any condition, the data for that participant was removed from EEG analysis.

ICA/PCA to get rid of eyeblinks

Independent Component Analysis (ICA) is an effective way to separate EEG data into neural activity and artifact because it isolates the data into components that are unique. With high density EEG, it is important to use PCA combined with ICA to reduce the number of components (see Appendix B for more details) [59,60]. Four factors were considered when identifying which components were a result of artifacts:

1) The scalp distribution of the component
2) Inspecting the original EEG signal overlaid on the recovered signal after the selected component is removed
3) Looking at the time locked activity of the component per trial
4) The power spectrum of the component

This is discussed in more detail in Appendix B. Artifacts removed included eye blinks, horizontal eye movements, EKG and button press artifacts.

3.2.2 ERP analysis

Event related potentials (ERPs) measure electrical activity in response to specific events and are thought to reflect the sum of post-synaptic potentials produced when several pyramidal cortical neurons fire in synchrony while processing information[61]. ERPs were constructed by averaging electrical activity across all subjects and trials per condition using ERPlab. Statistical analysis was performed via t-tests on the peak amplitude for each ERP component in regions of interest. Details on how the peak amplitude was calculated are in Appendix A. We also used a cluster permutation analysis to confirm that the conditions were indeed different in ERPs. This method is described below.

3.2.3 Cluster Permutation Analysis

With electrical information measured at 1000s of time points for 256 sensors, hdEEG experiments typically produce very high dimensional data. Whenever there is analyses with higher dimensions than the subject pool, the results may fail to fit additional data or predict future observations reliably. In other words, when numerous individual tests with the $p < 0.05$ threshold are conducted, the actual error rate greatly exceeds the nominal rate (5%)[62]. Correction for multiple comparisons must be applied, however many of these methods reduce power and curtail the likelihood of revealing a true effect – if there is one. Although increasing the subject pool to an appropriate level is not a feasible option, there are a few ways in which statistics can be performed on EEG data in a valid way [62,63]. One way to do this is to have certain electrodes or time points of interest set *a priori*. Another approach is to group certain electrodes and time points *a priori* and determine which parameters are significant with a correction that takes into account the smaller number of tests that are done. However, if a more exploratory approach is required, cluster permutation analyses can address the multiple comparisons problem[64,65].

Cluster permutation relies on the assumption that true effects should be clustered in both time and space[62] and has two major components:

1. One component is the cluster-forming algorithm, which reduces the high dimensional data into smaller units based on spatio-temporal clustering[66].
2. The other requires a null hypothesis, against which the observed data is compared to obtain p-values using permutation tests. Performing a full permutation test would be computationally intractable. However, a special class of approximations, Monte-Carlo sampling, can be done and yield satisfactory results[65]. The Monte-Carlo simulation repeats random sampling to better determine the underlying correlation.

It is important to recognize that because p-values are determined from cluster level statistics, the p-value of a cluster does not necessarily represent that of a single member of that cluster. Thus, cluster based statistics only provide weak [family-wise] error rate control[67], and other statistics would need to be performed to determine more precise location and timing of the effect. Cluster permutation was performed in the FieldTrip and specific steps are discussed in Appendix C.

3.2.4 Time Frequency Analysis

Previous studies have shown changes in mu-alpha and beta rhythms in mentalizing & mirroring tasks[50,52,68,69]. Time frequency analysis allows us to visually see how power varies in different frequencies across time. This is typically done relative to the onset of a particular time-locked event[70,71]. EEG signals capture phase-amplitude oscillations in the time domain. Time frequency analysis transforms this data into the frequency domain. With this transformation there is a trade-off between precision in the time domain to precision in the frequency domain formalized by the Heisenberg uncertainty principle[72]. As the time window used to estimate the data increases (reducing temporal resolution), the frequency resolution increases[46,72]. In other words the more you know about *when* some activity occurs, the less you know about *where* (at which frequency) it happens, and vice versa[46]. We selected a time range of 400 ms that optimized for an adequate time and frequency resolution. We used the fieldtrip toolbox in Matlab to get time frequency analysis on the data that had been cleaned in EEGlab. See more specifications in Appendix C. Differences were either processed by doing t-tests on the average mu or beta suppression for each subject, or by cluster permutation analysis as shown above.

3.2.5 Microstates

Beyond understanding *where* mirroring and mentalizing activity takes place, we want to understand *when* and *in what* combinations the activity takes place. With traditional ERPs we can view these peaks and troughs for specific electrodes at a time[73]. Microstates complement this analysis by identifying stable configurations of global activity using the brain's topographic activity. We use a toolbox called CENA within the brainstorm program to perform microstate analysis. Brainstorm is an open-source software written in java/Swing that can be run within or independently of matlab[74].

EEG data can be visualized as a time series of spatial patterns (or maps). A "microstate" refers to a momentary, stable global brain state, and is thought to reflect transient information processing in the brain[39,75]. Microstate algorithms search for a few stable spatial patterns that can capture the maximum amount of variance in the data[76,77]. The goal is to temporally classify neural activity into brain states that can capture most of the states the brain occupies.

We used root mean squared error (RMSE) analysis for identifying/clustering microstates. This analysis does not make use of an *a priori* hypothesis, thereby eliminating confirmatory bias of the experimenter[74]. The RMSE analysis uses noise levels detected during the baseline period to decompose the post stimulus waveform into stable microstates and the transitions between them. We follow up with a cosine similarity metric and global field power (GFP) analysis[74]. These two analyses collectively indicate whether or not the differences in activity between microstates are related to changes in cortical sources or in power respectively[39,73]. Finally, a bootstrapping procedure was conducted to determine whether or not these results are stable across subjects. The parameters we set are as follows: A lag of 10 ms, a baseline period from -500 ms to -1 ms pre-stimulus, a 99% CI to detect significant rises or falls in the RMSE function, and a 95% CI for the cosine metric analysis to determine whether a microstate differs significantly from another. See Appendix G for further specifications.

3.2.6 Source analysis

While EEGs can provide high temporal resolution, they are limited in the spatial domain[78]. The basic goal of all source localization methods with EEG is to use measurements on a scalp to determine dipolar source locations and strengths inside the brain[79]. There are two broad categories of source localization methods: current density estimates and beamformers[79]. Appendix C outlines a few common source localization methods that could be used within each category. We selected minimum norm estimate because it is typically used to localize evoked data[79]. Furthermore it is more robust when compared to other source localization methods without any a-priori information[81].

To perform the source localization, we need a head model that is created using, a template anatomical MRI was used to calculate a head model using the "forward solution”. The anatomical MRI was obtained from the FieldTrip templates folder. The EEG electrode location file was adjusted to fit the head model created and the EEG data was used to construct a noise covariance matrix. The noise covariance matrix, epoched EEG data, and head model were used together to get the source activity using the minimum norm estimate. The source data was then averaged across conditions and subjects and visualized within each stable microstate. See Appendix C for further specifications.

4 Results

4.1 Cleaning Data

Data was collected on 40 participants (mean age = 25 years old) at BC Children’s hospital. There were an equal number of males and females (20 each) in the subject pool. Subjects 26-35 were removed due to noisy data, because either more than 25 channels or 1/3rd of the epochs were removed. The majority of the removed subjects were male (9/10), which reduces our ability to make meaningful conclusions on sex-based analysis with EEG data.

4.2 Behavioural

We explore how reaction time varies across question condition (*how* vs. *why*) and stimulus (hand vs. face) in Figure 11. A two-way ANOVA (see Appendix D) shows that the interaction between reaction times for why/how and face/hand conditions is significant (F(1,4)=10.47, P=0.003). Participants were significantly faster for *how* (mirroring) than *why* (mentalizing) questions for both hands and face stimuli. Participants were also faster in response to face stimuli compared to hands stimuli. P-values from the t-tests comparing between conditions are shown in table 4. This implies that the stimulus type mediates the relationship between reaction times and the question condition.

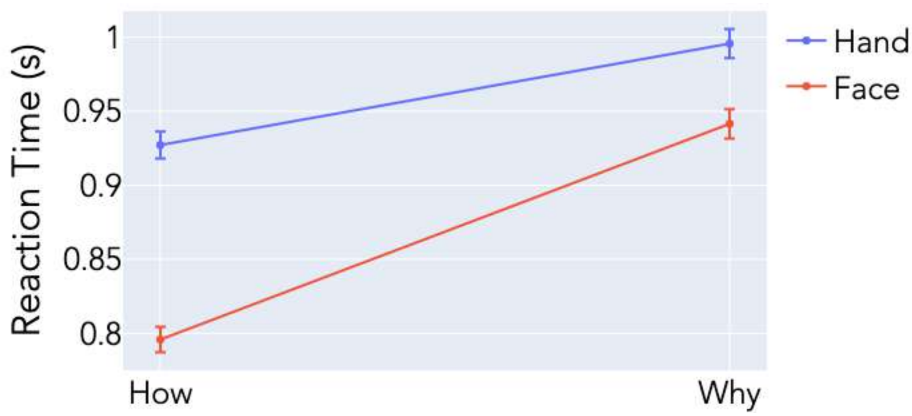


Figure 11: Reaction times across condition.

Table 4: P-Values from T-Tests comparing reaction times across conditions. Significant comparisons are highlighted in yellow.

	Why Hand	Why Face	How Face
How Hand	4.13E-07	0.29	4.78E-25
Why Hand		1.14E-04	2.01E-50
Why Face			6.18E-28

We also investigate how accuracy varies across question condition (*how* vs. *why*) and stimuli (face vs. hand) in Figure 12. Question conditions and stimuli both have a significant effect on accuracy independently. Results from T-Tests comparing task accuracy across conditions are shown in Table 5. However, a 2-way ANOVA (see Appendix D) shows that there is no interaction between the two (F(1,4)= 2.14, P=0.15).

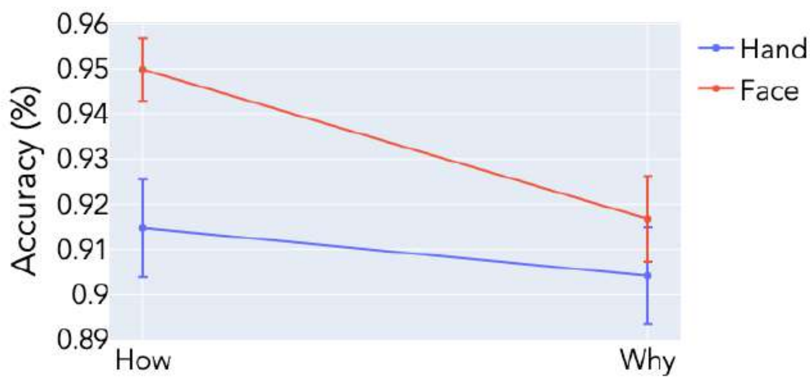


Figure 12: Accuracy for all conditions and t-test results below.

Table 5: P-Values from T-Tests comparing accuracy across conditions. Significant comparisons are highlighted in yellow.

	Why Hand	Why Face	How Face
How Hand	0.49	0.89	7.98E-3
Why Hand		0.38	6.23E-4
Why Face			6.46E-3

4.2.1 Do we see sex-based differences?

Reaction Time: To determine how sex plays a role in reaction times we do a separate analysis for each condition shown in Figure 13. There is a significant effect of sex on reaction time ($p<0.0125$), however a factorial ANOVA shows that there is no interaction between sex and the *how* or *why* condition on reaction time ($F(1,4)=0.043$, $P=0.54$).

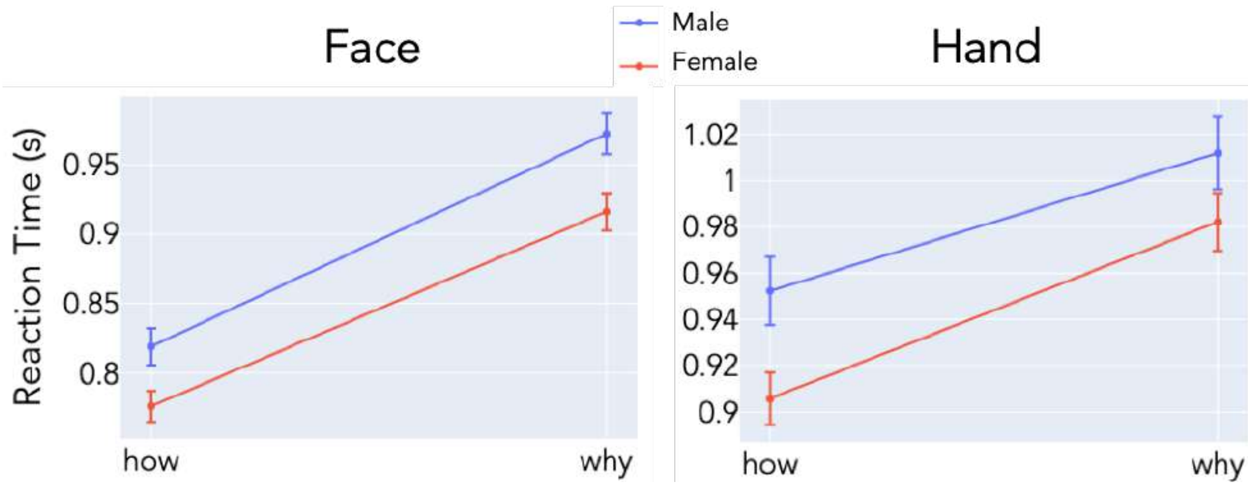


Figure 13: Reaction times for how and why conditions, shown for each sex. Females respond faster and are shown in red. Face and Hand conditions are analyzed separately.

Accuracy: Sex does not appear to play a significant role in determining accuracy for any condition (see figure 14). Neither the ANOVA or T-tests show any relationship between sex and accuracy, or in mediating the relationship between question condition (how vs why) and accuracy. See appendix D for more behavioural ANOVA and T-test results.

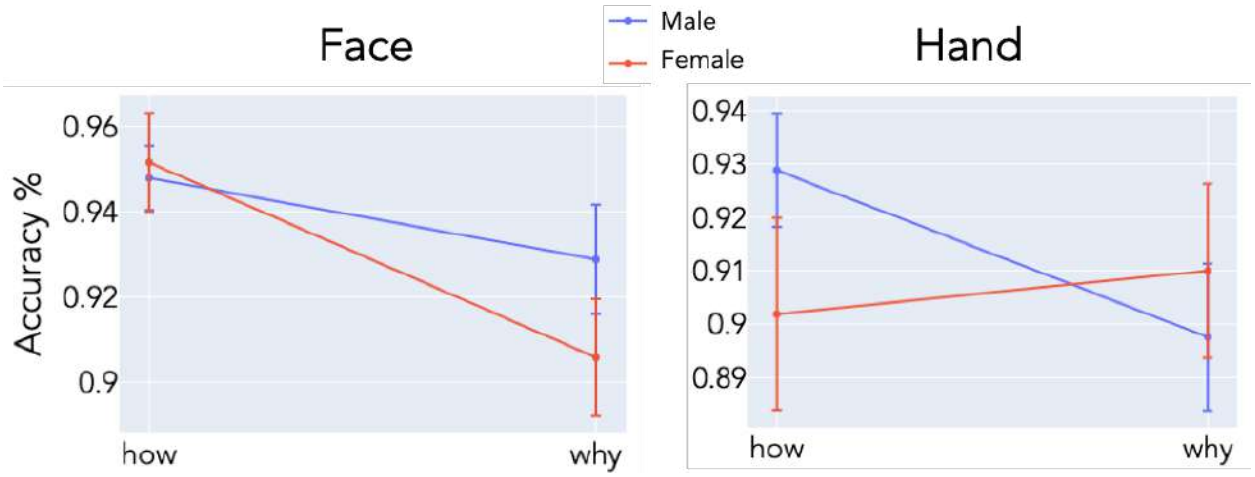


Figure 14: Accuracy on the photo judgement task, split up by sex. There is no correlation between sex and accuracy for either how or why conditions for hands or faces. Female accuracy is shown in red and male accuracy is shown in blue.

4.2.2 How many times is the correct answer "yes" in each condition?

Each time a question is paired with a photo, the answer can be "yes" or "no" according to that photo-question pairing. For example, a question could ask "is this person happy?" and be paired with a photo of a happy person (in which case the expected answer is yes), or a photo of a confused person (in which case the expected answer is no). Because there are several neurological effects associated with expectation violations, a pertinent question to explore is how many times the expected answer is "yes" for each condition. Figure 15 shows the average ratio of the number of intended "yes" responses over total number of questions per subject. There is no significant difference between how and why in this ratio, indicating that any differences we see between conditions would not be due to expectation violations.

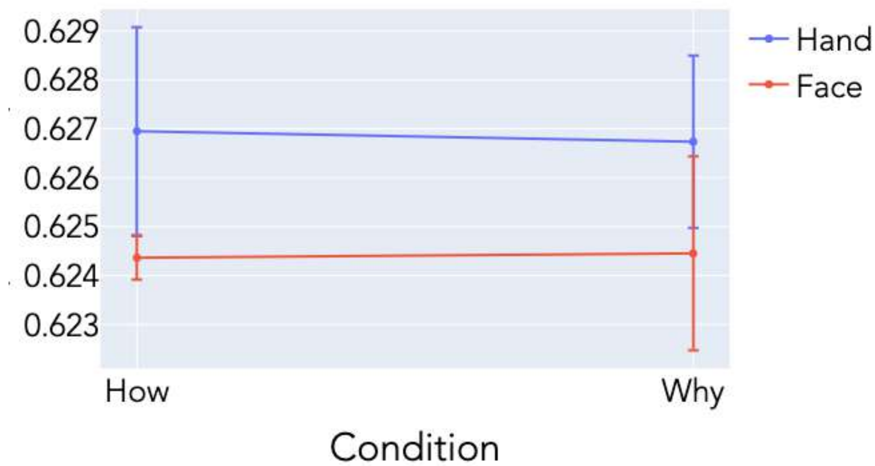


Figure 15: Number of expected "yes" responses per condition. The y axis shows the proportion of expected yes answers to each photo-question pair over the total number of questions. This is compared across how and why conditions for both hands and faces, and statistical analysis shows that there is no significant difference between any condition.

4.3 ERP Analysis

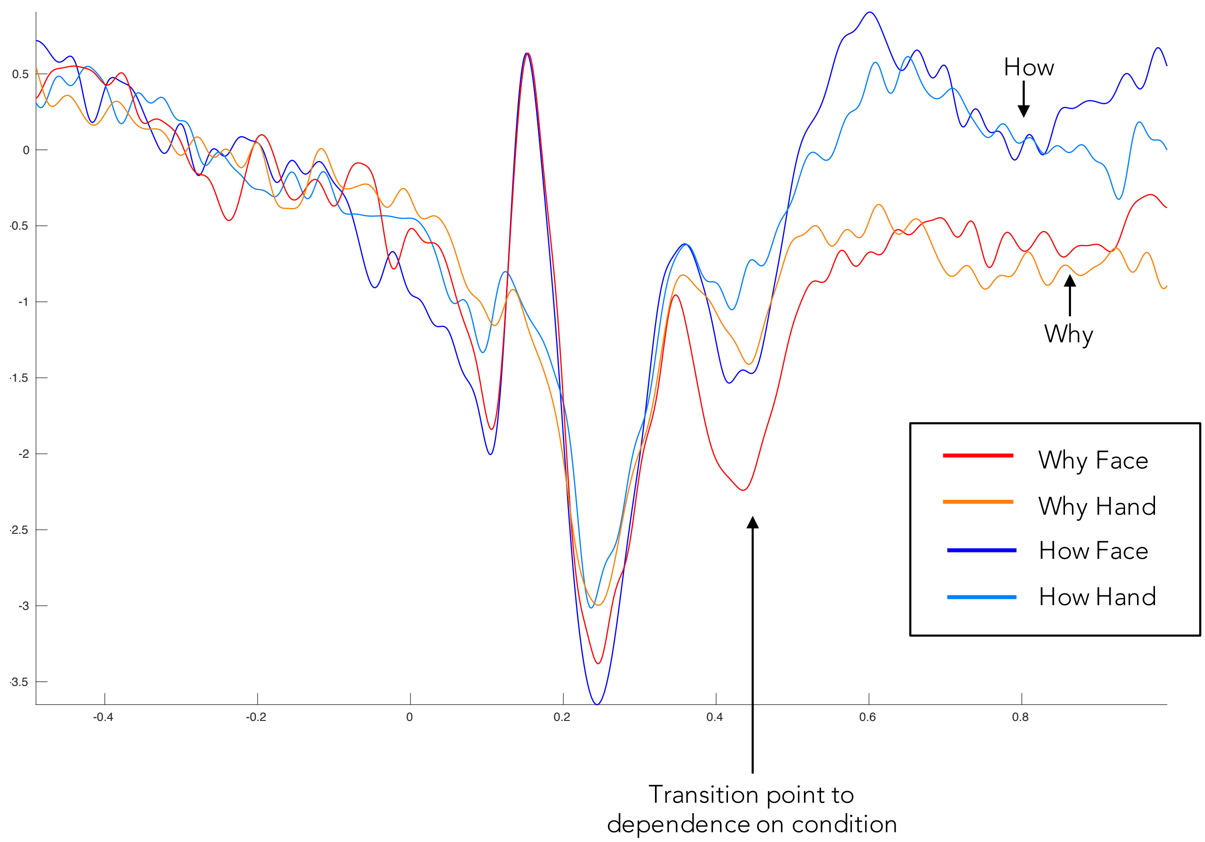


Figure 16: ERPs in all central electrodes averaged relative to event onset for each condition. ERPs are plotted positive up and timing is represented in seconds. Darker colours are associated with images of faces, and red/orange represents questions about intent (MZN), while blue represents questions about means (MNS).

ERP results showed significant difference in both stimulus and group in central electrodes. Figure 16 shows ERPs for each condition (How-Face, Why Face, How Hand, Why Hand) averaged across all central electrodes. The red and orange lines represent the *why*, or mentalizing condition, while the blue lines represent the "how", or mirroring condition. The darker lines (dark blue and red) represent response to images with faces, whereas the lighter lines (orange and light blue) represent responses to images with hands. For the first 250 milliseconds it seems that ERP components are highly dependent on stimulus. However, around 400 milliseconds this relationship switches and the ERP components seem to be more tied to the type of questions that were asked (why vs. how). We explore the difference between the why and how conditions by averaging across stimuli (face and hands) in the next section.

4.3.1 Mentalizing vs. Mirroring

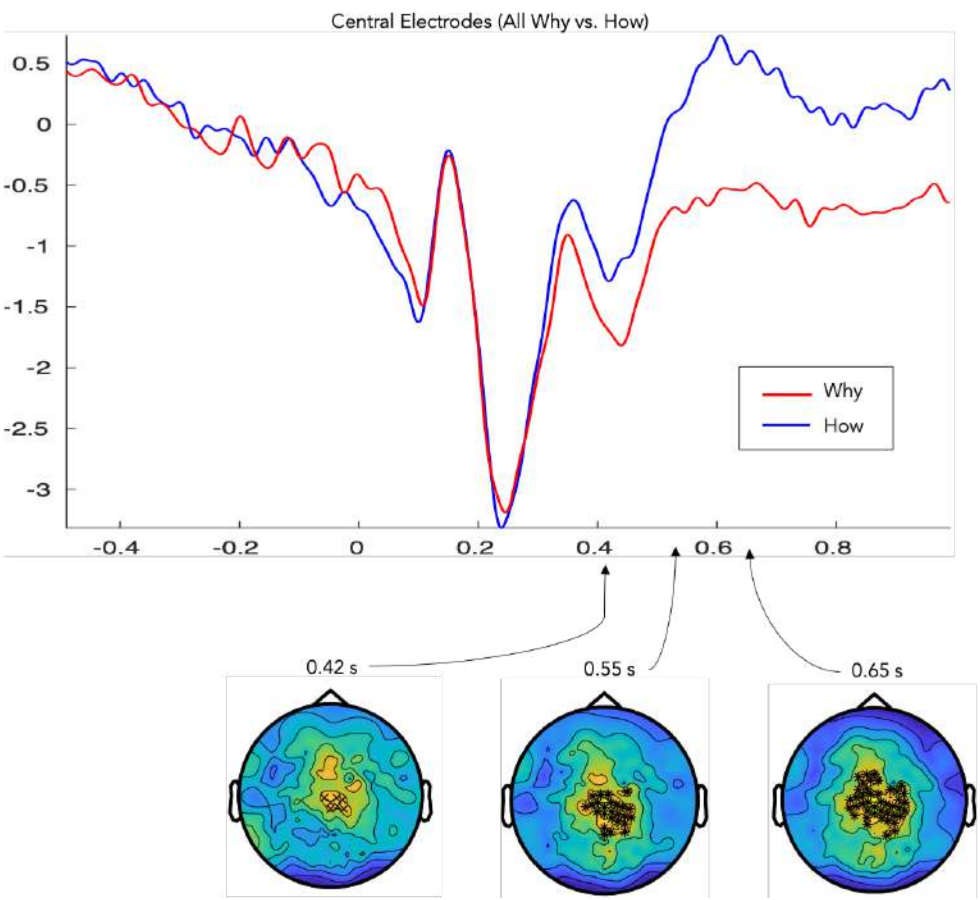


Figure 17: ERPs averaged across the "why" and "how" condition for all central electrodes. ERPs are plotted positive up with time shown in seconds on the x axis. Results from cluster permutation analysis are shown below corresponding to several time points (p<0.025). Each of the stars in the cluster permutation map represents significant differences between conditions.

Investigating differences between averaged “how” and “why” conditions shows ERPs that are almost identical before 300 ms. Figure 17 shows these results alongside cluster permutation results at several time intervals. At around 400 ms, the two conditions (why vs. how) start to diverge, and the cluster permutation analysis indicates where the significant differences are localized in the topographical maps of the brain (shown underneath the ERP plots of figure 18). Whereas differences are highly localized to a small central area at 400 ms, this difference appears to spread out over-time, indicating that this effect increases in either in intensity or location. Cluster permutation results cannot be used to make any conclusive statements about differences between regions to be made so we also applied t-tests to each condition to compare ERP peak amplitudes for each individual. We expected differences in central electrodes from the literature and our ERP plots indicate that the conditions do indeed diverge in these electrodes around 350+ms. T-tests show that these conditions are significantly different ($p<0.05$) in most central electrodes (see Table 6).

Electrode	Component	P-Value
45	N400	0.015
45	aP3b	0.041
45	LPP	0.001
59	N400	0.021
59	P3b	0.045
59	LPP	0.007
60	N400	0.032
60	LPP	0.004
66	N400	0.034
66	LPP	0.017
90	LPP	0.006

Table 6: ERP peaks by electrode and ERP component that are significantly different between why vs. how. All significant electrodes correspond to central electrodes.

4.3.2 Stimulus specific effects

Visual inspection of ERPs in Figure 18 demonstrates that the stimuli (hands vs faces) result in different ERP patterns. Facial expressions are amongst the first things noted about an individual, and they have been shown to elicit a measurable EEG response even before they reach conscious awareness[84]. Previous studies have shown that the N170 is sensitive to faces and emotional arousal[85]. By averaging ERPs across how and why components and plotting them in the same central electrodes as shown previously, we can see that there are clear stimulus specific differences. Figure 18 shows that ERPs are different at 170 ms as expected. Because the N170 response is typically seen in occipital regions, we plot the ERP averaged across occipital electrodes in figure 19. The face condition also appears to have an increased N450 and higher LPP than hands.

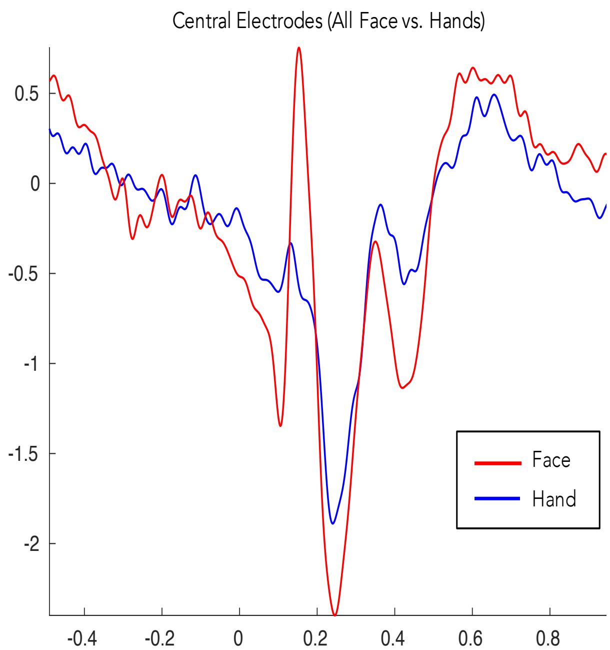


Figure 18: ERP waveforms averaged across all central electrodes, comparing the face and hands conditions.

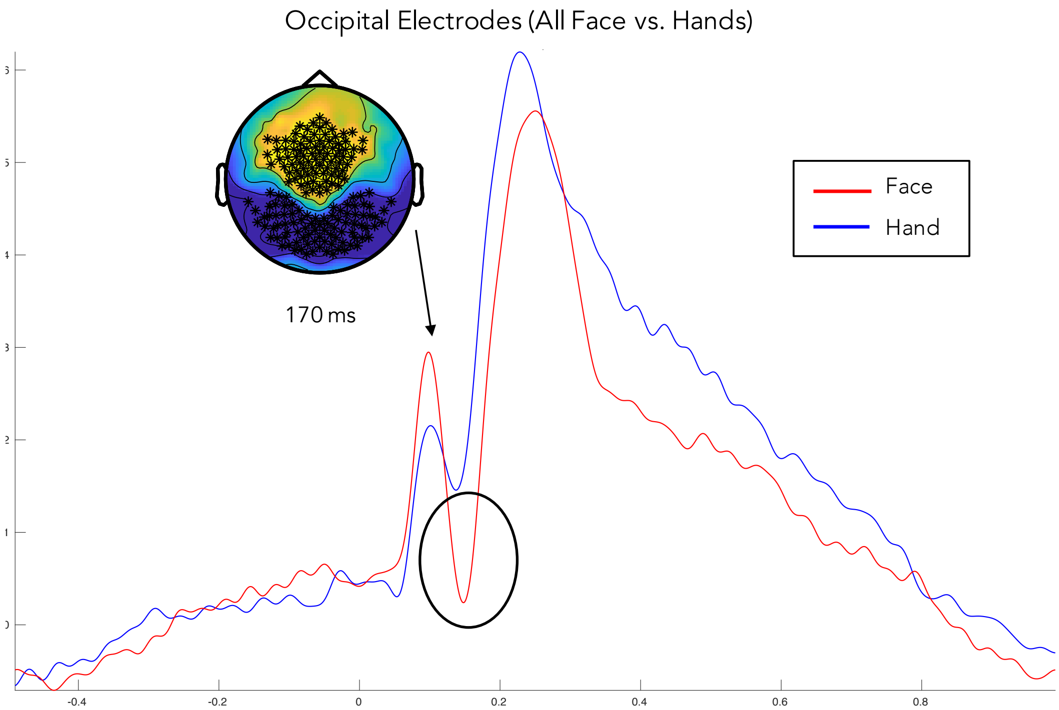


Figure 19: Comparison between hands and face ERP over all occipital electrodes. The stars on the topographical map show which electrodes are significantly different (p<0.025).

We see a larger N170 peak component in occipital areas for face stimuli as expected, but interestingly we see differences at later time points as well. A cluster permutation analysis around 170 ms (represented by the topographical map in figure 19) shows highly significant and stable differences between the two conditions globally. These results were also verified by t-tests of N170 peak amplitudes across several electrodes in each condition.

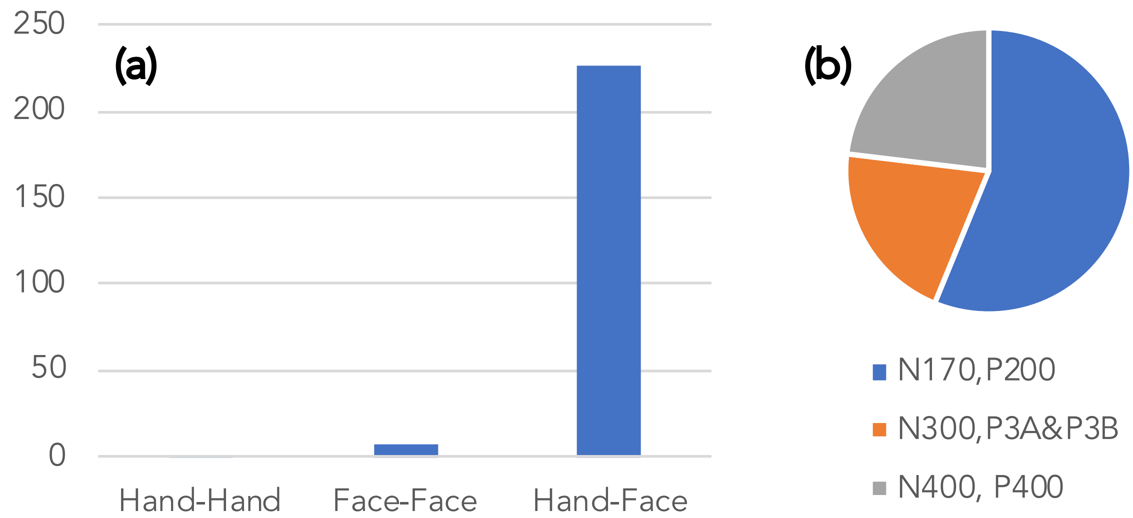


Figure 20: (a) The number of comparisons that are significantly different for each condition-pair (p<0.05). There is a separate t-test for each electrode and each ERP component of interest. (b) The ratio of each ERP component's contribution in the total number of significant Hand-Face comparisons from the bar graph.

We performed an analysis to determine how much varying stimuli can contribute to differences in ERPs across conditions. To do so we performed numerous independent t-tests for different condition pairs across all elecrodes and peak amplitudes within several ERP components. Pairings that had different photos/stimuli (i.e. How-Hand vs. How-Face; Why-Hand vs. How-Face; and Why-Hand vs. Why-Face) were many more comparisons with significantly different peaks ($p<0.05$), as shown in figure 20 a. Breaking down all pairings with different stimuli (hand vs face) into which ERP components contributed the most distinguishable peaks shows that early ERP components (N170 & P200) make up the bulk of the differences in all electrodes for the stimulus specific differences (Figure 20 b). We also found that the Why-Hand vs. How-

Face pairing was far more likely to have $p<0.05$ as an outcome to the t-tests than both the Why-Hand vs. Why-Face and How-Hand vs. How-Face pairings combined. This indicates that there may be an interaction effect between the type of stimuli shown and the condition, which we explore in the next section.

4.3.3 Is there a face specific effect that relates to mirroring or intent?

Facial expressions are representations of our inner emotional states and are central to non-verbal communication. There are at least 46 unique muscle movements that represent specific emotions represented expressed on the face[84]. There are different brain areas devoted to processing information in the face[86], so it is likely that the mentalizing network for face processing is distinct from the mentalizing network employed in the absence of facial expressions. In order to explore this question further we create 'face minus hand' contrasts and compare the corresponding why and how conditions (see appendix E on how this was done). Figure 21 shows ERPs across a few different electrode locations and shows where the significant differences are localized in a cluster permutation. Cluster permutation analysis shows that the there was a significant difference for this how vs face contrast from 400 and 600 ms seconds. Moreover, it suggests that the effects are most likely localized to left central/left parietal and left frontal electrodes (figure 21 c). ERP plots of these regions in figure 21 (b) (d) and (e) show that the *why* (face – hand) condition has a more negative deflection than the *how* (face – hand) condition.

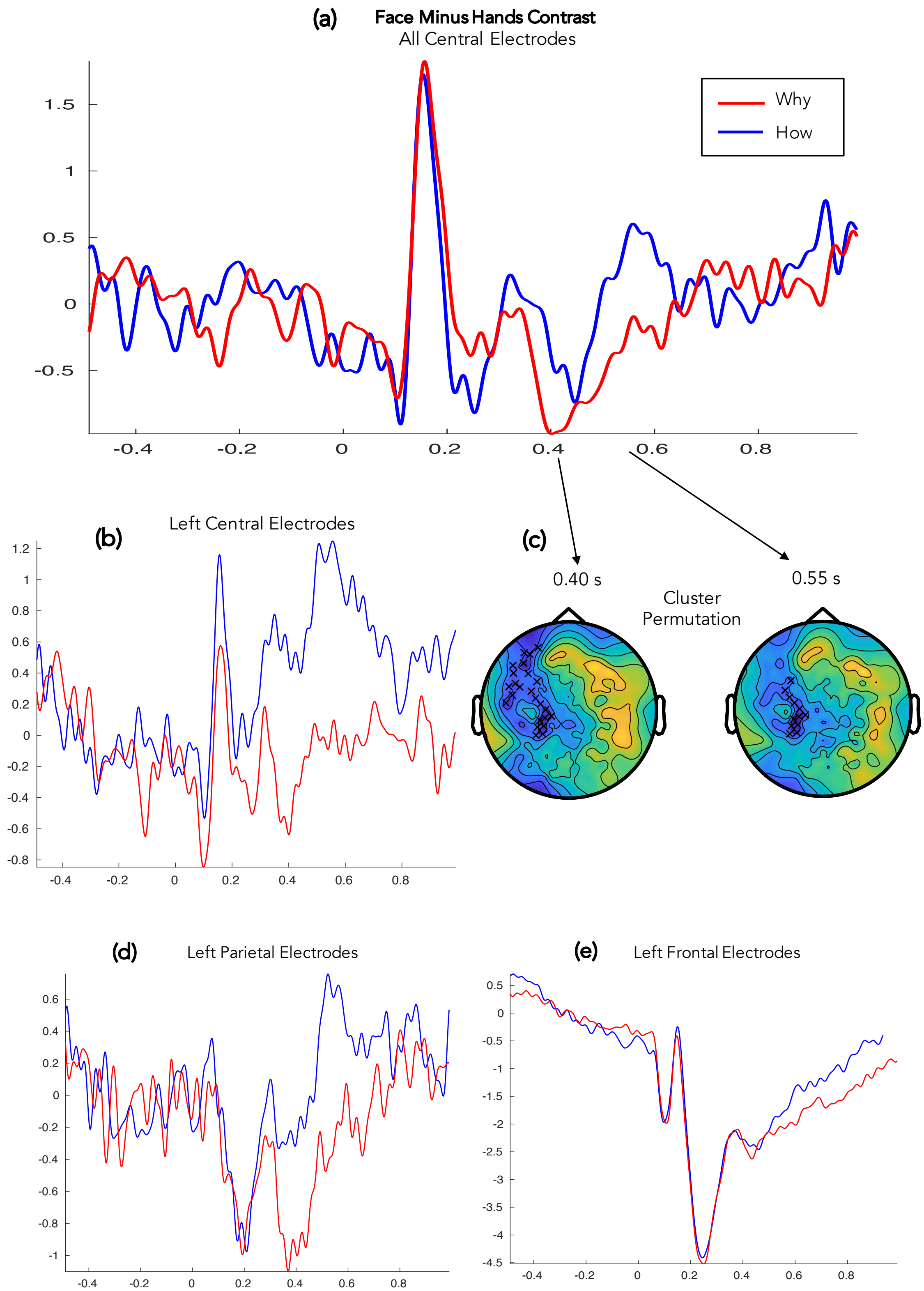


Figure 21: (a) shows the ERP for the how vs why conditions of the face minus hands contrast in central electrodes. Figures (b), (d) and (e) show the same in left-central, left-parietal and left-frontal electrodes respectively. Figure (c) shows which areas are significantly different at 0.40- and 0.55-seconds using cluster permutation analysis (p<0.0125) motivating plots (b), (d) and (e).

4.3.4 Sex-Based Differences in ERP

There were some issues in EEG data collection for participants 26-35, so these datasets were removed from ERP analyses due to excessive noise. Because most of these datasets (90%) happened to be males, we are limited in the interpretation of our following results. By visualizing ERPs in Figure 22 and 23, the only difference appears to be a possible higher N170 response to faces by males, however, these results are not significant when tested. Moreover a previous study has shown the opposite relationship, with females having a higher N170 response to faces than males[87].

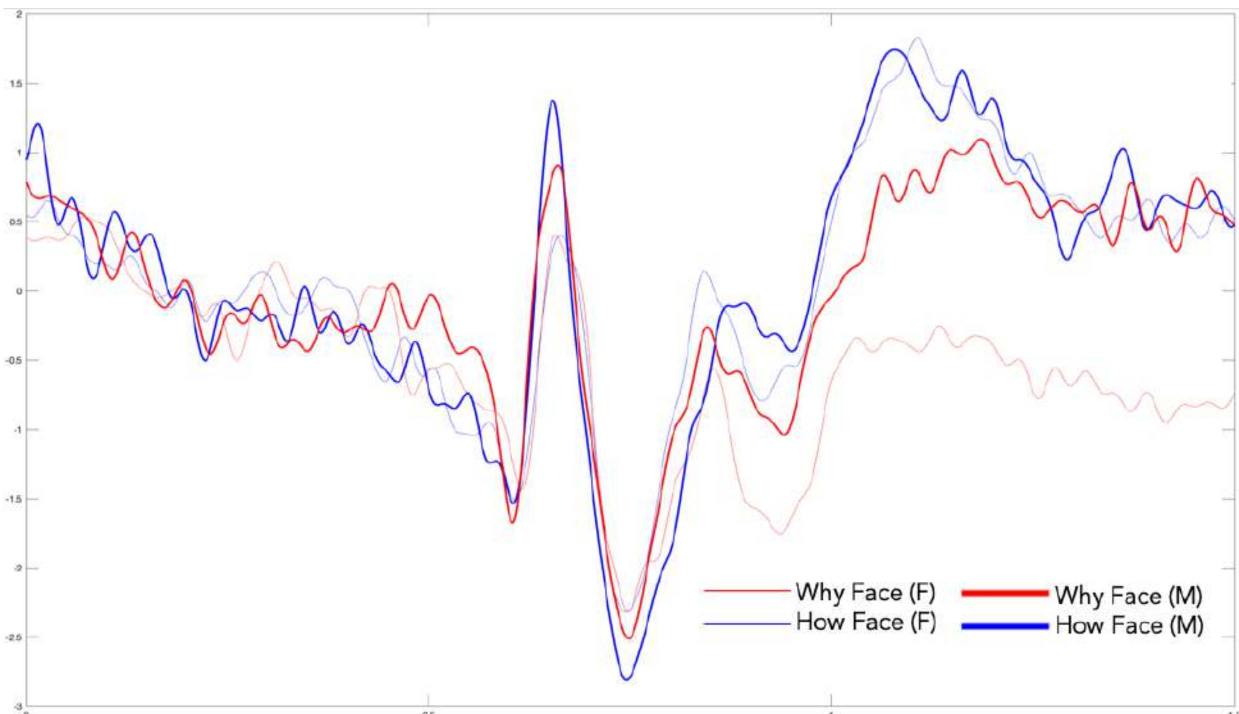


Figure 22: Averaged ERPs for How and Why conditions in response to face stimuli, split by sex in central electrodes. Female ERPs are presented with finer line thickness for each condition, and male ERPs are shown in thicker lines.

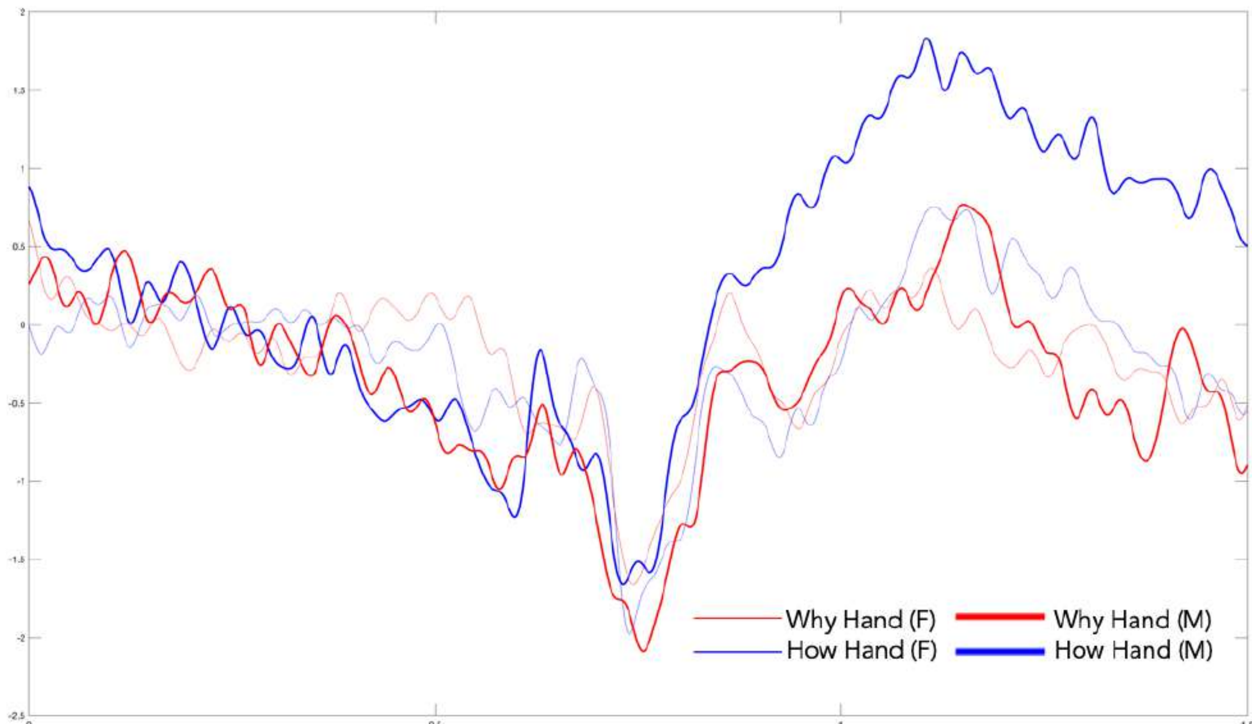


Figure 23: Averaged ERPs for How and Why conditions in response to hand stimuli, split by sex in left parietal electrodes. Female ERPs are presented with finer line thickness for each condition, and male ERPs are shown in thicker lines.

4.3.5 Is N400 amplitude correlated to behavioural results in any condition?

We were interested in determining whether or not N400 was associated with task accuracy, so we correlated N400 peak amplitude and task accuracy. For the *how hand* condition there is no correlation between accuracy and ERPs. There is a moderate negative correlation between accuracy and N400 amplitude for the *how face, why face* and *why hand* conditions (correlation of r^2=0.11, p=-0.33; r^2=-0.096, p=0.31; and r^2=0.12, p=0.34 respectively). In these cases, greater (negative, reversely coded) N400 amplitude is moderately associated with higher task accuracy. For the *how hand* condition however, we do not observe a strong or moderate correlation between N400 amplitude and task accuracy. All the plots are shown in figure 24.

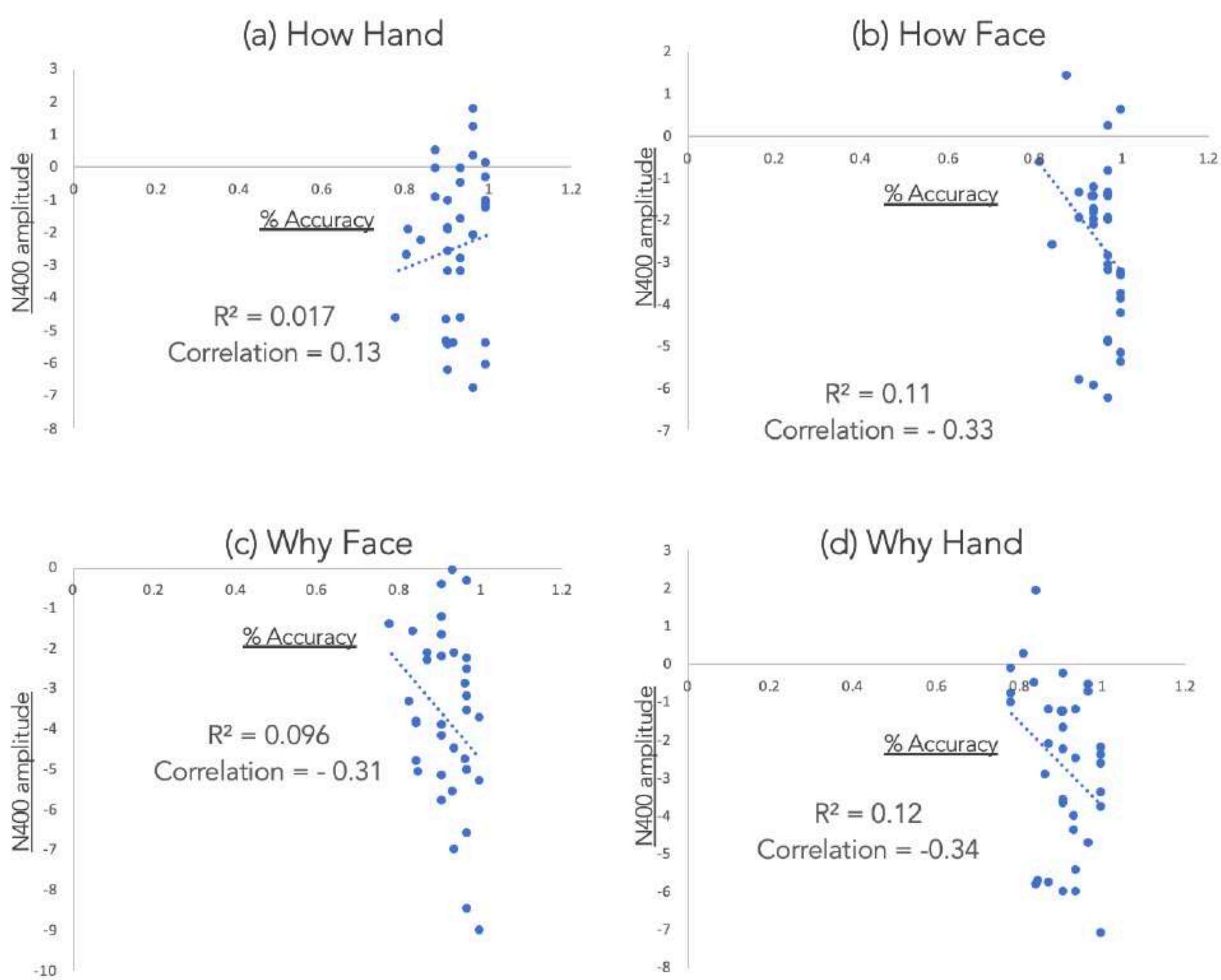


Figure 24: A scatterplot N400 amplitude and thee corresponding task accuracy for each subject in each condition. Dotted lines represent the line of best fit. Correlations and R^2 values are overlayed on each plot.

4.4 Do we see differences in alpha and beta suppression?

Figure 25 shows changes in power at various frequencies through time for the averaged how and why condition at central electrodes. By taking the alpha power of individual subjects at multiple time windows, we observed no significant differences between the *how* and *why* conditions in central electrodes. By performing a cluster permutation analysis, we can confirm that there are no significant differences in mu or beta power for the averaged *how* and *why* at *any* electrode cluster (two tailed $p<0.025$).

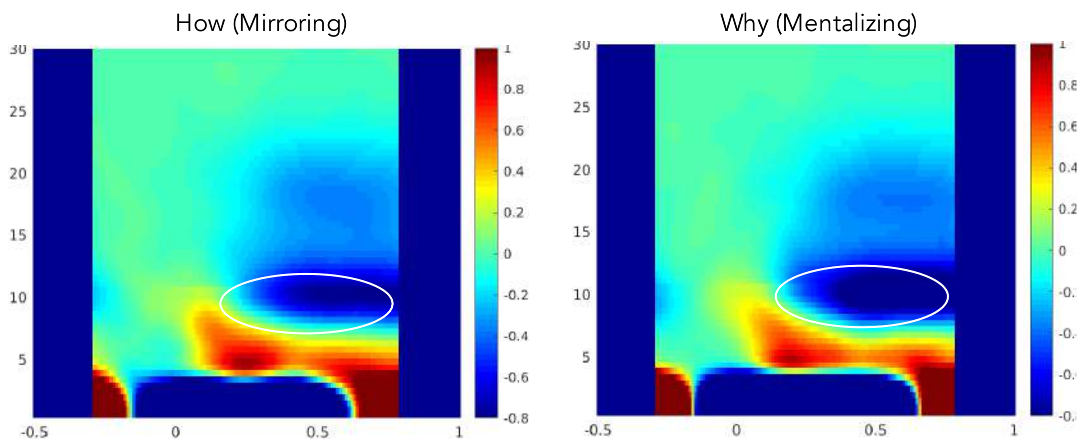


Figure 25: Time frequency plots for averaged how and why conditions at central electrodes.

In order to see if the lack of differences we see above can be explained by differences in face and hand processing by the MZN and MNS, we split the time frequency plots up by stimuli and mentalizing/mirroring condition, represented in figure 26. Statistical analysis at various time windows for both alpha and beta suppression shows that there are no significant differences between how and why conditions for either hand or face at central electrodes. Interestingly, there are several other regions with significant differences between the face and hand conditions, indicating that both mu and beta suppression are sensitive to stimulus type (See Appendix F).

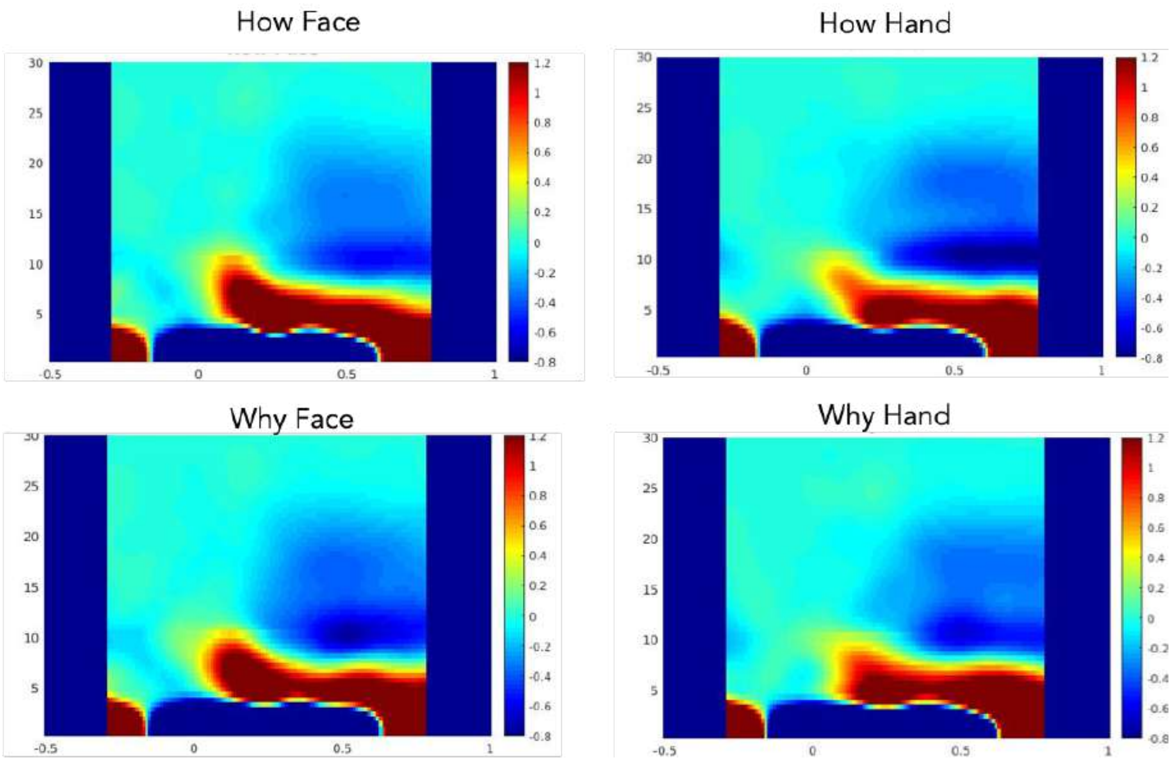


Figure 26: Time frequency plots for each individual condition at central electrodes. There are no significant differences in mu or beta frequencies between how vs why conditions in these electrodes.

There is evidence that mu suppression in right central electrodes is tied to face processing in particular. We performed another analysis to see if we could tease apart *how* and *why* conditions in this region for faces. We observe increased mu suppression in the *why* condition for right central electrodes that starts at around 400 ms and lasts for another 200 ms (see figure 27). However, this difference was not apparent in the hand condition for this region.

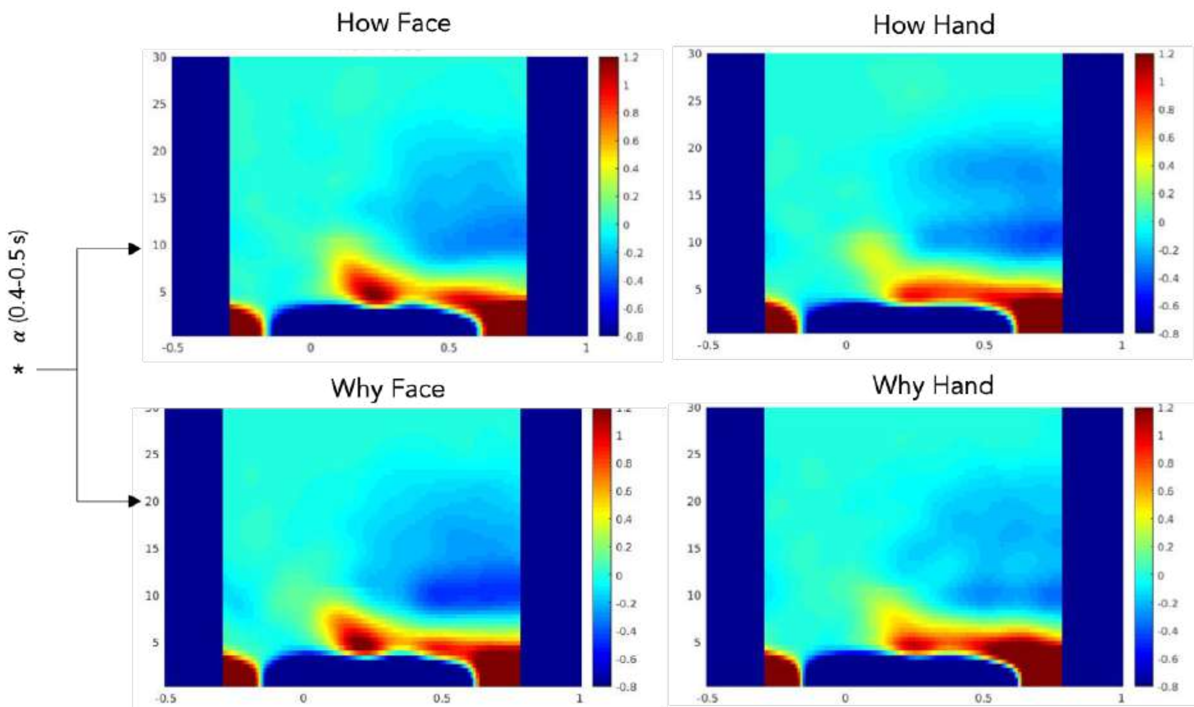


*Figure 27: Time frequency plots in **right** central electrodes. This is where we expect to see face specific effects. There is a significant difference (two tailed p<0.025) in alpha power (10-12 Hz) between How and Why Face conditions at 400 ms. There is no significant difference in alpha or beta power between How and Why Hand conditions.*

Beta power modulation has been observed in mentalizing tasks. For this reason, we investigate differences in beta suppression across why and how conditions for both face and hands at various sites of interest. We found that there is a significant task dependent beta modulation in left frontal electrodes. We observe a higher suppression of power in the beta band for the *why* condition than the *how* condition around 600 ms (see figure 28). However, this difference is only significant for the hand condition and does not exist for faces.

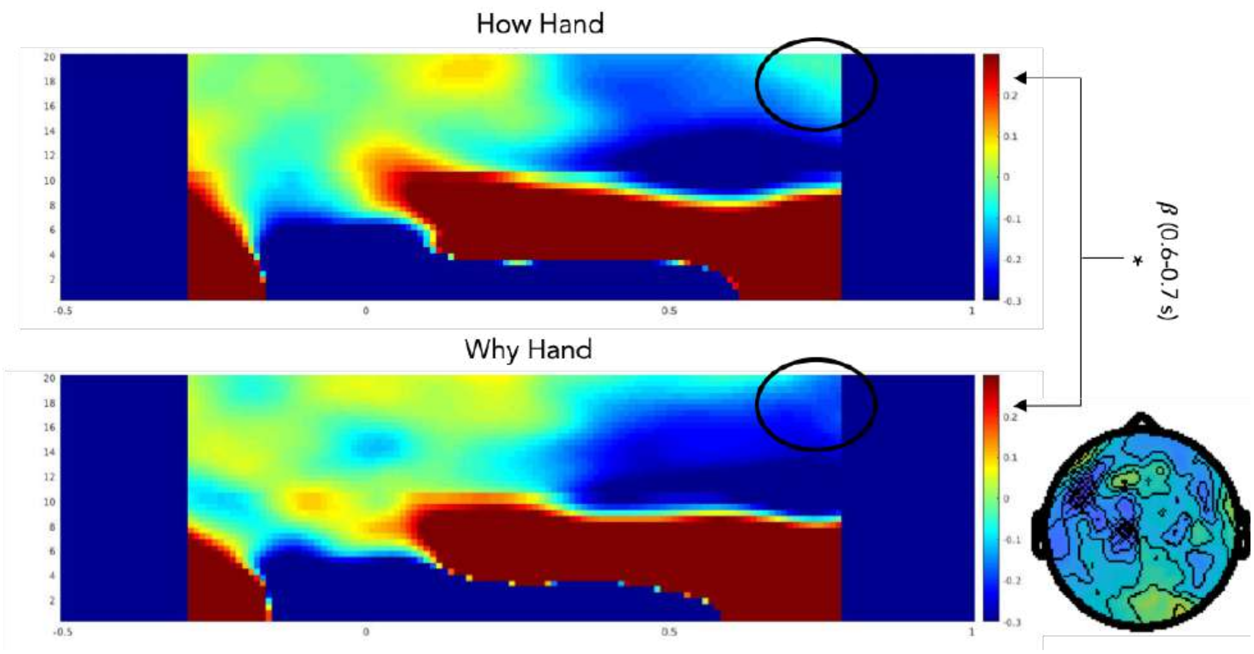


Figure 28: Time frequency plots in left frontal electrodes. We see differences in beta power (15-20 Hz) between how and why hand conditions at 600 ms (two tailed p<0.025). A cluster permutation shows where these differences are localized.

Appendix F shows the alpha and beta time frequency plots in left frontal electrodes where some have also seen changes in beta power associated with mentalizing. None of the differences in frontal electrodes were significant.

4.5 What is the sequence of state transitions in each condition?

In order to investigate our hypothesis of increased cognitive load and higher order processing for mentalizing, we investigate how many stable states the brain occupies for each condition. In order to do this, we take ERP waveforms for all the *how* and *why* conditions and perform microstate analyses on each. Figure 29 shows root mean squared error for each condition along with pointers to identify when a new stable brain state starts and ends.

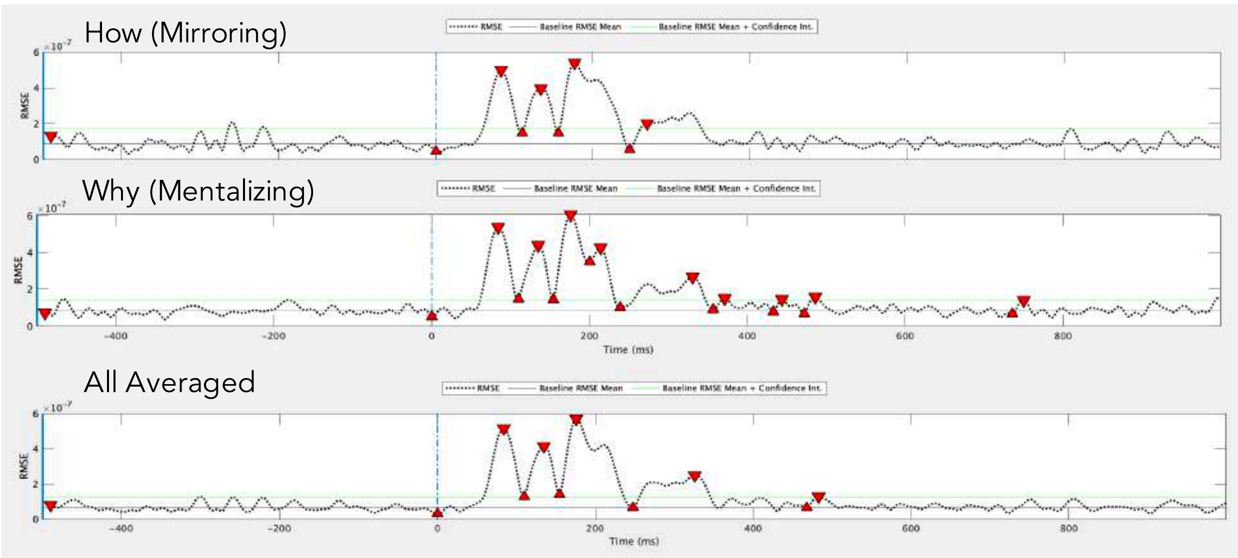


Figure 29: Plots of the root mean square error, alongside red pointers that are used to indicate where state transitions between microstates occur.

The *why* condition has 9 stable states whereas the how condition has only 4. In order to visualize these microstates, we plot each stable state as a block in time, with the transitions between them as empty spaces in figure 39. The first 3 microstates in the *how* condition strongly overlap with the first 4 in the *why* condition. There are several stable states that are occupied for the mentalizing condition (*why*) after 300 ms that are not present in the mirroring condition (*how*). Figure 31 shows the mid-point of each stable microstate, that are used for source plots. See appendix G for all microstate results.

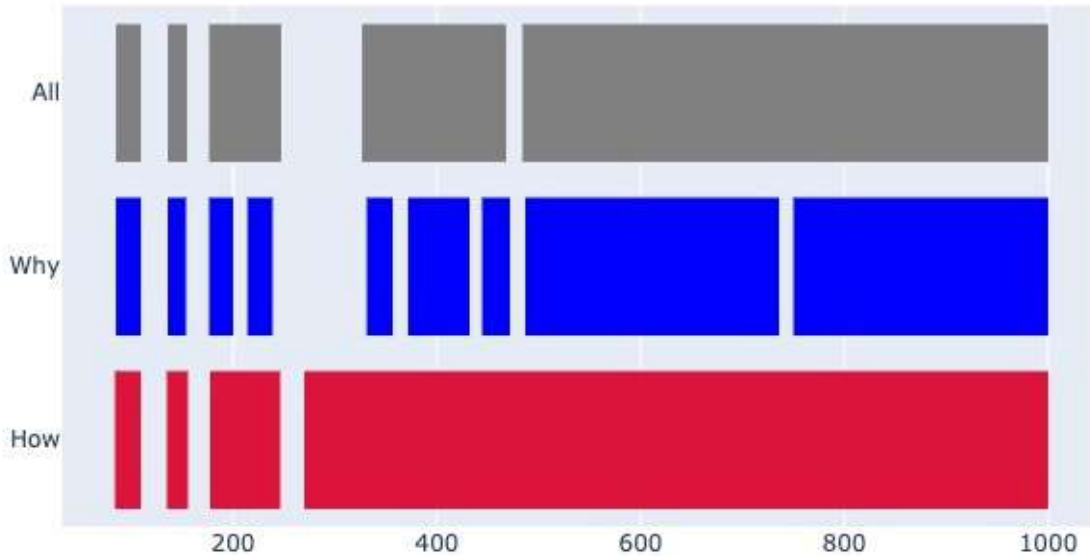


Figure 30: Each block represents a stable topological state, or microstates. Transitions between microstates are represented as empty spaces between blocks. The why condition is shown in blue and the how condition is shown in red. An average across both conditions is shown in grey.

Why

Microstate	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9
Mid-point (ms)	97	145	188	227	344	402	458	611	875

How

Microstate	State 1	State 2	State 3	State 4
Mid-point (ms)	97	145	212	634

All Averaged

Microstate	State 1	State 2	State 3	State 4	State 5
Mid-point (ms)	97	145	212	397	742

Figure 31: Microstate mid-points used for source localization. The why condition is shown in red, and the how condition is shown in blue. An average across both is shown in grey.

4.6 What sources are activated?

Understanding what sources are activated in each condition (*how* vs *why)* can provide some insight into which brain networks are associated with MNS and MZN. Source reconstruction was performed at the mid-point of each stable microstate (microstates determined based on the condition). Because the inverse problem is ill-posed and EEGs have characteristically low spatial resolution[3], we refrain from labelling activity at a very granular level and stick to identifying larger brain clusters less likely to be subject to noise. We start by exploring both *how* and *why* conditions to facial stimuli using microstate results. For both the *how* and *why* conditions we see that in the first microstate (97 seconds) there is predominantly occipital activity in primary visual areas with higher activity in the right hemisphere as shown in Figure 32 and 33 (figures for how hand and why hand are shown in appendix H). Figure 34 shows that for the first ~230 ms there are no differences between how and why conditions. We discuss condition contrasts in detail below.

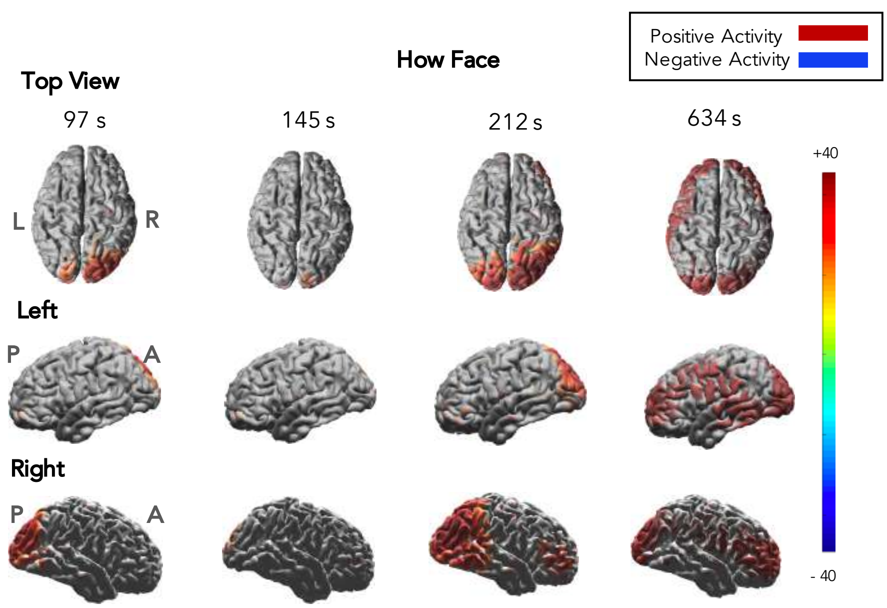


Figure 32: Source analysis of the How Face condition selected at the mid-point of each stable microstate.

[3] The aim of the inverse problem is to find the sources that generated the EEG patterns observed.

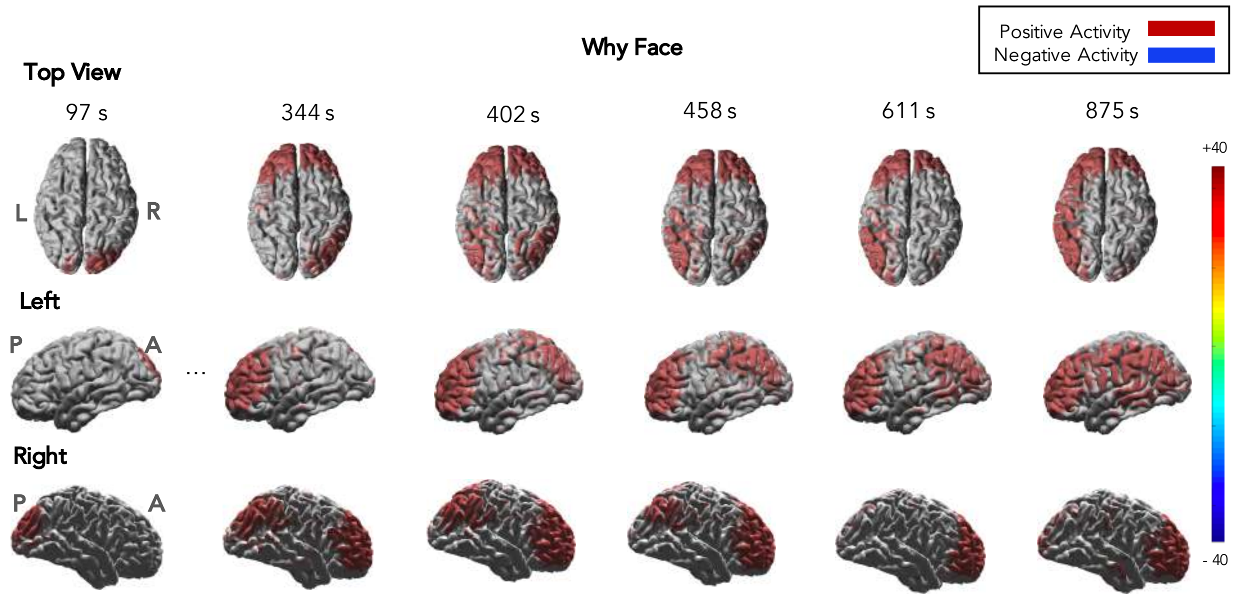


Figure 33: Source analysis of the Why Face condition selected at the mid-point of each stable microstate. Microstates 2 and 3 are excluded from this image but are shown below in Figure 34.

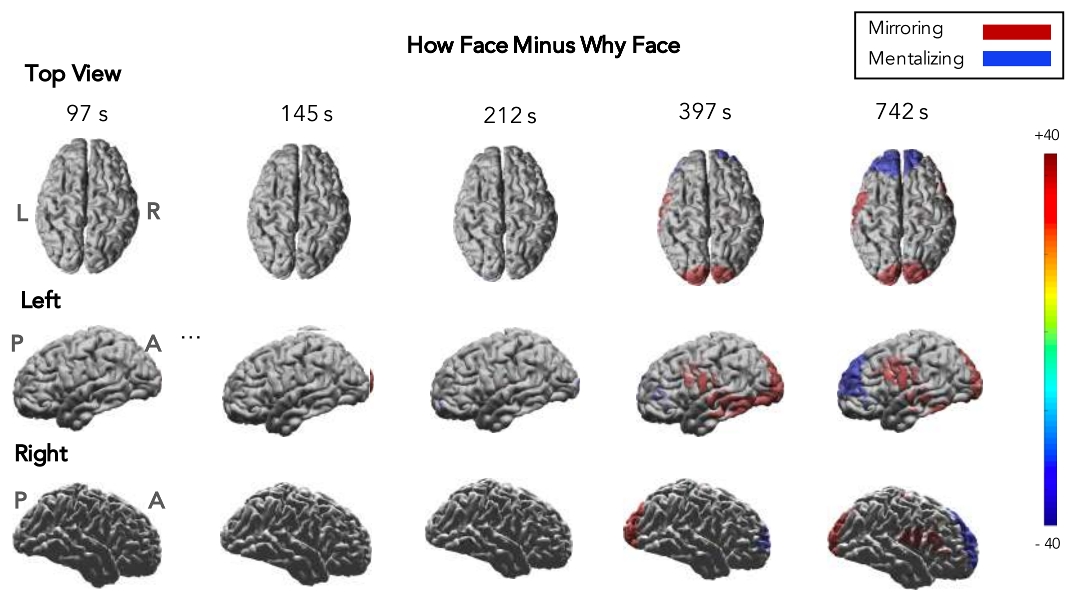


Figure 34: Source analysis of the How minus Why contrast for faces, selected at the mid-point of each stable microstate (microstate times determined from the entire population). Positive values, or red colours indicate when the how condition is greater than the why condition, which we use as a proxy for mirroring activity. In contrast, the blue indicates when the why condition is greater than how, which we use as a proxy for mentalizing activity.

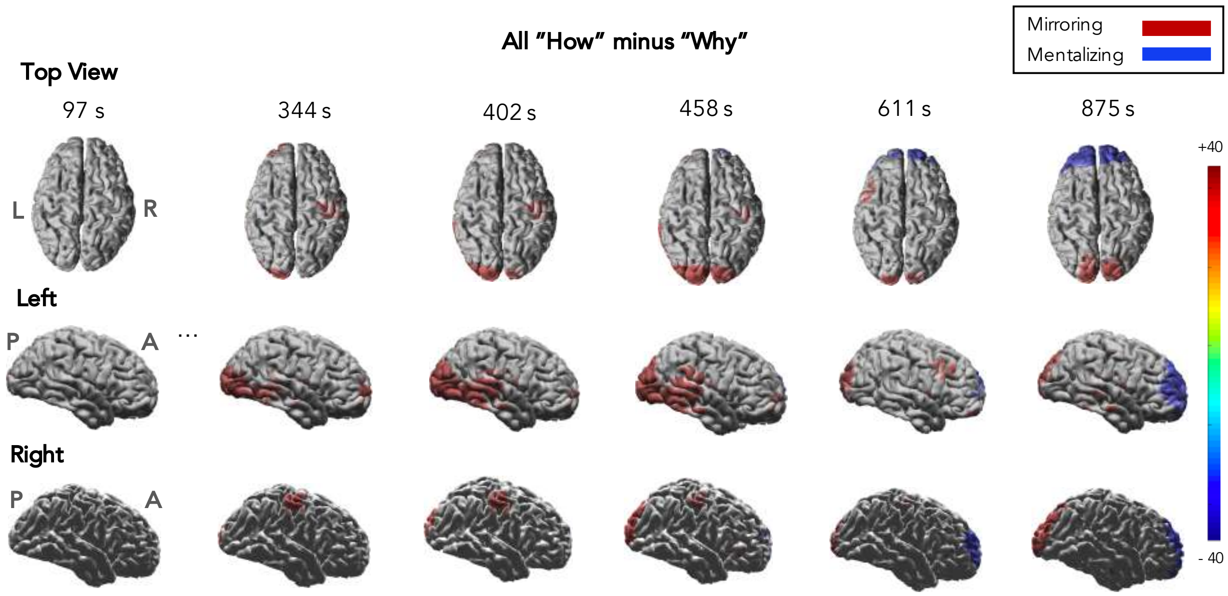


Figure 35: Comparison of how and why conditions at each stable microstate of the why condition. We exclude the first 3 microstates as there are no differences. Positive values, or red colours, indicate when the how condition is greater than the why condition, which we use as a proxy for mirroring activity. In contrast, the blue indicates when the why condition is greater than how, which we use as a proxy for mentalizing activity. Microstates 2 and 3 are shown in the next figure.

By comparing *how* and *why* conditions for just face stimuli using the 5 microstates from figure 35 we can note the following:

1. The first 3 microstates show no differences between conditions
2. There is higher activity in the occipital lobe, left precentral gyrus, left postcentral gyrus and left superior temporal gyrus for the *how* condition at the 4th microstate (397 s)
3. There is higher activity in the occipital lobe, the left postcentral gyrus, right postcentral gyrus and the left superior temporal gyrus for the *how* condition

at the 5th microstate (742 s). There is higher medial prefrontal cortex activity for the *why* condition at the 5th microstate (742 s)

By comparing *how* and *why* conditions for both face and hand stimuli using the 9 microstates from the *why* condition in figure 35 we can note the following:

1. The first 4 microstates show no differences between the conditions
2. There is increased activity in the occipital lobe and the right IFG for the *how* condition in microstates 5-7 (344-458 s).
3. There is increased activity in the left superior temporal gyrus and superior temporal sulcus for the *how* condition in microstates 6-7 (402-458s).
4. There is increased activity in the left inferior frontal gyrus for the *how* condition in microstate 8 (611 s).
5. There is increased activity in the medial prefrontal cortex for the *why* condition in microstates 8-9 (611-875 s).

Comparisons for the microstates before 300 ms are presented in figure 36 and show that there is no difference between stable *how* and *why* states then.

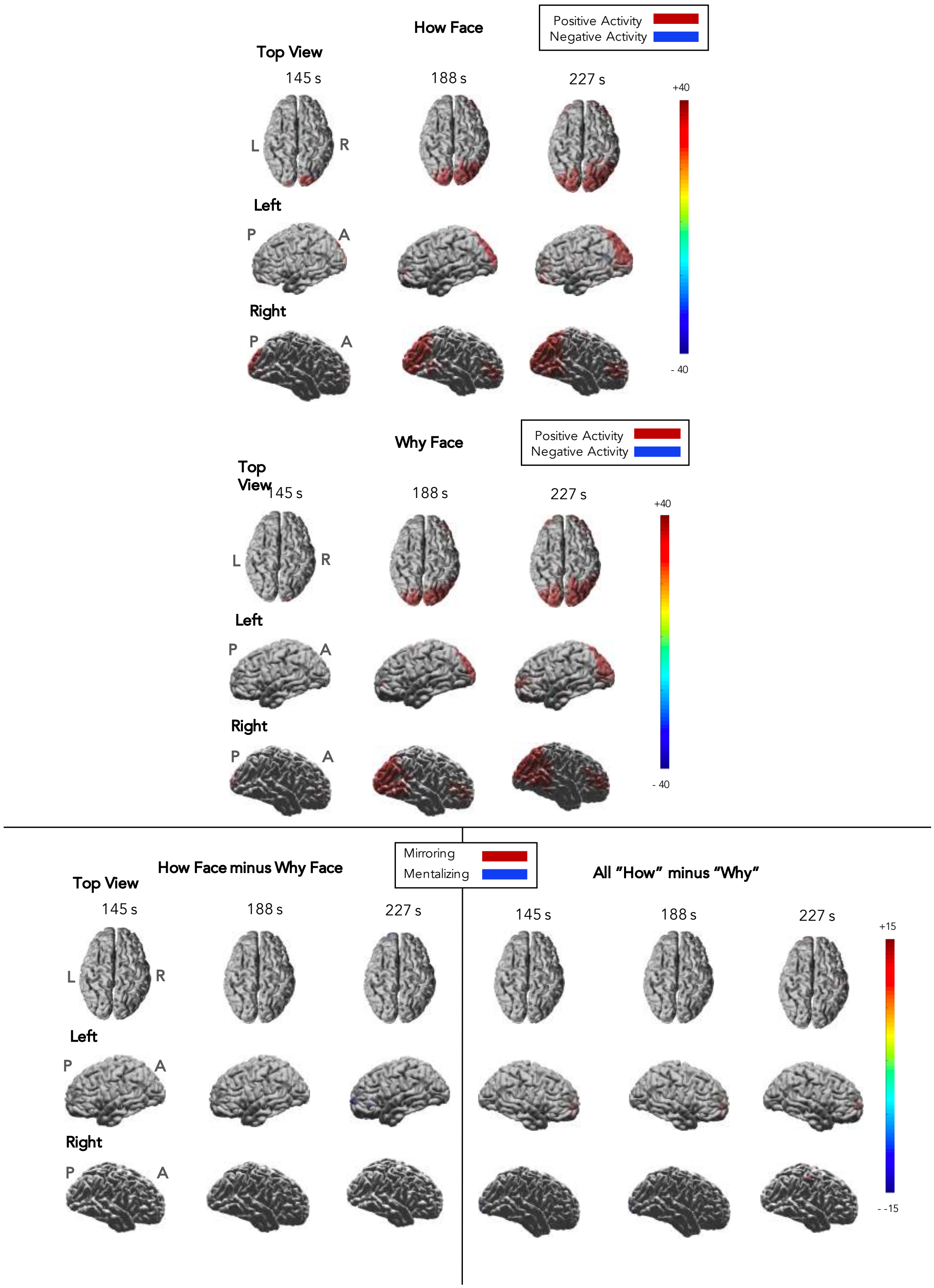


Figure 36: Source analysis for the first three microstates in multiple conditions. These results show that in the first 3 microstates there are no differences between the social brain condition. It is only after this time that subjects start to process the questions. The conditions over which the source activity is averaged are stated above the source plots.

5 Discussion

5.1 Behavioural Analysis

In the photo judgement task, we saw that participants were faster and more accurate in their responses to *how* questions than *why*. This is in line with results from Spunt and Adolfs[6]. We assume that the *why* and *how* conditions tap more into mentalizing and mirroring processes respectively, as was corroborated by the fMRI sources for each condition[6]. Given this assumption, we use the *why* condition as a proxy for mentalizing activity and *how* as a proxy for mirroring. Our results suggest that *how* questions may require less cognitive load, as they require less time to answer (reaction time) and were more correct (accuracy). These results may support the hypothesis that mentalizing is a more cognitively demanding task than mirroring. MEG and EEG studies have shown that mental effort can modulate long-distance functional connectivity between functionally district areas[88,89]. Moreover, there are a few EEG connectivity measures that can provide a measure for how many long-distance connections exist [89,90]. Using these properties, one study demonstrated that in contrast to mirroring tasks, mentalizing tasks modulate these network connectivity measures in the same way as increasing cognitive load does[39]. Taken together our results support the idea that the mentalizing processes are more cognitively demanding than mirroring ones. In the absence of explicit visual information to base judgements on, as required by mirroring functions, mentalizing tasks must require global integration and involvement from other task related brain areas, thereby increasing cognitive load.

In addition, our results demonstrate that the relationship between accuracy, reaction time and social brain network (how/MNS vs. why/MZN) are dependent on the type of stimuli used (face/hand). We see a relationship such that responses to pictures of faces are likely to be more accurate and take less time to judge than responses to pictures of hands. This could indicate that we process information about faces more easily than we process information about hands. The relationship between reaction time and social brain network was furthermore meditated by stimuli type. We see that images of faces reduce reaction time more than hands when responding to *how* questions, that rely on mirroring, relative to *why* questions, that involve mentalizing. The fusiform face area is involved in recognizing faces. This may confer an advantage that the mirroring system has in identifying faces relative to hands. This particular advantage may not be apparent in mentalizing tasks which require additional rate limiting processes to infer intent.

Splitting up behavioral results to see if there were any sex-based differences showed that females are significantly faster ($p<0.05$) than males at responding to questions in all conditions. Surprisingly, the quicker response times had no bearing on accuracy of the responses. Interestingly, a previous study investigating responses of males and females to an emotion identification task found that there were no significant differences in reaction time (RT) between males and females[91].

5.2 ERP analysis

We investigated ERPs in 3 categories of interest: the late components are thought to reflect information or intent processing; middle components are typically implicated in lower level goal inferences; and early components usually reflect physical parameters of the stimuli.

5.2.1 Late ERP components

A few studies on trait inferences documented that information about personality traits and intent, associated with mentalizing, occur at 400+ ms, irrespective of instructions[105,106]. As mirroring activity is thought to begin immediately after a stimulus is processed, this implies that mentalizing begins later on. Our results showed that the *why* condition exhibits a significantly more pronounced negative peak around 400 ms (N400) in central electrodes. Correspondingly, a positive peak following the N400, known as a late positive potential (LPP), was stronger in the *how* condition. As the electric fields picked up at the brain are a summation of both positive and negative potentials, it is hard to determine whether or not the difference in LPP is simply due to a stronger and longer lasting negative potential (N400) for the *why* condition, or if the two ERPs represent distinct processes at play. We can explore this question by investigating what previous literature show us about both LPP and the N400.

One study exploring the N400 had participants play a game in which they were instructed to make a choice to communicate a truth with good intent, lie, or tell a truth with deceptive intent. They found that a frontal and central N400 component was sensitive to the *intent* of the participant[107]. The same N400 deflection occurred when participants lied or told the truth with deceptive intent. They claimed that this negative deflection was result of an internal conflict that involved managing another person's mental state, while simultaneously deceiving them, irrespective of the actual 'truth' [107]. In addition to these findings, the N400 has been found to be significantly more negative following trait inconsistences. Trait inconsistencies are when information is presented about someone that conflicts with personality traits previously are associated with them[100]. This result shows that the N400 response may be involved in understanding unanticipated behaviour by relying on mentalizing processes. Our results tie in with these findings, as we show that the N400 was more active in the condition in which participants were asked to make judgements about other's mental states.

The relationship with LPP that we see is less clear. A previous study investigating autism found that LPP activity is different for individuals with ASD in frontal and parietal electrodes[14]. They first showed that typically developing individuals have a higher LPP amplitude in response to images that show pain than ones that do not. However, individuals with ASD did not show this LPP modulation by painful stimuli[14]. In fact, there was no significant difference in LPP between painful and non-painful stimuli for individuals for them at all. Another study found that while individuals with ASD demonstrated relatively smaller LPP amplitude to social stimuli, they generated larger LPP amplitude to non-social stimuli[108]. This indicates that the LPP is implicated in social processes. In addition it has been shown in multiple experiments, that emotional pictures are associated with higher LPP than neutral ones[101] and that LPP activity can be used for emotion classification[109]. These results indicate that an LPP is involved in interpreting information that has high emotional valence, and this activity is impaired in ASD.

Our research shows that the N400 is larger when evaluating the intent of an action (*why* condition), while LPP is enhanced when interpreting means (*how* condition). As the stimuli we use across *why* and *how* conditions have the exact same emotional valence, the difference in LPP we see between them may be explained by a different function than emotional valence. The same study linked N400 amplitude to trait inconsistencies also showed that the LPP was more positive in response to trait inconsistencies[100]. Sensitivity of LPP to incongruent stimuli (that does not align with expectations or preceding images) has been well documented[104]. Our behavioural results show that expectation violations should be the same across conditions, so this would also not explain the differences we see directly. However, another study that showed LPP enhancement with incongruent endings used source localization to implicate the temporo-parietal junction (part of the MZN) in its generation[102]. They also showed that the N400 is more negative when an image appears for the first time (implying that it plays a role in primary context comprehension and mentalizing), whereas LPP amplitude was the same regardless of the number of times the incongruent stimuli appeared. Source analysis found that the IFG, mPFC and an area adjacent to the TPJ were implicated in the generation of the N400, showing that it is also involved in mentalizing functions. Given the results above we suspect that in the *why* condition, there is a significant and sustained N400 response and that this response diminishes a corresponding LPP that follows. In fact, another study in which participants were asked to think about an artists' intent while looking at paintings, found similar results in which N400 was more negative and LPP was lower for the mentalizing condition as opposed to ones in which they were shown pictures in the absence of any instructions[110]. There may have been a sustained N400 from mentalizing activity that resulted in reduced LPP amplitude following the N400 negative peak. Another interpretation is that the LPP we see represents increased mirroring activity around 400 ms. Face processing centers are found predominantly in areas that are implicated in the MNS (IFG, IPL, STS, MTG)[111–113], and it has also been demonstrated that the LPP is highly selective for facial expressions. This could imply that in *how* conditions, there is a feedback system that amplifies mirroring activity in order to correctly identify actions. It would also indicate that there is an interaction of

LPP with the how condition and faces. This also ties in with our behaviour results as we discuss in the *face-how interaction* section.

We also showed that increasing N400 amplitude is weakly associated with higher task accuracy for the *How Face, Why Face* and *Why Hand* conditions. We had previously shown that the N400 amplitude is higher for faces than hands and higher for the *why* than *how* condition. Because the correlation coefficients were very similar, we are unable to compare across conditions. However, these results do indicate that N400 may be implicated in mentalizing functions, both directly playing a role in processing responses to *why* questions and also in processing faces.

5.2.2 Middle ERP components

An ERP study on lower level goal inferences (part of the mirroring system) reported that goal inferences were made after about 250 ms irrespective of the implicit or explicit instruction[106,114]. The P300 component is split into two components. The P3a is an early attention driven process and the P3b is sensitive to task demand and expectancy[94]. In addition, studies have shown that individuals with ASD have altered P3a and P3b amplitudes relative to controls when differentiating emtions[115]. Moreover it has been demonstrated that tasks with higher cognitive demand reduce P3b amplitude[94]. For our study we find that P3b is lower in the *why* (mentalizing) condition of our task than *how*. This could reflect increased cognitive load in mentalizing processes relative to the mirroring[94].

However, because of the role that expectancy violations play in P300 ERP components, we investigate this as a potential confound. For our photo judgement task, when a photo is presented it can either match the question (answer = yes) or violate the expectation (answer = no). The fact that these components are sensitive to expectation violations prompted us to conduct a behavioural analysis to see how often question stimulus pairings would result in an expected answer of "yes". For example, a question could ask "is this person happy?" and be paired with a photo of a happy person (in which case the expected answer is yes), or a photo of a confused person (in which case the expected answer is no). Figure 16 shows the average ratio of the number of intended "yes" responses over "no" per subject. There is no significant difference between how and why in this ratio, indicating that any differences we see between conditions would not be due to expectation violations.

5.2.3 Early ERPs

Early ERP components are thought to be related to stimuli processing. The N170 is a face-specific parieto-occipital component that occurs within 200 ms poststimuli[116] Our results find that the N170 is higher for face stimuli, which is in line with previous results.. High-arousal faces (e.g., fearful, sad and happy) evoke larger N170 amplitudes than low-arousal faces (i.e., neutral)[85], suggesting an enhanced attentional allocation to arousing/relevant stimuli[117]. One study found that early neural correlates (under 200 ms) support automatic processing that is not under voluntary and are implicated in automatic imitation during strategic games. They also showed that IPL activity (part of the MNS) was higher during this time period and also linked to face and action perception[118]. In addition, there is no difference between averaged *how* vs. *why* conditions in these early ERP components.

5.2.4 Importance of keeping the stimuli the same

We showed that between different stimuli (face vs hands), the waveform appears to be different not only at the N170 component, but at most other time points before and after. In order to see how much the stimuli contributes to differences in ERP relative to varying the social brain condition, we performed some additional comparative analyses. We started by caclulating the peak amplitude for each ERP component per person & electrode. A t-test was used to compare across groups to see how often each comparison was significant. Because we are not using this data to determine specific locations, the results were not corrected for multiple testing. Pairings that had different photos/stimuli (ie. How-Hand vs. How-Face; Why-Hand vs. How-Face; and Why-Hand vs. Why-Face) were significantly more likely to have different peaks ($p<0.05$), as shown in figure 20a than pairings that varied on question condition (*why* vs *how*).

In order to localize *when* the stimili based (*hand* vs *face*) changes were most likely to occur we counted the number of ERP components with significant comparisons ($p<0.05$). Early ERP components (N170 & P200) made up the bulk of the differences that we saw in all electrodes (Figure 22b). These results emphasize the importance of keeping stimuli constant and point to an advantage of using the photo-judgement task.

5.2.5 Face-How Interaction effects.

Using the analyses described above we also show that that the *Why Hand vs. How Face* pairing had a significant outcome more often than the rest of the pairings combined. Coupled with other results we suspect that there may be an interaction between the type of stimuli shown and condition. We have shown that there is an interaction between stimuli and *how/why* conditions of our behavioural results. Results demonstrated that images of faces sped up response time for *how* questions in particular. For this reason we create a contrast by subtracting the ERP waveforms of the *hand* condition from the *face* condition. We then compare the *face minus hand* contrast across social brain conditions (*how* vs. *why).* As expected, we see a number of significant differences between these waveforms indicating that faces may mediate the relationship between MZN and MNS. In particular we note that around 300 ms, activity begins to diverge significantly in left central and left parietal electrodes for this comparison. For example, we showed that the left central and left parietal N400 is selective for faces as the *why* condition has a signficiantly higher N400 peak in the face-hand contrast. The differences in N400 peaks between *how* and *why* conditions are more pronounced in the face-hand contrast than when all conditions are averaged, as evidenced both by the magnitude of the difference and the p-value of the cluster permutation, which was thresholded at 0.0125. Our results also show that *how* condition is assocated with a much higher LPP delta in response to images of faces relative to hands. This matches previous findings that indicate that LPP is activated more strongly in response to facial stimuli, emotional salience and mirroring function. It is important to note, however, that this difference could be a result of carry-over from highly negative potential in the *why (face minus hands)* condition.

5.2.6 Sex based differences

Many studies have revealed that women show a greater empathic attitude and interest for social information than men[119]. There are previous studies that have shown sex-based differences for ERP components in responses to images of faces[91]. These changes were shown in the N170 and P450 components[87,120]. For example, one study showed that females had significantly higher P450 amplitudes to facial expressions than males[91]. However, these differences were specific to types of emotions, with the amplitude and duration varying between happy and sad faces. One study revealed that females showed a more negative amplitude of N170 when discriminating orientations (right or left) of faces while males did not[120]. Another study corroborated these findings and showed that the N170 was significantly more negative for females than males[87]. We have shown that females responded significantly faster to the questions in the task than males, with no changes in accuracy. In order to investigate whether or not these behavioural differences manifest in ERP results we visualized the ERPs of each sex. In conflict with previous results, as shown in figure 24 and 25, females do not appear to have a more negative N170 or higher P450 than males. We did not see any significant differences either; however, it is worth noting that after removing the noisy EEG data, we were left with roughly half the number of males as females. It is likely that there is simply not enough power with the number of trials and subjects we had to show differences at this level.

5.3 Task based oscillatory frequencies

Time frequency analysis allows us to visualize EEG data in the frequency domain. This is important because we know that certain populations of neurons tend to oscillate at particular frequencies. Getting information about what frequencies are associated with regions of interest at different times can shed some light on what subcortical areas may be involved. Although there is a plethora of research on mirroring activity marked by mu (alpha) suppression, there are very few linking mentalizing function to time frequency analysis. The few mentalizing studies that do report a modulation of time-frequency have found changes in beta power following a mentalizing task. In addition, several neuropsychiatric diseases that are characterized by social deficits, such as ASD, schizophrenia and frontotemporal dementia, are associated with alterations in alpha and beta oscillations[50,121,122]. At an aggregate level, there were no significant differences between averaged *how* and *why* conditions for mu or beta suppression at any electrode cluster and time window of interest. These results are interesting given that we saw clear differences in ERP components between conditions. While a few studies have shown mu suppression in other areas too, most show it in central electrodes[123]. In particular, tasks that involve a social components tend to specifically show modulation of mu suppression in central electrodes[70].

In order to look at these results at a more granular level we tried to tease apart differences within the face and hand stimuli groups separately. It could be that mentalizing and mirroring functions process face and hand stimuli differently at the level of distinct neural populations. We found that for the face condition there was heightened mu suppression for *why* questions in right central areas around 500 ms. This relationship did not exist for the hands condition. This may indicate that for stimuli that involve faces, questions about a representative internal state rely more heavily on mirroring functions that are reflected in higher mu suppression at right central electrodes. In contrast for conditions with images of hands we did not see any significant differences between *how* and *why* conditions for right central electrodes. We investigated this area a-priori as one study found that the right central electrodes were particularly sensitive to faces and that the exact power could be used to discriminate between type of facial expression at around 500 ms[124]. In addition, as we have discussed above there are areas of the brain devoted entirely to face processing. The neurons involved with those areas could be oscillating at different frequencies to process information about hands relative to faces.

Studies consistently find that processing intent (mentalizing) places a higher cognitive load on the brain[30,38]. It could be that part of the extra cognitive load when processing intent specifically associated with facial expressions is actually a result of increased mirroring activity. Faces provide a lot of information that is helpful in making judgements about internal states, and our mirror systems have likely associated certain expressions with internal mental states. It is likely that participants have been exposed to specific questions about internal judgments (such as '*is this person admiring someone')* in the context of facial expressions in their lives prior to the study. Consequently, the mirroring system may contain some sort of an action identification model that can link internal states more directly to facial expressions. In comparison, photos of hands don't immediately contain all the information necessary to make these judgements. In turn, it is likely that to do so requires greater mentalizing activity, as the participant would need to (a) interpret the activity, (b) why they might be participating in it, and correspondingly (c) what that could imply about their internal state. If this is true, it could imply that the mentalizing system is more heavily involved in primary context comprehension. Once stimulus-intent associations have been solidified by repeated exposure, it is possible the mirroring system kicks in to interpret the same question. We have already shown above that N400 ERP component, which is heightened when involved in primary context comprehension, is heavily linked to mentalizing activity. This could indicate that the mentalizing system is more active in response to questions of intent paired with images of hands, whereas the mirroring system is more active with intent inferences made on facial expressions. In fact, another study that involved participants judging the intent, emotion and gender of a moving person found similar results[68]. Emotion and gender identification would fall in the *how* category of our photo judgement task. The intent condition on the other hand would fall in our *how* group. They found that mu suppression was larger in the intention than in the emotion and gender conditions, with no difference between the latter two[68]. These results map exactly to ours, however we were able to show that this relationship does not hold up in images of hands.

In addition to the above results, we also investigated mentalizing activity by exploring power changes in beta frequencies. One study found that beta power modulations were associated with a joint attention (JA) task in neurotypicals, however individuals with ASD did not show this relationship. They also were able to show through source analysis that the IPS and temporoparietal regions (implicated in mentalizing) were the main brain areas associated with this beta power modulation[50]. JA tasks are often used as indicators for mentalizing activity, so this indicates that beta power may be indicative of mentalizing abilities[50]. Another study also showed that beta power suppression was associated with an explicit mentalizing activity[68]. Our results indicated that differences in beta suppression between the *how* and *why* condition were only present in response to photos of hands. This fits with the hypothesis that increased mentalizing activity may compensate to fill in information that the mirroring system does not have. As a result, we see that the mirroring system may be able to help the mentalizing system answer *why* questions when it comes to faces, but it may fall short when it comes to photos of hands. In this case, the mentalizing system may need to fill in.

Our time frequency results also indicate that mentalizing activity arises later on. We see that beta power modulations occur around 600 ms whereas the mu suppression occurs at around 400 ms. This ties into our ERP results where face specific mirroring functions appear significantly before the waveforms can be differentiated based on mirroring or mentalizing function. In order to see how exactly the brain transitions between different states for each condition, we use microstate analysis to complement our current analyses.

5.4 Microstate Analysis

The predominant theory of action intent suggests that the MZN functions in a top-down manner, leading to slower but more deliberate processing. The MNS on the other hand functions as a bottom-up processor that acts automatically in the presence of visuomotor stimuli, with a feed-forward pathway to the MZN. One way to investigate this hypothesis is to look into how brain states transition in the mirroring and mentalizing conditions of our photo judgement task. Microstate analysis uses the brains topographic activity through time to identify stable configurations of global activity, as well as the transition periods between them. A “microstate” refers to a momentary, stable global brain state, and is thought to reflect transient information processing in the brain[39,75].

There has been plenty of evidence that microstate activity can be reliably linked to several disease symptoms for a variety of conditions. One class of microstate analysis involves classifying brain states into a few previously identified common topographic states (labelled with letters A through G). Research has shown that varying the amount that each microstate appears is linked with either different task states or psychological conditions. Studies have demonstrated that individuals with ASD lack a certain microstate that is often associated with social communication[125,126]. The fact that these results have been cross validated strongly suggests that microstate analysis can be used to detect both deviant functions of large-scale cortical activities in ASD, and how normal social cognition functions in neurotypicals.

EEG signals represent coordinated electrical activity in groups of neurons. One possibility about the neural substrate of microstates suggests that each microstate comes from a small, local group of neurons that becomes transiently coordinated. However, the topographic maps show that there are very few completely distinct microstates, and each is associated with a well-defined structure that suggests there is global coordination across the entire cortical surface. It is much more likely that microstates emerge from coordinated activity of neural assemblies that span large areas of the cortex. A change in the topographical map represents a change in orientation of underlying active dipoles in the brain that generate that topography, and it is likely that transitions between microstates represent when activation of new neural networks occurs. In this interpretation one of the most illuminating aspects of microstate analysis is the time course at which microstates switch. For this reason, we chose to perform a class of microstate analysis that simply tries to categorize EEG activity into stable and transition states so that we can understand this time course. This analysis does not try to fit each time point into one of the few pre-defined microstates from A-G and simply emphasizes the identification of stable brain states.

When coupled with source reconstruction microstate analysis allows us to know *when*, *where* and *in what* combination the MNS and MZN are being activated for each condition. Our results show that in the *how* (mirroring) condition, there are only 4 distinct microstates, while in the *why* (mentalizing) condition there are 9. We also see that the time course of the first 4 microstates in the *why* condition overlap almost perfectly with the first 3 microstates in the *how* condition. Microstate 3 for both start at the same time, while microstate 4 in the mentalizing condition ends at the same time as microstate 3 in the mirroring condition. We can interpret this as some additional activity that results in an extra state transition for mentalizing functions that is not present in the mirroring condition. In particular, it may indicate feed-forward information from the MNS to the MZN that does not produce significant changes in source activation but can be picked up by the microstate analysis as a unique state. The last microstate in the *how* condition stays stable for an extended period of time during which time there are 5 unique stable microstates in the *why* condition. This suggests that mentalizing may recruit several other brain areas sequentially to process information about action intent. One study had shown that in mentalizing tasks there is increased cognitive load for tasks that require explicit mentalization as evidenced by graph theory metrics such as modularity. Many other studies have shown that the mirroring system is automatically activated early on, whereas the mentalizing system is elicited in later stages. Our results support these findings as microstate transitions occur in the same time frames for both *how* and *why* conditions soon after stimulus presentation and diverge later on. In addition, the rapid sequence of changing state transitions in the mentalizing task may indicate that there is higher cognitive load and more processing pathways involved.

5.5 Source Analysis

Understanding which regions are activated through time can provide insight on how the MZN and MNS interact. A minimum norm estimate provides a conservative estimate of source activation in the absence of a-priori hypotheses. We are unable to localize to deep subcortical areas with this method, so we estimate the active regions based on surface maps. Another important note is that there is always some level of activity in all areas of the brain. We set a threshold so that only the strongest sources were visible and refer to these when we talk about source activity. In line with the rest of our results, we see that the how and why conditions activate the same sources for the first 300 ms. In particular we see that activity predominantly in the occipital lobe for the first 2 microstates (first 188 seconds). At 227 seconds, for all the conditions we see that there is additional activity in what appears to be the right inferior frontal gyrus (IFG). The IFG has been implicated as one of the key regions of the MNS. In fact, one study even showed that disrupting activity in the inferior frontal gyrus using transcranial magnetic stimulation increased reaction times to an emotion recognition task and eliminated mu suppression, which is an index for mirroring activity[127].

Looking specifically at the *how face* condition, we see that although the first 3 microstates involve the MNS, by the fourth microstate, the mPFC, a key region of the mentalizing system, is also active. Although this activity is smaller in magnitude and breadth than mPFC activity in the *why* condition, this indicates that the mentalizing system may be automatically activated in the presence of social stimuli. In order to make a correct comparison between the two conditions, we create a contrast and visualize those sources in Figure 34. In line with the rest of our results, we see that the how and why conditions activate the same sources for the first 212 ms at-least (3 microstates). At the next microstate (4th microstate, 397 s), activity is higher activity in the occipital lobe, left precentral gyrus, left postcentral gyrus and left superior temporal gyrus (STG) for the *how* condition than the why. These are all regions that have been associated with the MNS. Mirroring activity in the precentral gyrus is well documented[128]. In addition, autistic patients have lower activity in the STG, which several studies have found to have mirroring properties[128]. In particular one study found that STG activity with ASD was impaired in the left hemisphere specifically[129], which ties in well with the fact that we also see a left hemispheric lateralization. Moving on to the 5th microstate (742 s), we also observe that all regions that display *how>why* activity, are associated with the mirroring system and are more left lateralized. However, we also see that the mPFC has higher activity for the *why* than *how* condition in this microstate. This indicates that although mPFC activity begins around 350 seconds after the presentation of facial stimuli, activity in the *why* condition only supersedes that of the *how* condition at a later time point. This may indicate that there is both an automatic component of mentalizing in the presence of facial stimuli, and a more deliberate component that begins later on, that is associated with answering the task questions about intent.

Investigating the aggregate *how* and *why* contrast for all 9 microstates in the *why* condition, we have the same findings from the *how face – why face* contrast. The first 4 microstates show no differences between conditions. For microstates 5-8 (344-458s), activity is predominantly higher in the *how* condition and all regions are associated with the mirroring system. For microstates 8-9 (611-875), there is increased activity in the mPFC, part of the MZN, for the *why* condition. In line with the rest of our results, the regions that are more active in the *how* condition are all part of the MNS. These results are also in line with previous findings of condition contrasts in the photo judgement task. Because of the low spatial resolution offered by EEGs, we are limited in our ability to identify sources at a granular level. This may be why we don't see specific changes in sources for some sequential microstates. While the source plots do look different in the magnitude and precise locations of larger brain areas identified across microstates, in the interest of validity, we refrain from making distinctions at this level.

5.6 Limitations

This study is not without limitations. First and foremost, we must recognize that EEGs are notoriously sensitive to noise. We had to eliminate 10 subjects from our EEG findings due to issues in data collection, which reduces the power by increasing susceptibility to this noise. Another problem is that there is a plethora of EEG functional connectivity metrics that can be used and a tremendous amount of variability on which are selected between experiments. Studies that have explored our research questions with EEG have used a large range of analyses from mu suppression index to graph theory metrics. However, there is variability even within a particular method; for example, studies that used microstate analysis on individuals with ASD used a different variant that identifies ratios of pre-defined states that are occupied. Our analysis is different in that we are concerned simply with when stable and transition states occur. On one hand, using robust EEG functional connectivity methods that are consistent between experiments can increase the power of findings. However, there is a trade-off between exploring new questions and validating answers to previous ones. We conducted a combination of both in order to be confident in our results, while also exploring new questions.

Another methodological concern with our preprocessing concerns using PCA to limit the number of dimensions before applying ICA to identify and remove artifacts. One study has shown that doing so can adversely affect both the number of independent components and their stability under repeated decomposition[59]. However, given our choice to use high density EEG, doing an ICA without a PCA was not feasible. Not only would this drastically increase computational time, but correctly identifying each artifact component when they are split into numerous sub-components would not be possible. Another alternative to removing epochs that contain artifacts would reduce our power so significantly that it would render our results meaningless. Moreover, in light of our findings, we suspect that this limitation would not change the interpretation of our results as it would equally affect all conditions that we compared.

Furthermore, there are a few limitations regarding source analysis. Firstly, proper source analysis relies on accurate head models and cap placement. We did not collect fMRI data or have skull measurements, so getting an accurate head model is difficult.

Consequently, we are reliant on correct and consistent cap placements across all the subjects. In addition, as described above, minimum norm estimates cannot localize deep sources and have low spatial resolution. In order to limit these concerns, we are conservative in our estimates of observed source activity.

Regarding our conclusions about face specific effects, it is possible that there are different contexts and associations with the particular images that we selected for hands vs. faces. It could be that the differences we find related to face specific effects (in ERP, TFA and behavior) are related to these different contexts or associations. However, this is unlikely as many of our results match previous findings.

6 Conclusion

Aim 1: *Do we detect reliable differences in event related potentials (ERPs) between mirroring and mentalizing conditions that align with previous results?*

Our results show that there are clear differences in *how* (mirroring) and *why* (mentalizing) conditions that can be detected by EEG. Brain activity appears to diverge around 300 ms after stimulus onset. For the first ~300 ms, ERPs, time frequency, microstate and source analyses all lend evidence to the idea that brain activity is largely the same regardless of whether the participant is instructed to think of intent or means. This supports the hypothesis that the mentalizing task is a higher order system that follows and relies on mirroring processes. Once the waveforms begin to diverge, mentalizing is associated with a higher (negative) N400 peak in central electrodes. In line with previous literature, we believe that N400 amplitude can be used as a marker for mentalizing activity. The N400 peak was also correlated with task accuracy for the *why hand, why face, how face* conditions. *Why* questions were used as a proxy for mentalizing, so it is expected that the *why face* and *why hand* conditions should both be correlated with N400. However, many studies, including our own show that faces automatically induces the MZN even when evaluating action means, such that *how face* condition is also correlated with N400. We also note that LPP (a sustained positive peak that follows N400) is higher for *how* conditions, in particular for *how face.* We suspect that LPP is indicative of both emotional salience and mirroring functions.

Aim 2: *How are mirroring and mentalizing systems modulated by photos of faces vs. hands?*

Face and hand stimuli are associated with several differences in ERP components. In particular, we see that the face condition is associated with an increased N170 peak, which fits with the literature on face processing. It also seems that faces also mediate both LPP and N400 activity. We find that photos of faces increase LPP amplitude for the mirroring condition and N400 amplitude for the mentalizing condition. Moreover, the type of stimulus also modulates behaviour, such that individuals respond faster to questions associated with pictures of faces and the fastest to *how face* questions in particular. As mirroring activity precedes mentalizing, we can expect that questions more heavily reliant on the MNS require less time to process. Since many face processing areas in the brain are linked directly to the MNS, it is possible that this confers an added benefit in response time.

Aim 3: *What particular frequencies may be implicated in the differences we see in ERPs? Can mu and beta suppression provide some insight in the contributions of the MNS and MZN respectively for each condition?*

We are able to show that for faces in particular mu suppression (an index for mirroring activity) is higher in the *how* than *why* condition around 400ms. However, no difference was found in response to photos of hands for mu suppression. This may indicate that the MNS is more active when processing action means for faces than hands.

In addition, we showed that beta suppression (associated with mentalizing functions) is higher for *why* than *how* in response to photos of hands around 600 ms. We did not see any relationship between the two conditions using beta suppression in response to photos of faces. This implies that photos of hands specifically require significantly higher mentalizing processes to interpret intent and is in line with previous studies that show the mentalizing system is more active when subjects must answer questions about intent in the absence of detailed information.

Aim 4: *How do brain-states transition in mentalizing vs. mirroring conditions? How many stable states are achieved in each task and what are the cortical sources of each stable state?*

Using microstate analysis, we are able to show that ~300 ms after stimulus presentation the brain undergoes several state transitions while processing intent (mentalizing). Conversely, while processing action means (mirroring), after ~300 ms the brain appears to stay in one stable state for an extended period of time. Investigating source space results shows that indeed the *how* condition activates mirroring areas more strongly. Likewise, the *why* condition activates mentalizing regions more strongly.

6.1 Future Work

In order to further investigate whether or not the mirroring system precedes and informs the mentalizing system, it is important to investigate direction of information transfer between brain systems in both. In addition, graph theory analysis can provide some insight into how cognitive load and connectivity patterns change across time for both conditions. Exploring how all these measures are altered in patients who have disrupted mentalizing or mirroring function can provide some insight into the interplay between the two systems. This is a promising field that holds many unanswered questions with insightful directions to pursue them in.

Overall Summary

Background: Human brains are shaped by their interactions with other people, which can be both non-verbal (actions, gestures, posture and expressions), and verbal[130]. Interpreting these cues is critical for motor learning and social cognition. Mirror neurons, which respond to both the observation and execution of an action[1–3] are found in the ventral premotor cortex, inferior parietal lobe[131] and the superior temporal sulcus[132]. These regions form a complex network in which the visual representation of motion fashions a corresponding motor representation. Another essential feature of successful social functioning is the ability to reason about the minds of others, by inferring not just *what* they are doing but *why* they are doing it[4]. The mentalizing system (MZN) is implicated in the process of inferring other's mental states and includes the temporoparietal junction and the ventromedial prefrontal cortex[5]. To date most studies of the mirror neuron system (MNS) and MZN have investigated each system independently. In order to understand whether deficits in social functioning arise from disruptions in the MNS or the MZN, it is important to be able to better delineate the two. Previous studies leave three gaps that we explore: 1) How the MNS and MZN differ within the same subjects 2) How the timing of activation patterns differs between the two systems and 3) How EEGs can be used to investigate these questions as opposed to fMRI which is frequently used to study the MZN and MNS.

Objectives: We are interested in determining whether or not we can detect reliable differences in event related potentials (ERPs) between mirroring and mentalizing conditions that align with previous results. We also consider whether measuring mu and beta suppression can provide some insight in the contributions of the MNS and MZN respectively for each condition. Moreover, we investigate if the mirroring and mentalizing systems are modulated by photos of faces vs. hands. Lastly, we aim to add to the literature by inspecting how brain-state transition in the mentalizing vs. mirroring conditions. For each of the stable brain states, we would like to know what system and sources contribute most strongly.

Methods: This study utilizes high-density EEG (hd-EEG) to quantify dynamics in functional brain networks supporting mirroring and mentalizing processes in neurotypical adults. We use a task that has been previously shown to differentiate the mirroring and mentalizing activity with fMRI. Participants as shown pictures of faces or hands and asked pictures about *how* (mirroring) or *why* (mentalizing) the actions in the photos are being performed. Classic EEG methods such as ERP are used to confirm that EEG can detect differences between the two systems. Time Frequency analysis is then used to characterize these differences within frequency bands. In order to characterize the interplay between the two systems in time, microstate analysis is used. Lastly, source reconstruction is performed to identify the regions implicated in each condition along with the timing of their activations.

Results: We show that there are clear differences in how (mirroring) and why (mentalizing) conditions that can be detected by EEG. Brain activity appears to diverge around 300 ms after stimulus onset. Once the waveforms are distinguishable, mentalizing is associated with a higher (negative) N400 peak in central electrodes. The N400 peak was also correlated with task accuracy for the why hand, why face, how face conditions. We also note that LPP (a sustained positive peak that follows N400) is higher for when questioned about action means and in particular when paired with images of faces. The nature of the stimulus (face or hands) used is associated with several differences in ERP components. In particular, we see that the face condition is associated with an increased N170 peak, which fits with the literature on face processing. It also seems that faces also mediate both LPP and N400 activity. We find that photos of faces increase LPP amplitude for the mirroring condition and N400 amplitude for the mentalizing condition. Moreover, the type of stimulus also modulates behaviour, such that individuals respond faster to questions associated with pictures of faces and the significantly faster to *how face* questions in particular. As mirroring activity precedes mentalizing, we can expect that questions more heavily reliant on the MNS require less time to process. Since many face processing areas in the brain are linked directly to the MNS, it is possible that this confers an added benefit to response time. We are able to show that for faces in particular mu suppression (an index for mirroring activity) is higher in the *how* than *why* condition around 400ms. However, no difference was found for mu suppression in response to photos of hands. This may indicate that the MNS is more active when processing action means for faces than hands. In addition, we showed that beta suppression (associated with mentalizing functions) is higher for *why* than *how* conditions in response to photos of hands around 600 ms. Using microstate analysis, we are able to show that ~300 ms after stimulus presentation the brain undergoes several state transitions while processing intent. Conversely, while processing action means (mirroring), after ~300 ms the brain appears to stay in one stable state for an extended period of time. Investigating source space results shows that indeed the how condition activates mirroring areas more strongly. Likewise, the why condition activates mentalizing regions more strongly.

Significance: Exploring the MNS and MZN together can help us understand the precise functions of and differences between each. Exploring how all these measures are altered in patients who have disrupted mentalizing or mirroring function can provide some insight into the interplay between the two systems. This is a promising field that holds many unanswered questions with insightful directions to pursue them in.

Appendix A

Filtering EEG data:

Specifications	Band-Pass Filter	Notch-Filter
Toolbox	EEGlab	ERPlab
Filter Type	FIR filter [133]	Parks-McClellan Notch
Frequencies Filtered (Hz)	0.05-30	60 Hz
Filter Order	3*Sampling Rate/Lower Cutoff	180
Additional notes		Remove DC offset

Calculating Peak Amplitude:

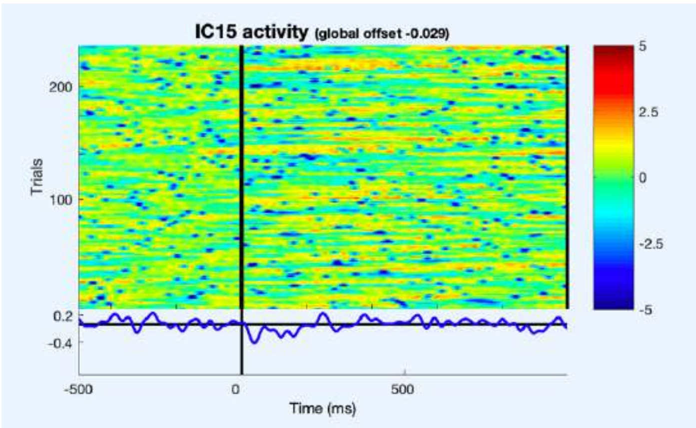
Peak amplitude was calculated for each person & ERP component by obtaining the peak latency (either positive or negative peak), within a certain time frame, and calculating the amplitude at that point. T-Tests were then constructed for each ERP component by taking the peak amplitudes and comparing across each condition.

Appendix B

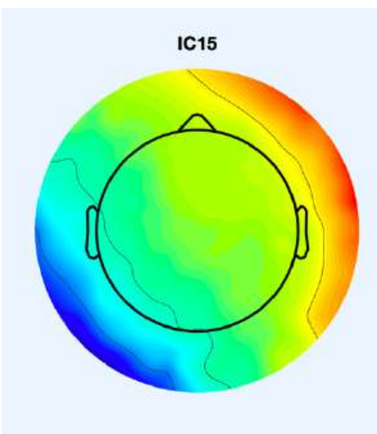
As a general rule, finding stable components from N channels typically requires far more than N^2 data sample points[59,60]. Because our data was too short for this given the high number of electrodes, we used ICA with PCA to narrow it down to 32 components. Below we outline different things that we look for when identifying whether each of the 32 components are due to noise or brain activity.

EKG Artifact:

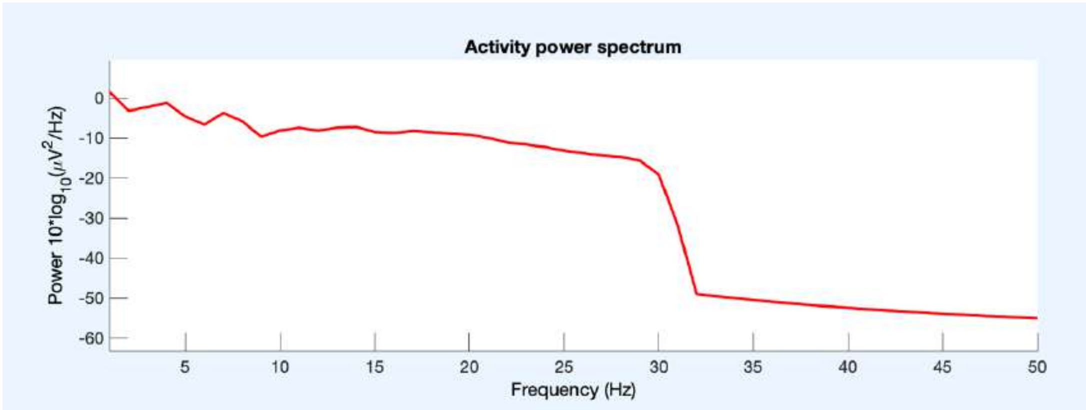
For heartbeats, we see electrical activity peak periodically every ~1000ms as in appendix figure 1 a (can range from 500 to 1200 ms). The topographic maps show a gradient that looks similar to appendix figure 1b below. In addition, the power spectrum activity has peaks around 4 and 6 Hz, although the main way to identify that there is no brain activity mixed into the signal is to confirm that there is no strong alpha peak (see appendix figure 1c). The last way to confirm that this is the right component is to plot the EEG that is reconstructed after the specified component is removed on top of the original EEG as in appendix figure 1d. The reconstructed signal should be very close to the original EEG except for where you expect the deflections from the specific artifact to occur.



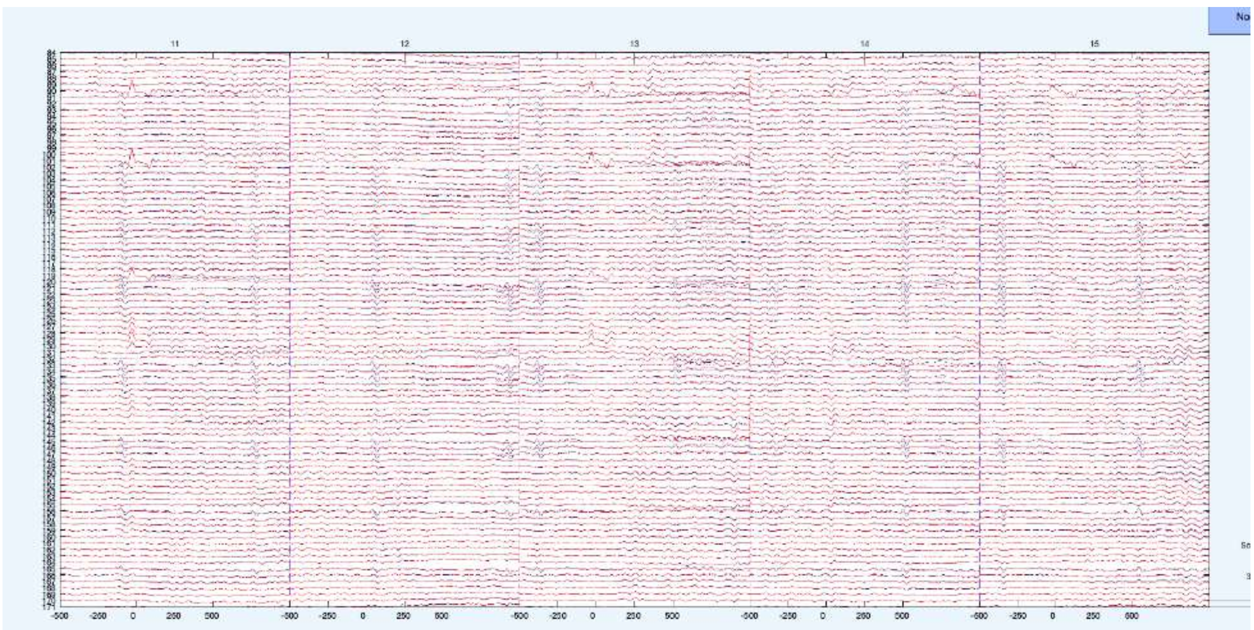
Appendix Figure 1a: Electrical Activity in Time



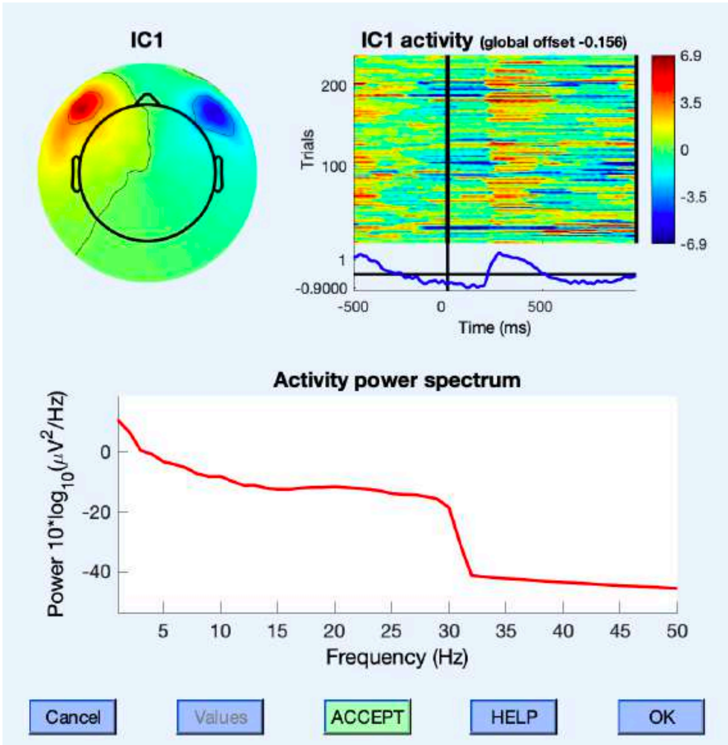
Appendix Figure 1b: Topographic Map of Activity



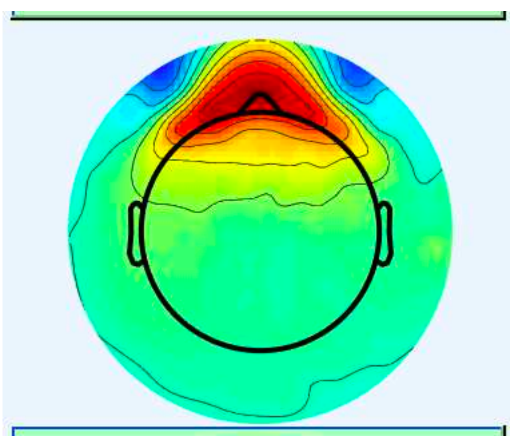
Appendix Figure 1c: Power Spectrum Map



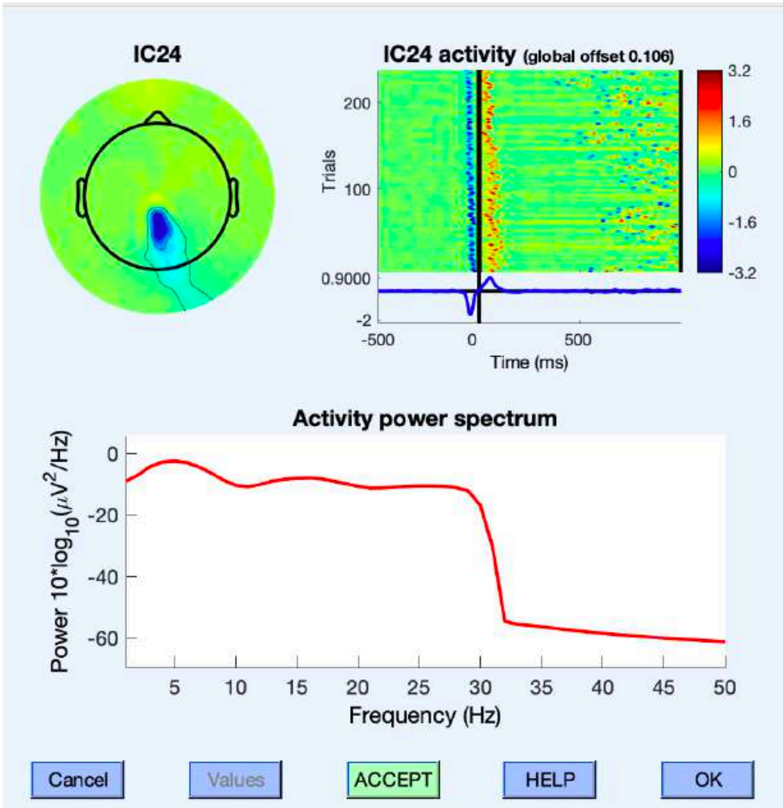
Appendix Figure 1d: EEG signal plotted against component removed EEG signal. The original signal is in blue and the reconstructed signal is in red.



Appendix Figure 2 Typical ICA patterns for horizontal eye movements.



Appendix Figure 3 ICA topography for eyeblinks



Appendix Figure 4 Button Press artifact centered around button press events only

The component specific attributes were validated based on previous research and all 4 attributes were inspected before removing any components.

Appendix C

The specific parameters that we used for cluster permutation are outlined below. Cluster-based permutation tests do not, however, control the false alarm rate at the level of the (channel, frequency, time)-triplets or the (channel, frequency)-pairs[62]. According to a review of permutation tests by Groppe et al., (2007):

"It is important to note that because p values are derived from cluster level statistics, the p value of a cluster may not be representative of any single member of that cluster. For example, if the p value for a cluster is 5%, one cannot be 95% certain that any single member of that cluster is itself significant […]. One is only 95% certain that there is some effect in the data. Technically, this means that cluster-based tests provide only weak [family-wise error rate] control"[67].

Cluster Permutation Parameters

Specification	Description	Selected Value
cfg.latency	Whatever time range you expect to see changes in (in seconds after event onset)	Variable
cfg.frequency	The frequency(s) at which we expect to see differences	Variable

cfg.method	We select monte carlo stimulation to sample several times and increase the accuracy.	'montecarlo'
cfg.statistic	What statistical method is used? The one we selected calculates the dependent samples T-statistic	'ft_statfun_depsamplesT'
cfg.correctm	The method for correction	'cluster'
cfg.clusteralpha	The alpha threshold for each cluster	0.025 or lower
cfg.minnbchan	The minimum number of channels in a cluster	2
cfg.alpha	The overall alpha threshold	0.025
cfg.numrandomization	The number of randomizations in the monte carlo stimulation. Over 1000 is ideal	1000
cfg_neighb.method	This parameter specifies how to construct the neighborhood. Triangulation selects the nearest	'distance'

	direct neighbors whereas distance selects the electrodes within a 3-D Euclidean distance	

Time Frequency Parameters

Specification	Description	Selected Value
cfg.t_ftimwin	Decreasing this number increases temporal resolution at the cost of frequency resolution. This specifies the sliding time window in seconds.	0.4
cfg.epochbaseint	Specifies the baseline interval for the data	-0.50
cfg.continuous	If the data is epoched it is not continuous.	'no'
cfg.taper	Decides whether or not frequency analysis is performed using a single taper (hanning) or multiple.	'hanning'
cfg.method	Different methods of calculating the spectra. The option 'mtmconvol'	'mtmconvol'

	implements multitaper time-frequency transformation based on multiplication in the frequency domain.	
cfg.foi	Frequency of interest (use frequencies at which data was filtered)	0.5:0.5:30
cfg.toi	Times of interest (epoch length)	0.5:0.01:1.5;
cfg.pad	The length in seconds to which the data can be padded out. The padding will determine your spectral resolution. The option 'nextpow2' rounds the maximum trial length up to the next power of 2. By using that amount of padding, the FFT can be computed more efficiently.	'nextpow2'

Potential Source Analysis methods:

Current Density Estimates:

- Least-squares minimum norm estimate (MNE): MNE is favored for analyzing evoked responses and for tracking the wide-spread activation over time[80,81].

- Dynamical statistical parametric mapping (dSPM): Uses minimum-norm inverse maps weighted by estimates of noise at that location[80,82]. The noise is obtained by applying the inverse operator to the signal covariance matrix[80]. dSPM can localize deeper sources more accurately than standard minimum norm procedures, but the spatial resolution remains low[82].
- Low resolution brain electromagnetic tomography (LORETA): Also uses minimum-norm inverse maps weighted by estimates of noise, however the noise is obtained by using the diagonals of the model resolution matrix[80]

Beamformers:

- Linear constrained minimum variance (LCMV): in the time domain[83]
- Dynamic imaging of coherent sources (DICS): in the frequency domain[83]

Source Analysis Parameters

Specification	Description	Selected Value
MRI	Need a template MRI for the source analysis from fieldtrip	'standard_mri' in fieldtrip's templates folder
cfg.method	Method of source analysis used	'mne'

cfg.mne.scalesourcecov	Scale the source covariance matrix	'yes'
cfg.grid	The grid specifies the forward model that was created.	'leadfield' or whatever you named the forward model
cfg.lambda	A regularization parameter	3

Appendix D

In order to investigate whether or not there is an interaction between the stimulus and mentalizing or mirroring condition we include a 2-way ANOVA for reaction time and accuracy separately.

Results from the ANOVA for Accuracy:

	F value	Num df	Den df	Pr>f
Condition (how/why)	6.82	1	37	0.013
Stimulus (face/hands)	10.33	1	37	0.003
Cond:stim	2.14	1	37	0.152

The interaction is not significant.

Results from the ANOVA for Reaction Time:

	F value	Num df	Den df	Pr>f
Condition (how/why)	59.55	1	37	0
Stimulus (face/hands)	79.30	1	37	0
Cond:stim	10.47	1	37	0.003

The interaction is significant which means that the stimulus MEDIATES reaction time in each condition.

To determine how sex plays a role in reaction times we do separate analyses for each condition.

Results from the ANOVA for Reaction Time using only hand stimuli:

	Sum_sq	Df	F	Pr(>f)
C(how/why)	13.27	1.00	124.03	3.80E-28
C(sex)	1.51	1.00	14.15	1.73E-04
C(how/why):c(sex)	0.03	1.00	0.24	0.62
Residual	266.12	2,488	Nan	Nan

Results from the ANOVA for Reaction Time using only face stimuli:

	Sum_sq	Df	F	Pr(>f)
C(how/why)	2.93	1.00	25.99	3.69E-07
C(sex)	0.90	1.00	7.95	4.84E-03
C(how/why):c(sex)	0.04	1.00	0.38	0.54
Residual	279.02	2,478	NaN	NaN

Sex has a very significant effect on reaction time, but the interaction does not. Now we can run t-tests to see which specific groups have significant differences between them.

How Face: 0.012*

Why Face: 0.005*

To determine how sex plays a role in accuracy we do separate analyses for each condition.

Results from the ANOVA for accuracy using only hand stimuli:

	Sum_sq	Df	F	Pr(>f)

C(how/why)	0.002	1.00	0.48	0.49
C(sex)	0.001	1.00	0.23	0.64
C(how/why):c(sex)	0.01	1.00	1.67	0.20
Residual	0.32	72	NaN	NaN

Results from the ANOVA for accuracy using only face stimuli:

	Sum_sq	Df	F	Pr(>f)
C(how/why)	0.02	1.00	7.85	0.01
C(sex)	0.002	1.00	0.67	0.42
C(how/why):c(sex)	0.003	1.00	1.27	0.26
Residual	0.19	72.00	NaN	NaN

Sex does not affect accuracy, nor does it mediate the relationship between question condition and accuracy.

Appendix E

The following figure outlines how the face minus hands contrast is constructed.

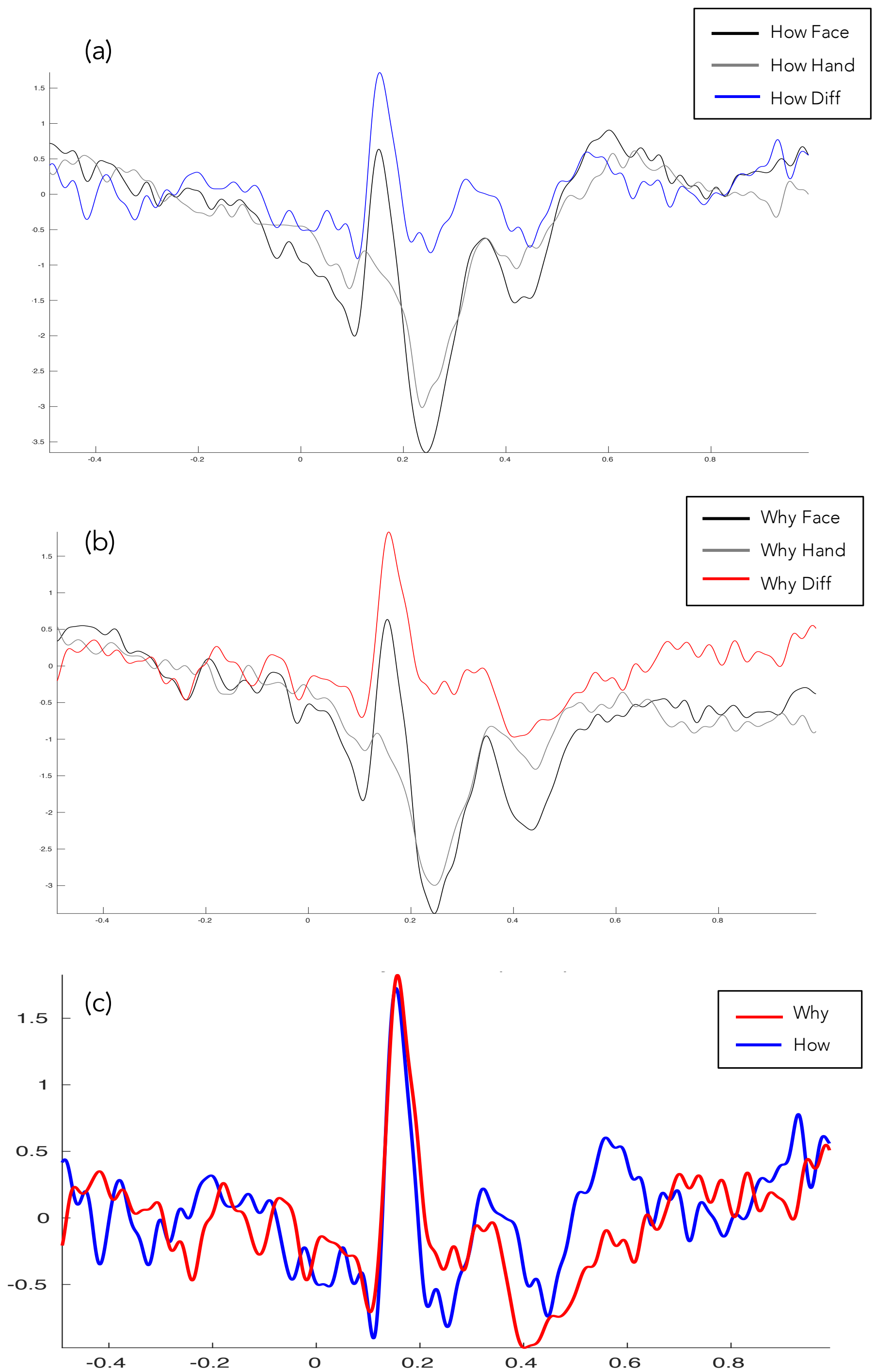


Figure 6: Figure (a) shows the How Face ERP as a black line and the How Hand ERP in light grey. The How contrast is constructed by subtracting How Hand from How Face and the resulting ERP is shown in blue. (b) shows the Why

Face ERP as a black line and the Why Hand ERP in light grey. The Why contrast is constructed by subtracting Why Hand from Why Face and the resulting ERP is shown in red. (c) Shows the 2 contrasts for Why and How in red and blue respectively.

Appendix F

In order to investigate whether or not there are any differences in time frequency across frontal electrodes, we plot them below. Statistical analysis shows that there is no difference between any of the conditions.

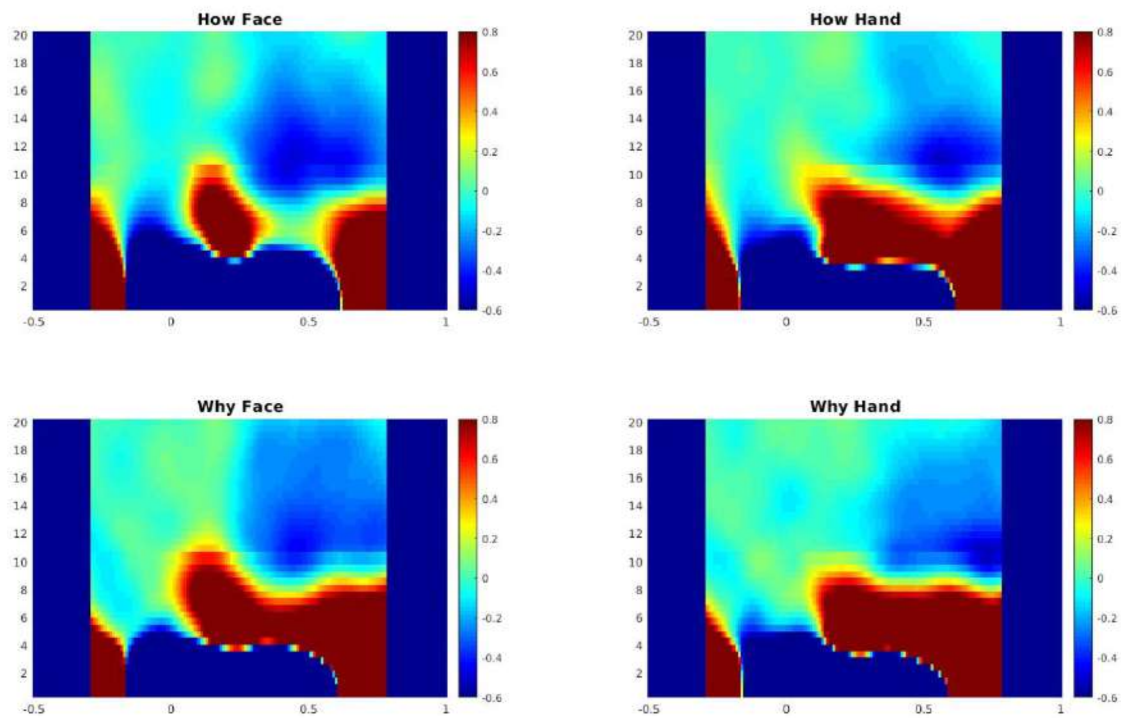


Figure 7: Time frequency analysis averaged across frontal electrodes.

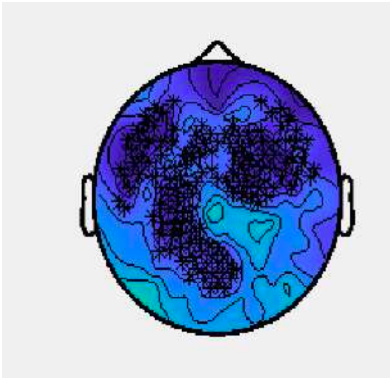


Figure 8: A cluster permutation analysis compares differences in mu suppression for face and hand condtions at 400 ms.

We also explore differences in mu suppression between face and hands. There are numerous differences, which shows that mu suppression is likely heavily linked to the type of stimulus. As these regions weren't associated with our hypotheses so we do not investigate them further.

Appendix G

The following 3 charts show the maximum and average global field potentials (GFP), as well as the start and end of each state. The GFP represents the strength of the electric field over the brain at each instant, and so is often used to measure the global brain response to an event or to characterize rapid changes in brain activity[134].

How:

	Baseline	State 1	State 2	State 3	State 4
Start	-490	83	134	177	269
End (n+1)	0	110	156	247	1000
Max GFP	1.0000	2.4427	1.6873	4.1347	3.8036
Avg GFP	0.6010	2.3128	1.4581	3.4859	2.0168
Stddev GFP	0.2494	0.1400	0.1061	0.6030	0.6340

Why:

	Baseline	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9
Start	-490	84	135	176	214	331	371	444	486	750
End (n+1)	0	110	154	200	239	357	433	472	736	1000
Max GFP	0.9683	2.4722	1.7685	3.3611	4.2590	2.8409	2.6767	2.5232	2.4120	2.3713
Avg GFP	0.5385	2.3355	1.5797	3.0027	4.1705	2.7243	2.5727	2.4753	2.3101	2.0062
Stddev GFP	0.2263	0.1434	0.0839	0.2392	0.1055	0.0442	0.0513	0.0152	0.0560	0.2471

All:

	Baseline	State 1	State 2	State 3	State 4	State 5
Start	-490	84	135	176	326	483
End (n+1)	0	110	155	248	468	1000
Max GFP	0.9758	2.4543	1.7100	4.1859	2.8769	2.3382
Avg GFP	0.5624	2.3287	1.5122	3.5785	2.5534	1.9249
Stddev GFP	0.2323	0.1356	0.0891	0.5833	0.0950	0.3414

The cosine distance between template maps can provide information on how different each of the states are from one another[73]. The standard deviation and confidence intervals add to this information and are shown below:

HOW: Cosine distance between template maps

	State 1	State 2	State 3	State 4
Baseline	0.10	0.20	0.08	0.28
State 1	-	0.05	0.01	0.08
State 2	-	-	0.09	0.06
State 3	-	-	-	0.09
State 4	-	-	-	-

HOW: Standard deviation of cosine distances of topomaps in each template map

	Template Magnitude	Standard Deviation	Confidence Interval
Baseline	1.79	1.229	2.41
State 1	369	0.003	0.005
State 2	232	0.003	0.007
State 3	555	0.006	0.012
State 4	304	0.105	0.206

WHY: Cosine distance between template maps

	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9
Baseline	0.26	0.36	0.30	0.25	0.36	0.32	0.31	0.49	0.65
State 1	-	0.04	0.02	0.02	0.03	0.02	0.03	0.10	0.34
State 2	-	-	0.07	0.10	0.07	0.05	0.05	0.06	0.25
State 3	-	-	-	0.01	0.01	0.02	0.04	0.13	0.42
State 4	-	-	-	-	0.03	0.02	0.04	0.14	0.42
State 5	-	-	-	-	-	0.01	0.03	0.09	0.37
State 6	-	-	-	-	-	-	0.00	0.06	0.29
State 7	-	-	-	-	-	-	-	0.04	0.24
State 8	-	-	-	-	-	-	-	-	0.11
State 9	-	-	-	-	-	-	-	-	-

WHY: Standard deviation of cosine distances of topomaps in each template map

	Template Magnitude	Standard Deviation	Confidence Interval

Baseline	1.09	1.166	2.285
State 1	372	0.003	0.006
State 2	252	0.002	0.004
State 3	479	0.002	0.003
State 4	666	2E-04	4E-04
State 5	435	0.001	0.001
State 6	410	0.002	0.004
State 7	395	0.001	0.001
State 8	363	0.019	0.038
State 9	318	0.010	0.020

Appendix H

Cluster	Anatomical region
1	Left inferior frontal gyrus Left superior parietal lobule Left inferior parietal lobule Left superior temporal gyrus
2	Right precuneus Right inferior parietal lobule Right postcentral gyrus Right insula Right superior parietal lobule
3	Right inferior frontal gyrus Right middle frontal gyrus
4	Left inferior temporal gyrus
5	Left middle frontal gyrus
6	Left cerebellum
7	Right middle temporal gyrus Right inferior temporal gyrus
8	Right cingulate gyrus
9	Left insula
10	Left superior parietal lobule
11	Right inferior frontal gyrus
12	Right cerebellum
13	Left cingulate gyrus
14	Left medial frontal gyrus

Anatomical clusters of the MNS from a large meta-analysis[9] that identified significant anatomical clusters associated with the MNS in fMRI studies are shown in the table to the left.

Sources for the *why hand* and *how hand* conditions at the mid-point of each of the corresponding microstates are shown below.

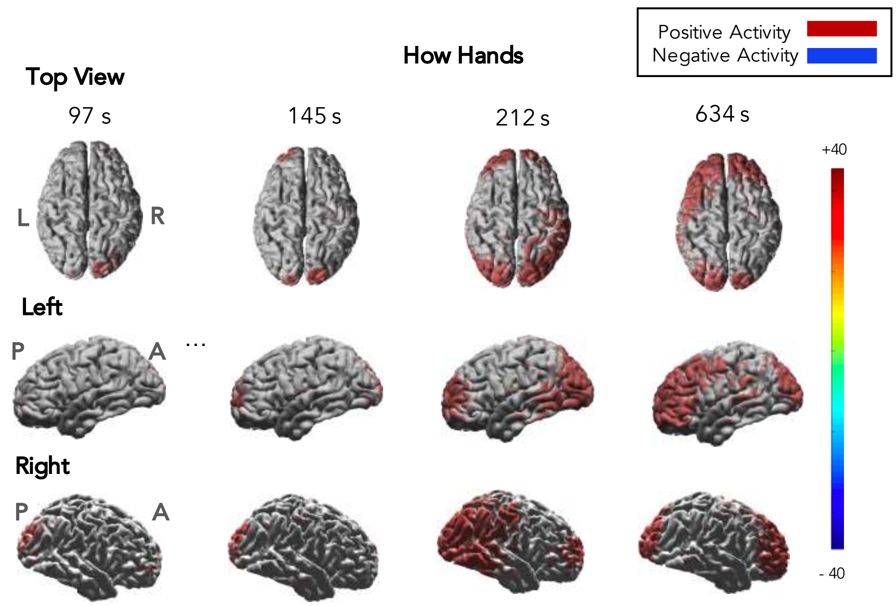


Figure 9: Source analysis for the how hand condition for each microstate.

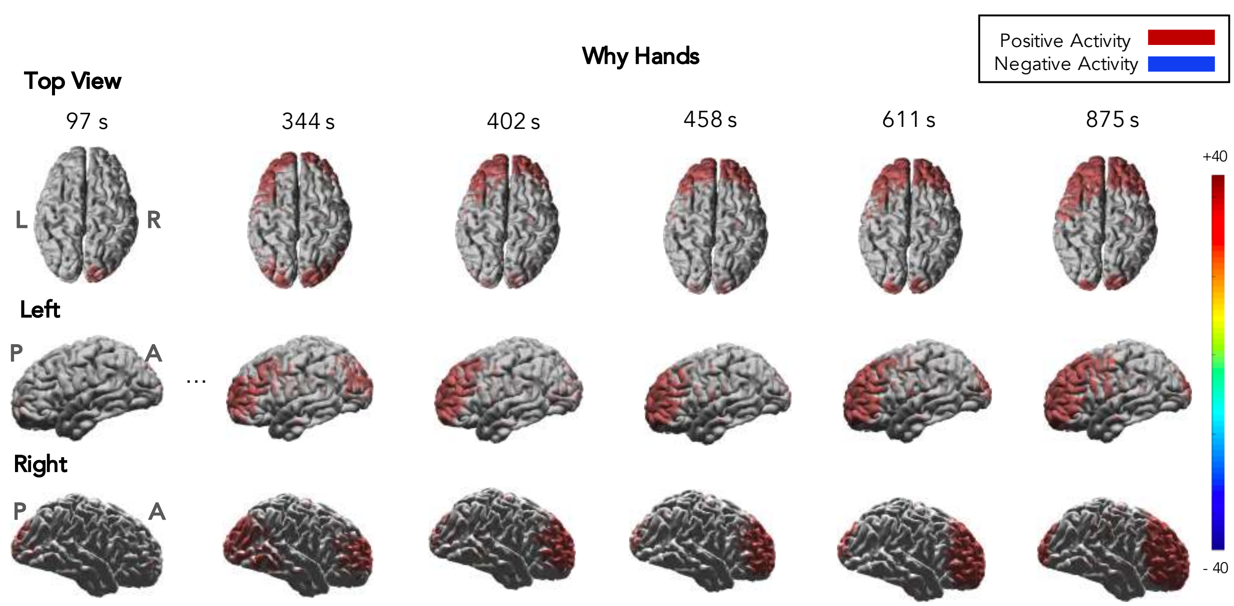


Figure 10: Source analysis for the why hand condition for each corresponding microstate.

Appendix I

__Future analyses to investigate automaticity of the MZN and MNS.__

A previous study showed that the MZN was modulated by cognitive load, whereas the MNS was not. An independent study has shown that there are measurable changes in brain topology (using EEG) in response to increasing cognitive load[89]. Using graph theory, they were able to show that greater cognitive effort resulted in a more globally efficient, less clustered, and less modular network configuration, with more long-distance synchronization between brain regions. The MNS appears to act automatically while the MZN seems to show increased cognitive load if specifically triggered[30]. We should therefore expect that graph theory measures of MZN activity show increased global efficiency, and reduced clustering or modularity. Both the tasks in this study can have graph theory measures applied on the ERP data to see if there are differences in the measures of interest.

Works Cited

1. Spapé, M. M. *et al.* Keep your opponents close: Social context affects EEG and fEMG linkage in a turn-based computer game. *PLoS One* (2013). doi:10.1371/journal.pone.0078795

2. Schippers, M. B. *et al.* Mapping the information flow from one brain to another during gestural communication. *Source Proc. Natl. Acad. Sci. United States Am.* **107**, 9388–9393 (2010).

3. Frith, C. D. & Hasson, U. Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *R. Soc. Philosphical Trans. B* **371**, 20150366 (2016).

4. Koster-Hale, J. *et al.* Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *Neuroimage* **161**, 9–18 (2017).

5. Atique, B., Erb, M., Gharabaghi, A., Grodd, W. & Anders, S. Task-specific activity and connectivity within the mentalizing network during emotion and intention mentalizing. *Neuroimage* **55**, 1899–1911 (2011).

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

7. Ramachandran, V. S. *The tell-tale brain: A neuroscientist's quest for what makes us human*. (WW Norton & Company, 2012).

8. Gallese, V. Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification. *Psychoanal. Dialogues* **19**, 519–536 (2009).

9. Molenberghs, P., Cunnington, R. & Mattingley, J. B. Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neurosci. Biobehav. Rev.* **36**, 341–349 (2012).

10. Libero, L. E. *et al.* The role of mirroring and mentalizing networks in mediating action intentions in autism. *Mol. Autism* **5**, 50 (2014).

11. Hamilton, A. & Marsh, L. Two systems for action comprehension in autism: mirroring and mentalizing. in *Understanding Other Minds* (eds. Baron-Cohen, S., Lombardo, M. & Tager-Flusberg, H.) (Oxford University Press, 2013). doi:10.1093/acprof:oso/9780199692972.001.0001

12. Sperduti, M., Guionnet, S., Fossati, P. & Nadel, J. Mirror Neuron System and Mentalizing System connect during online social interaction. *Cogn. Process.* **15**, 307–316 (2014).

13. Decety, J. & Jackson, P. L. The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* **3**, 71–100 (2004).

14. Fan, Y. T., Chen, C., Chen, S. C., Decety, J. & Cheng, Y. Empathic arousal and social understanding in individuals with autism: Evidence from fMRI and ERP measurements. *Soc. Cogn. Affect. Neurosci.* **9**, 1203–1213 (2014).

15. Spunt, R. P. & Lieberman, M. D. An integrative model of the neural systems

supporting the comprehension of observed emotional behavior. *Neuroimage* **59**, 3050–3059 (2012).

16. Spunt, R. P. & Lieberman, M. D. Dissociating Modality-Specific and Supramodal Neural Systems for Action Understanding. *J. Neurosci.* **32**, 3575–3583 (2012).

17. Van Overwalle, F. & Baetens, K. Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage* **48**, 564–584 (2009).

18. Fishman, I., Keown, C. L., Lincoln, A. J., Pineda, J. A. & Müller, R.-A. Atypical Cross Talk Between Mentalizing and Mirror Neuron Networks in Autism Spectrum Disorder. *JAMA Psychiatry* **71**, 751 (2014).

19. Marsh, L. E., Mullett, T. L., Ropar, D. & Hamilton, A. F. de C. Responses to irrational actions in action observation and mentalising networks of the human brain. *Neuroimage* **103**, 81–90 (2014).

20. Sheline, Y. I. *et al.* The default mode network and self-referential processes in depression. *Proc. Natl. Acad. Sci.* **106**, 1942–1947 (2009).

21. Gallese, V. Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* **2**, 493–501 (1998).

22. Przyrembel, M., Smallwood, J., Pauen, M. & Singer, T. Illuminating the dark matter of social neuroscience: Considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. *Front. Hum.*

Neurosci. **6**, (2012).

23. Paulus, F. M., Müller-Pinzler, L., Westermann, S. & Krach, S. On the distinction of empathic and vicarious emotions. *Front. Hum. Neurosci.* **7**, 196 (2013).
24. Catmur, C. Understanding intentions from actions: direct perception, inference, and the roles of mirror and mentalizing systems. *Conscious. Cogn.* **44**, (2015).
25. Ansuini, C., Santello, M., Massaccesi, S. & Castiello, U. Effects of end-goal on hand shaping. *J. Neurophysiol.* **95**, 2456–2465 (2006).
26. Ansuini, C., Giosa, L., Turella, L., Altoè, G. & Castiello, U. An object for an action, the same object for other actions: effects on hand shaping. *Exp. Brain Res.* **185**, 111–119 (2008).
27. Marteniuk, R. G., MacKenzie, C. L., Jeannerod, M., Athenes, S. & Dugas, C. Constraints on human arm movement trajectories. *Can. J. Psychol. Can. Psychol.* **41**, 365 (1987).
28. Naish, K. R., Reader, A. T., Houston-Price, C., Bremner, A. J. & Holmes, N. P. To eat or not to eat? Kinematics and muscle activity of reach-to-grasp movements are influenced by the action goal, but observers do not detect these differences. *Exp. Brain Res.* **225**, 261–275 (2013).
29. Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E. & Lieberman, M. D. Evidence for social working memory from a parametric functional MRI study. *Proc. Natl. Acad. Sci.* **109**, 1883–1888 (2012).

30. Spunt, R. P. & Lieberman, M. D. The Busy Social Brain. *Psychol. Sci.* **24**, 80–86 (2013).

31. Wagner, D. D., Kelley, W. M. & Heatherton, T. F. Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cereb. cortex* **21**, 2788–2796 (2011).

32. Rameson, L. T., Morelli, S. A. & Lieberman, M. D. The neural correlates of empathy: experience, automaticity, and prosocial behavior. *J. Cogn. Neurosci.* **24**, 235–245 (2012).

33. Senju, A., Southgate, V., White, S. & Frith, U. Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science (80-.).* **325**, 883–885 (2009).

34. Mochizuki, H., Franca, M., Huang, Y.-Z. & Rothwell, J. C. The role of dorsal premotor area in reaction task: comparing the "virtual lesion" effect of paired pulse or theta burst transcranial magnetic stimulation. *Exp. brain Res.* **167**, 414–421 (2005).

35. Ortigue, S., Thompson, J. C., Parasuraman, R. & Grafton, S. T. Spatio-temporal dynamics of human intention understanding in temporo-parietal cortex: a combined EEG/fMRI repetition suppression paradigm. *PLoS One* **4**, e6962 (2009).

36. Manera, V., Becchio, C., Cavallo, A., Sartori, L. & Castiello, U. Cooperation or

competition? Discriminating between social intentions by observing prehensile movements. *Exp. Brain Res.* **211**, 547–556 (2011).

37. Arioli, M. *et al.* Affective and cooperative social interactions modulate effective connectivity within and between the mirror and mentalizing systems. *Hum. Brain Mapp.* **39**, 1412–1427 (2018).

38. Cole, E. J., Barraclough, N. E. & Enticott, P. G. Investigating Mirror System (MS) Activity in Adults with ASD When Inferring Others ' Intentions Using Both TMS and EEG. *J. Autism Dev. Disord.* **48**, 2350–2367 (2018).

39. Zhang, L., Gan, J. Q., Zheng, W. & Wang, H. Spatiotemporal Phase Synchronization in Adaptive Reconfiguration from Action Observation Network to Mentalizing Network for Understanding Other's Action Intention. *Brain Topogr.* **31**, 447–467 (2018).

40. Baron-Cohen, S., Leslie, A. M. & Frith, U. Does the autistic child have a 'theory of mind'? *Cognition* **21**, 37–46 (1985).

41. Neuroscience, H. *et al.* Reconceptualizing functional brain connectivity in autism from a developmental perspective. **7**, 1–11 (2013).

42. Cole, E. J., Barraclough, N. E. & Andrews, T. J. Reduced connectivity between mentalizing and mirror systems in autism spectrum condition. *Neuropsychologia* **122**, 88–97 (2019).

43. Libero, L. E. *et al.* The role of mirroring and mentalizing networks in mediating

action intentions in autism. *Mol. Autism* **5**, 1–13 (2014).

44. Woodward, N. D., Giraldo-Chica, M., Rogers, B. & Cascio, C. J. Thalamocortical Dysconnectivity in Autism Spectrum Disorder: An Analysis of the Autism Brain Imaging Data Exchange. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2**, 76–84 (2017).

45. Roach, B. J. & Mathalon, D. H. Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophr. Bull.* **34**, 907–926 (2008).

46. Cohen, M. X. *Analyzing Neural Time Series Data: Theory and Practice*. (The MIT Press, 2014).

47. Brodbeck, V. *et al.* Electroencephalographic source imaging: a prospective study of 152 operated epileptic patients. *Brain* **134**, 2887–2897 (2011).

48. Chu, C. J. High density EEG—What do we have to lose? *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.* **126**, 433 (2015).

49. Lystad, R. P. & Pollard, H. Functional neuroimaging: a brief overview and feasibility for use in chiropractic research. *J. Can. Chiropr. Assoc.* **53**, 59 (2009).

50. Soto-Icaza, P., Vargas, L., Aboitiz, F. & Billeke, P. Beta oscillations precede joint attention and correlate with mentalization in typical development and autism. *Cortex* **113**, 210–228 (2019).

51. Cooper, N. R., Simpson, A., Till, A., Simmons, K. & Puzzo, I. Beta event-related

desynchronization as an index of individual differences in processing human facial expression: further investigations of autistic traits in typically developing adults. *Front. Hum. Neurosci.* **7**, 1–8 (2013).

52. Park, J., Kim, H., Sohn, J. W., Choi, J. R. & Kim, S. P. EEG beta oscillations in the temporoparietal area related to the accuracy in estimating others' preference. *Front. Hum. Neurosci.* **12**, 1–11 (2018).

53. Luck, S. J. *An introduction to the event-related potential technique*. (MIT press, 2014).

54. Nolan, H., Whelan, R. & Reilly, R. B. FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *J. Neurosci. Methods* **192**, 152–162 (2010).

55. Soufflet, L. *et al.* A statistical evaluation of the main interpolation methods applied to 3-dimensional EEG mapping. *Electroencephalogr. Clin. Neurophysiol.* **79**, 393–402 (1991).

56. Yao, D. A method to standardize a reference of scalp EEG recordings to a point at infinity. *Physiol. Meas.* **22**, 693–711 (2001).

57. Geselowitz, D. B. The zero of potential. *IEEE Eng. Med. Biol. Mag.* **17**, 128–136 (1998).

58. O. Bertrand, F. P. and J. P. A theoretical justification of the average reference in topographic evoked potential studies. **62**, 462–464 (1985).

59. Artoni, F., Delorme, A. & Makeig, S. Applying dimension reduction to EEG data by Principal Component Analysis reduces the quality of its subsequent Independent Component decomposition. *Neuroimage* **175**, 176–187 (2018).

60. Radüntz, T., Scouten, J., Hochmuth, O. & Meffert, B. EEG artifact elimination by extraction of ICA-component features using image processing algorithms. *J. Neurosci. Methods* **243**, 84–93 (2015).

61. Sur, S. & Sinha, V. Event-related potential: An overview. *Ind. Psychiatry J.* **18**, 70 (2009).

62. Sassenhagen, J. & Draschkow, D. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* **56**, 1–8 (2019).

63. Luck, S. J. & Gaspelin, N. How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology* **54**, 146–157 (2017).

64. Popov, T., Oostenveld, R. & Schoffelen, J. M. FieldTrip Made Easy : An Analysis Protocol for Group Analysis of the Auditory Steady State Brain Response in Time , Frequency , and Space. **12**, 1–11 (2018).

65. Pernet, C. R., Latinus, M., Nichols, T. E. & Rousselet, G. A. Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. *J. Neurosci. Methods* **250**, 85–93 (2015).

66. Hayasaka, S. & Nichols, T. E. Combining voxel intensity and cluster extent with

permutation test framework. **23**, 54–63 (2004).

67. Groppe, D. M., Urbach, T. P. & Kutas, M. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology* **48**, 1711–1725 (2011).

68. Perry, A., Troje, N. F. & Bentin, S. Exploring motor system contributions to the perception of social information: Evidence from EEG activity in the mu/alpha frequency range. *Soc. Neurosci.* **5**, 272–284 (2010).

69. David, N. *et al.* Differential involvement of the posterior temporal cortex in mentalizing but not perspective taking. *Soc. Cogn. Affect. Neurosci.* **3**, 279–289 (2008).

70. Horan, W. P., Pineda, J. A., Wynn, J. K., Iacoboni, M. & Green, M. F. Some markers of mirroring appear intact in schizophrenia: Evidence from mu suppression. *Cogn. Affect. Behav. Neurosci.* **14**, 1049–1060 (2014).

71. Cuellar, M. E. & del Toro, C. M. Time-frequency analysis of Mu rhythm activity during picture and video action naming tasks. *Brain Sci.* **7**, (2017).

72. Griffiths, D. J. *Introduction to electrodynamics.* (AAPT, 2005).

73. Cacioppo, S. & Cacioppo, J. T. Dynamic spatiotemporal brain analyses using high-performance electrical neuroimaging, Part II: A step-by-step tutorial. *J. Neurosci. Methods* **256**, 184–197 (2015).

74. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: A

User-Friendly Application for MEG/EEG Analysis. *Comput. Intell. Neurosci.* **2011**, 1–13 (2011).

75. Koenig, T. *et al.* Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages. *Neuroimage* **16**, 41–48 (2002).

76. von Wegner, F. & Laufs, H. Information-Theoretical Analysis of EEG Microstate Sequences in Python. *Front. Neuroinform.* **12**, 1–10 (2018).

77. Williams, N. J., Daly, I. & Nasuto, S. J. Markov Model-Based Method to Analyse Time-Varying Networks in EEG Task-Related Data. *Front. Comput. Neurosci.* **12**, 1–18 (2018).

78. Burle, B. *et al.* Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *Int. J. Psychophysiol.* **97**, 210–220 (2015).

79. Halder, T., Talwar, S., Jaiswal, A. K. & Banerjee, A. Quantitative Evaluation in Estimating Sources Underlying Brain Oscillations Using Current Source Density Methods and Beamformer Approaches. *eneuro* **6**, ENEURO.0170-19.2019 (2019).

80. Hauk, O., Wakeman, D. G. & Henson, R. Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *Neuroimage* **54**, 1966–1974 (2011).

81. Hauk, O. Keep it simple: A case for using classical minimum norm estimation in the analysis of EEG and MEG data. *Neuroimage* **21**, 1612–1621 (2004).

82. Stropahl, M., Bauer, A.-K. R., Debener, S. & Bleichner, M. G. Source-modelling

auditory processes of EEG data using EEGLAB and Brainstorm. *Front. Neurosci.* **12**, 309 (2018).

83. Belardinelli, P., Ortiz, E. & Braun, C. Source Activity Correlation Effects on LCMV Beamformers in a Realistic Measurement Environment. *Comput. Math. Methods Med.* **2012**, 1–8 (2012).

84. Manfredonia, J. *et al.* Automatic recognition of posed facial expression of emotion in individuals with autism spectrum disorder. *J. Autism Dev. Disord.* **49**, 279–293 (2019).

85. Blau, V. C., Maurer, U., Tottenham, N. & McCandliss, B. D. The face-specific N170 component is modulated by emotional facial expression. *Behav. Brain Funct.* **3**, 1–13 (2007).

86. McGugin, R. W., Newton, A. T., Tamber-Rosenau, B., Tomarken, A. & Gauthier, I. Thickness of deep layers in the fusiform face area predicts face recognition. *bioRxiv* 788216 (2019).

87. Choi, D. *et al.* Gender difference in N170 elicited under oddball task. *J. Physiol. Anthropol.* **34**, 7 (2015).

88. Bassett, D. S., Meyer-Lindenberg, A., Achard, S., Duke, T. & Bullmore, E. Adaptive reconfiguration of fractal small-world human brain functional networks. *Proc. Natl. Acad. Sci.* **103**, 19518–19523 (2006).

89. Kitzbichler, M. G., Henson, R. N. A., Smith, M. L., Nathan, P. J. & Bullmore, E. T.

Cognitive Effort Drives Workspace Configuration of Human Brain Functional Networks. *J. Neurosci.* **31**, 8259–8270 (2011).

90. Zhang, L., Gan, J. Q., Zheng, W. & Wang, H. Spatiotemporal Phase Synchronization in Adaptive Reconfiguration from Action Observation Network to Mentalizing Network for Understanding Other's Action Intention. *Brain Topogr.* **31**, 447–467 (2018).

91. Orozco, S. & Ehlers, C. L. Gender differences in electrophysiological responses to facial stimuli. *Biol. Psychiatry* **44**, 281–289 (1998).

92. Ito, T. A. & Urland, G. R. The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cogn. Affect. Behav. Neurosci.* **5**, 21–36 (2005).

93. Rousselet, G. A., Macé, M. J.-M. & Fabre-Thorpe, M. Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *J. Vis.* **4**, 2 (2004).

94. Polich, J. Updating P300: An integrative theory of P3a and P3b. *Clin. Neurophysiol.* **118**, 2128–2148 (2007).

95. Kourtis, D., Sebanz, N. & Knoblich, G. Predictive representation of other people's actions in joint action planning: An EEG study. *Soc. Neurosci.* **8**, 31–42 (2013).

96. Qu, C., Wang, Y. & Huang, Y. Social exclusion modulates fairness consideration in the ultimatum game: an ERP study. *Front. Hum. Neurosci.* **7**, 1–8 (2013).

97. Maguire, M. J., Magnon, G., Ogiela, D. A., Egbert, R. & Sides, L. The N300 ERP component reveals developmental changes in object and action identification. *Dev. Cogn. Neurosci.* **5**, 1–9 (2013).

98. Franklin, M. S., Dien, J., Neely, J. H., Huber, E. & Waterson, L. D. Semantic priming modulates the N400, N300, and N400RP. *Clin. Neurophysiol.* **118**, 1053–1068 (2007).

99. Wu, Y. C. & Coulson, S. How iconic gestures enhance communication: An ERP study. *Brain Lang.* **101**, 234–245 (2007).

100. Baetens, K., Der Cruyssen, L. Van, Achtziger, A., Vandekerckhove, M. & Van Overwalle, F. N400 and LPP in spontaneous trait inferences. *Brain Res.* **1418**, 83–92 (2011).

101. Schupp, H. T., Schmälzle, R., Flaisch, T., Weike, A. I. & Hamm, A. O. Affective picture processing as a function of preceding picture valence: An ERP analysis. *Biol. Psychol.* **91**, 81–87 (2012).

102. Giglio, A. C. A., Minati, L. & Boggio, P. S. Throwing the banana away and keeping the peel: Neuroelectric responses to unexpected but physically feasible action endings. *Brain Res.* **1532**, 56–62 (2013).

103. Koelstra, S., Muhl, C. & Patras, I. EEG analysis for implicit tagging of video data. in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* 1–6 (IEEE, 2009). doi:10.1109/ACII.2009.5349482

104. Proverbio, A. M. & Riva, F. RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neurosci. Lett.* **459**, 142–146 (2009).

105. Van Duynslaeger, M., Van Overwalle, F. & Verstraeten, E. Electrophysiological time course and brain areas of spontaneous and intentional trait inferences. *Soc. Cogn. Affect. Neurosci.* **2**, 174–188 (2007).

106. Van Overwalle, F. & Vandekerckhove, M. Implicit and explicit social mentalizing: dual processes driven by a shared neural network. *Front. Hum. Neurosci.* **7**, 1–6 (2013).

107. Carrión, R. E., Keenan, J. P. & Sebanz, N. A truth that's told with bad intent: An ERP study of deception. *Cognition* **114**, 105–110 (2010).

108. Benning, S. D. *et al.* Late Positive Potential ERP Responses to Social and Nonsocial Stimuli in Youth with Autism Spectrum Disorder. *J. Autism Dev. Disord.* **46**, 3068–3077 (2016).

109. Mehmood, R. M. & Lee, H. J. A novel feature extraction method based on late positive potential for emotion recognition in human brain signal patterns. *Comput. Electr. Eng.* **53**, 444–457 (2016).

110. Beudt, S. & Jacobsen, T. On the role of mentalizing processes in aesthetic appreciation: An ERP study. *Front. Hum. Neurosci.* **9**, 1–13 (2015).

111. Haxby, J. V & Gobbini, M. I. *Distributed neural systems for face perception.* (The

Oxford Handbook of Face Perception, 2011).

112. Haxby, J. V *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (80-.).* **293**, 2425–2430 (2001).

113. van der Gaag, C., Minderaa, R. B. & Keysers, C. Facial expressions: What the mirror neuron system can and cannot tell us. *Soc. Neurosci.* **2**, 179–222 (2007).

114. Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A. & Van Overwalle, F. ERP time course and brain areas of spontaneous and intentional goal inferences. *Soc. Neurosci.* **4**, 165–184 (2009).

115. Frederick, J. *et al.* ERP and EEG Oscillations Study of Facial Expression Processing Deficits in Autism. *Appl Psychophysiol Biofeedback* **40**, 129–130 (2015).

116. Rossion, B. Understanding individual face discrimination by means of fast periodic visual stimulation. *Exp. Brain Res.* **232**, 1599–1621 (2014).

117. Wu, X. *et al.* Gender-specific effects of vasopressin on human social communication: An ERP study. *Horm. Behav.* **113**, 85–94 (2019).

118. Möhring, N., Brandt, E. S. L., Mohr, B., Pulvermüller, F. & Neuhaus, A. H. ERP adaptation provides direct evidence for early mirror neuron activation in the inferior parietal lobule. *Int. J. Psychophysiol.* **94**, 76–83 (2014).

119. Proverbio, A. M. Sex differences in social cognition: The case of face processing. *J. Neurosci. Res.* **95**, 222–234 (2017).

120. Sun, Y., Gao, X. & Han, S. Sex differences in face gender recognition: an event-related potential study. *Brain Res.* **1327**, 69–76 (2010).

121. Melloni, M. *et al.* Your perspective and my benefit: multiple lesion models of self-other integration strategies during social bargaining. *Brain* **139**, 1–19 (2016).

122. Billeke, P. *et al.* Paradoxical Expectation: Oscillatory Brain Activity Reveals Social Interaction Impairment in Schizophrenia. *Biol. Psychiatry* **78**, 421–431 (2015).

123. Hobson, H. M. & Bishop, D. V. M. Mu suppression - A good measure of the human mirror neuron system? *Cortex.* **82**, 290–310 (2016).

124. Moore, A., Gorodnitsky, I. & Pineda, J. EEG mu component responses to viewing emotional faces. *Behav. Brain Res.* **226**, 309–316 (2012).

125. D'Croz-Baron, D. F., Baker, M., Michel, C. M. & Karp, T. EEG Microstates Analysis in Young Adults With Autism Spectrum Disorder During Resting-State. *Front. Hum. Neurosci.* **13**, 173 (2019).

126. Jia, H. & Yu, D. Aberrant Intrinsic Brain Activity in Patients with Autism Spectrum Disorder: Insights from EEG Microstates. *Brain Topogr.* **32**, 295–303 (2019).

127. Keuken, M. C. *et al.* The role of the left inferior frontal gyrus in social perception: an rTMS study. *Brain Res.* **1383**, 196–205 (2011).

128. Rajmohan, V. & Mohandas, E. Mirror neuron system. *Indian J. Psychiatry* **49**, 66 (2007).

129. Bigler, E. D. *et al.* Superior temporal gyrus, language function, and autism. *Dev.*

Neuropsychol. **31**, 217–238 (2007).

130. Hari, R. & Kujala, M. V. Brain basis of human social interaction: from concepts to brain imaging. *Physiol. Rev.* **89**, 453–79 (2009).

131. Rizzolatti, G., Keysers, C. & Gallese, V. A unifying view of the basis of social cognition. *Trends Cogn. Sci.* **8**, 396–403 (2004).

132. Iacoboni, M. *et al.* Reafferent Copies of Imitated Actions in the Right Superior Temporal Cortex. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13995–13999 (2001).

133. Procházka, A., Mudrová, M., Vyšata, O., Háva, R. & Araujo, C. P. S. Multi-channel EEG signal segmentation and feature extraction. *INES 2010 - 14th Int. Conf. Intell. Eng. Syst. Proc.* 317–320 (2010). doi:10.1109/INES.2010.5483824

134. Khanna, A., Pascual-Leone, A., Michel, C. M. & Farzan, F. Microstates in resting-state EEG: Current status and future directions. *Neurosci. Biobehav. Rev.* **49**, 105–113 (2015).