

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Interpreting Actions: The Role of Sensory Encoding of Predictive Cues in the AON

Permalink

<https://escholarship.org/uc/item/6b75m0zk>

Author

Zhou, Xiaojue

Publication Date

2023

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Interpreting Actions: The Role of Sensory Encoding of Predictive Cues in the AON

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Neuroscience

by

Xiaojue Zhou

Dissertation Committee:
Professor Emily Grossman, Chair
Professor Jeffrey Krichmar
Professor Megan Peters

2023

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES	viii
ACKNOWLEDGMENTS	ix
VITA	x
ABSTRACT OF THE DISSERTATION	xii
1 Where are our action predictions? Action Understanding as a Predictive Process: Integrative Model And Neuroimage Evidence	1
1.1 Introduction	1
1.2 Action hierarchy and action prediction	3
1.2.1 Action Hierarchy	3
1.2.2 Action-to-Goal and Goal-to-Action	5
1.2.3 Inferences in Action Prediction	8
1.3 Action Observation in the brain: The Action Observation Network (AON) .	11
1.3.1 Neural pathway for action recognition	11
1.4 Action Prediction Model: Affordances and Generative Model	14
1.4.1 Predictions based on physical inference of objects and environment: affordance interacting with context	14
1.4.2 Neural evidence of affordances interact with context	16
1.5 Prediction in the Brain	19
1.5.1 Predicted action outcome in inferior frontalparietal cortex	19
1.5.2 Irrational Action: action prediction error	20
1.6 Conclusion and specific aims of the current work	23
2 Configuration of the Action Observation Network Depends on the Goals of the Observer	25
3 Neural Decoding and the Influence of Predictive Cues on Action Observation Network as Revealed by Multivariate Pattern Analysis	47
3.1 Introduction	47
3.2 Methods	50

3.3	Results	54
3.4	Discussion	58
4	Diffusion Model Insights into Evidence Accumulation During Action Prediction: The Role of Predictive Cues Brain Responses within the Action Observation Network	61
4.1	Introduction	61
4.2	Methods	63
4.3	Results	65
4.4	Discussion	68
5	Conclusion	71
	Bibliography	72
	Appendix A Appendix Title	88
	Appendix B Appendix	89

LIST OF FIGURES

	Page
1.1 Hierarchical predictive model. Action is the observable layer. State and Trait are hidden layers. The arrow indicating probabilistic mapping between two layers. (Tamir and Thornton, 2018)	10
1.2 Variable and Stable Affordance neural pathway (?)	16
2.1 Schematic of trial sequence. Participants were cued as to which aspect of the vignette to attend (attend to action, goal or identity). The three second vignette depicted an open room as an avatar approaches a shelf, directs their gaze to one of the two boxes, then either jumps or crouches to retrieve an object. Participants were prompted to discriminate the action (jump or crouch), the goal (the box positioned high or low) or the actor (man or boy) in accordance with the attention cue.	30
2.2 Defining ROI and Connectivity Method. a) The pSTS is identified using an independent localizer and a mixed effects model that treated individual subjects as random effects. The GLM analysis was conducted on vertices in a cortex-based aligned surface space. b) Functional connectivity to the seed identified large regions of interest, each of which was further divided into smaller parcels as defined by Glasser et al. (2016). Functional systems labels for each parcel were defined using the modal network system as defined by Schaefer et al. (2018). The 17 networks in that classification were combined into eight for simpler interpretation (i.e. the dorsal attention A and B were combined into a single dorsal attention). c) Functional connectivity between two parcels was calculated using the beta-series approach in which the connectivity was computed as the correlation (circles) between the timeseries of trialwise beta estimates (rectangles). Regions of interest used this approach with trials unlabeled by task condition. The main connectivity analysis further split the trials into unique types based on the task labels.	32
2.3 ROI Identification. Left: Group-level maps of functional connectivity ($z \geq 1.96$, $p < 0.05$ uncorrected) of the whole brain using the right pSTS (purple) as a seed. All trials regardless of attention manipulation were included in this analysis, which served to localize large regions of interest that were further subdivided into parcels. Right: Connectivity maps divided into regions of interest: inferior frontal cortex (IFC), posterior parietal cortex (PPC) and lateral and anterior superior temporal sulcus (STS+).	35

2.4	Modulation of Connectivity from pSTS between pairs of tasks (attend to action, goals or identity) in the pSTS (the seed ROI) and the other ROIs. Each bar indicates the magnitude of the difference in functional connectivity strength in a single parcel, computed across participants using the pair of conditions as indicated on the left. Right and left sides of each polar plot indicate connectivity to the right and left hemisphere pSTS, respectively. Top and bottom quadrants in each polar plot specify left and right hemispheres of the target ROIs. Top row: Attending to action (red) vs identity (gray). Middle row: attending to actions (red) versus goals (blue). Bottom row: attending to goals (blue) versus identity (gray) of the actors. Dashed line indicates the statistical significance threshold ($p < 0.05$).	36
2.5	Functional Network of the Modulated Connectivity. Panel a) IFC and STS+ parcels with significant modulation in the strength of functional connectivity when attending to actions versus goals. Panel b) The magnitude of functional connectivity modulation (as shown in Figure 2.4), colored by functional network assignment as derived from Schaefer et al. (2018). Bars with solid border have magnitude indicating stronger connectivity during actions versus goals. Bars with dashed borders indicates parcels with strong connectivity during goals as compared to actions.	38
2.6	Chordgram showing the functional system engaged between IFC, STS+, and PPC ROIs when subjects attend to actions more than goals (left) and attend to goals more than actions (right). Each strand represents the proportion of connections between two functional networks that are significantly modulated ($p < 0.1$) by the attentional instruction. In each of the chordgram, colors of the strand represent the functional network as assigned from the IFC (shown on the bottom ring). Target strands are ordered as STS+ (central) to PPC (peripheral).	39
2.7	Univariate Analysis Results. Univariate response modulation within each region of interest, all parcels combined. The BOLD modulation is computed as the difference in beta amplitude on trials with attention directed to goals versus action, and vice versa. Amplitude above zero indicates a stronger BOLD response for trials with attention directed to the goals versus actions in the vignettes. Scores below zero scores indicate stronger BOLD when attending to actions (versus goals). Each bar reflects the average ROI univariate response across all subjects and parcels within the ROI. Lighter colored bars show the univariate response in the left hemisphere and the darker colored bars show the BOLD response in the right hemisphere. Asterisks (*) indicated significant modulation of the BOLD response by the attention condition (in the bilateral PPC and right IFC ROIs).	40
2.8	Univariate response results in PPC, IFC and STS+ ROI. each bar represent one parcel and the color indicates whether the response is stronger for action(red) or goal(blue).	41

3.1	Stimuli and Design For each trial, participants were first given a cue about the upcoming actions during cued trials, and empty circle for uncued trials. Anticipatory interval was the time window following the cue and lasted from 3 to 6 seconds pseudo-randomized for each trial. After the anticipatory interval, participants then watched a 2s action vignette with 83% of the trials depicted action correspond to the cue given in the current trial. Participants responded with button to the action depicted in the stimuli both during the stimuli and after the stimuli if they haven't responded. At the end of each trial, the intertrial interval was randomly assigned between 3 to 6s.	51
3.2	MVPA analysis in AON ROI A) Multivariate pattern analysis (MVPA) was trained and test to classify the actions depicted in the stimuli for activation patterns during anticipatory interval and perceptual encoding interval separately. B) Regions of Interest in the current study includes bilateral pSTS, bilateral aSTS, bilateral IFG, and bilateral IPL. Each region was further divided into smaller parcels defined by Glasser et al. (2016)	52
3.3	Hypothesis Hypothesis of the current study. Left: During the anticipatory interval, we expect the sensory driven right pSTS to not be able to classify the actions but right IFG would have the action represented. Right: During the movie interval, we expect that the right pSTS would classify the action and even better when given a valid cue. We also expect improved classification for validly cued trials in the right IFG, but less than the pSTS due to the specificity of action representations in the pSTS.	54
3.4	MVPA Results. Left: Classification accuracies during the anticipatory interval for cued trials, including valid and invalid cues, and for uncued trials. Classifiers trained on each ROI using the parcels within the ROI as features. Right: Classification accuracy during the perceptual encoding interval. Here the left pSTS showed significant difference between cued and uncued trials but note the uncued condition did not classify significantly above chance.	55
3.5	Univariate Analysis Results. MVPA results for each parcel. Figure shows the parcels in which the the MVPA classifiers trained on the activation patterns both classified the actions above chance across both cueing conditions and also the classification accuracies that were significantly higher during the cued trials than uncued trials by testing the within subject differences using a paired t test ($p < 0.05$, uncorrected). These parcels are p9-46v during the perceptual encoding interval. AVI, FEF in the right IFG, and STV in the lateral left STS during anticipatory interval.	56

3.6	MVPA Results. Left: Average classification results when the classifiers were trained and tested in each parcel using the vertices within the parcel as the features in the classifiers. The cross-validated classification accuracies were then averaged across the subjects and parcels in each ROI. The asterisks indicate ROIs with significant differences between the classification accuracy for the validly cued and uncued conditions, evaluated using a paired t-test. The tilde indicates ROIs in which classification accuracy (across all conditions) is significantly above chance (0.5). Right: Average classification accuracy across all the parcels with the classifiers trained and tested on each parcel independently and then averaged across all the subjects and parcels.	57
3.7	Univariate Analysis Results. Left: BOLD response during anticipatory interval for different cueing conditions. Right: BOLD response during perceptual encoding interval.	58
4.1	Hierarchical Diffusion Modeling Results. Left: The posterior distribution of bias in starting point according to each cueing condition. Right: Posterior distributions of bias in starting point according to each cueing condition and also the length of the anticipatory interval.	66
4.2	Hierarchical Diffusion Modeling with Neural Measurements. a) and b) show the distribution plot of difference (Cued - Uncued) in BOLD activation during the anticipatory interval in right IFG and right pSTS against the difference between posterior estimates of bias in starting point fitted from the HDDM models for each subjects. c) and d) show the distribution plot of difference (Cued - Uncued) in MVPA classification accuracies when classifying the action cues during the anticipatory interval in right IFG and right pSTS against the difference between posterior estimates of bias in starting point fitted from the HDDM models for each subjects.	67

LIST OF TABLES

	Page
2.1 Univariate results table showing the repeated measures t test tat compares the average BOLD response between the two attention tasks. A: action, G: goal, I: identity.	40

ACKNOWLEDGMENTS

Studies conducted in chapter 2 to 4 were funded by the National Science Foundation BCS-1658560 to EG and BCS-1658078 to JP.

I extend my heartfelt thanks to Daniel Stehr, John Pyles, and Emily Grossman, who played a pivotal role as a co-author in Chapter 2 to Chapter 4. Their invaluable insights, expertise, and guidance significantly enriched the content and depth of these chapters.

I extend my profound gratitude to Daniel Stehr, whose collaboration was pivotal in the data collection process and who skillfully led the design of the experiment for Chapter 2. His insights and dedication were invaluable in shaping the foundational aspects of our research.

Special thanks are due to John Pyles for providing the funding and also his indispensable technical support and methodological expertise in the collection of fMRI data. His proficiency in handling complex technical aspects greatly enhanced the quality and reliability of our findings.

I am deeply indebted to Emily Grossman for providing the funding and her expert advice and supervision throughout the study. Her guidance and oversight were instrumental across all chapters, providing a cohesive and robust structure to the entire thesis.

My personal contribution primarily focused on leading the study for Chapters 3 and 4, an endeavor that was made possible and enriched by the collective efforts and support of my esteemed colleagues.

VITA

Xiaojue Zhou

EDUCATION

Doctor of Philosophy in Cognitive Neuroscience **2023**
University of California Irvine *Irvine, CA*

Bachelor of Science in Statistics and Psychology **2016**
University of Wisconsin - Madison *Madison, WI*

PROFESSIONAL EXPERIENCE

Graduate Research Assistant **2018–2023**
University of California, Irvine *Irvine, California*

Technical Consultant **2023**
Deep Valley Labs *San Jose, California*

Research Fellow **2022**
Devcom ARL

Research Specialist **2016–2018**
UW Health and Hospital *Madison, Wisconsin*

TEACHING EXPERIENCE

Teaching Assistant **2012–2018**
University of California, Irvine *Irvine, CA*

REFEREED JOURNAL PUBLICATIONS

Heterogeneous behavioral, endocrine and neural outcomes associated with early life rearing adversity in primates **Under review**
Psychoneuroendocrinology

Configuration of the Action Observation Network Depends on the Goals of the Observer **2023**
Neuropsychologia

Heterogeneous behavioral, endocrine and neural outcomes associated with early life rearing adversity in primates **Under review**
Psychoneuroendocrinology

Category-specific activations depend on imaging mode, task demand, and stimuli modality: An ALE meta-analysis	2021
Neuropsychologia	
Top-down attention guidance shapes action encoding in the pSTS	2021
Cortex	
Consequences of Altering Prefrontal-Temporal Lobe Connectivity in Young Nonhuman Primates.	2017
Biological Psychiatry	

REFEREED CONFERENCE PUBLICATIONS

Decode the Unseen: Classifying Action Representations During the Anticipatory Interval in the Action Observation Network (AON)	March 2023
Cognitive Neuroscience Society	
Integration of the AON and executive networks during action vs goal inference	March 2022
Cognitive Neuroscience Society	
Functional connectivity of the pSTS in the action observation network (AON) modulated by attention to action features during action and goal inference, an fMRI experiment	November 2021
Society for Neuroscience	
. Functional Connectivity During Action Recognition Modulated By Top-down Goals	November 2019
Society for Neuroscience	

ABSTRACT OF THE DISSERTATION

Interpreting Actions: The Role of Sensory Encoding of Predictive Cues in the AON

By

Xiaojue Zhou

Doctor of Philosophy in Cognitive Neuroscience

University of California, Irvine, 2023

Professor Emily Grossman, Chair

We examined the role of top-down influences, including predictive coding, in the action observation network (AON) and extended brain regions. To achieve this, we examined functional connectivity within the network, multivariate pattern classification of regional representations, and diffusion modeling to reveal indicators of predictive decision bias. In the first study, subjects were exposed to three-second action vignettes involving avatars performing one of two different actions. The participant was cued in advance to attend to the action, the outcome of the action (goal) or the identity of the actor. The results indicated that attention could modulate precision with which the actions were encoded in the AON, specifically the right pSTS and IFG. These findings are consistent with the hypothesis that the IFG provides top-down feedback to the pSTS and implicates a right-lateralized system in processing action kinematics. Study 2 seeks to evaluate whether information about anticipated actions is apparent in the neural signal prior to observation, as a potential source action prediction. In this second study, in half of the trials subjects were cued as to which action they would see before observing the action vignettes. The findings revealed that, when given a valid cue, the right IFG had a higher classification accuracy during cued trials than uncued trials (uncorrected for multiple comparisons), when measured during the anticipatory interval (prior to encoding). These results suggest that predictive cues can indeed impact the neural decoding processes within the AON. Study 3 further examines this predictive process in terms of

the association between response time and brain responses. We used hierarchical diffusion modeling to model trial-wise posterior estimates of decision bias, which showed clear evidence for observer's using the cue during the anticipatory interval. These studies collectively demonstrate how top-down influences from the IFG enhance the pSTS's reconstruction of action dynamics. These results are also consistent with the proposal that predictive cues can bias both behavioral responses and neural activity, providing a deeper understanding of the neural pathways involved in interpreting socially relevant actions.

Chapter 1

Where are our action predictions?

Action Understanding as a Predictive Process: Integrative Model And Neuroimage Evidence

1.1 Introduction

Action understanding guides inferences about our environment and their intentions without conscious awareness. For example, we perceive hand shape, direction of movement, and object's functional properties in object-oriented hand actions. From these cues observers can extract features that are diagnostic to the performer's different action goals. Often this is described as "action understanding", but in fact it is a very complex task that depends largely on the goals of the observer.

Initially, action understanding was considered a bottom-up process (Rizzolatti and Matelli,

2003). From the simulation theory point of view, the anticipated outcomes of observed actions are derived in a bottom-up fashion based on the observer’s own prior knowledge and motor experience (Tamir and Thornton, 2018). The main debate about which neural system supports action understanding revolves around whether mirror neurons, which initially were proposed to encode action intention, are also necessary to understand action kinematics. In this view, the observer needs to possess the knowledge of the action-stimulus association in order to predict action outcomes before understanding the complete action sequences (Csibra and Gergely, 2007). However, we are capable of predicting outcomes even for new actions and goals not previously seen. Also, because of the lack of one-to-one mapping between actions and their outcomes, which means multiple action sequences can achieve the same goal and vice versa, understanding actions needs more than simple stimuli-action association. Therefore simulation theory is limited in its ability to explain understanding action goals.

During natural action observation, we are doing causal inference at multiple levels to interpret actions, predict their outcomes and infer the actor’s goals. For example, participants can use subtle differences in the hand movement patterns to discriminate between two different immediate goals when the performers in a movie had just finished lifting the objects but before they actually reach the immediate end goal – pouring or drinking from the cup (Koul et al., 2018). This is an example of action goals derived from bottom-up kinematic cues: identifying goal from perceived action features. On the other hand, we can also use top-down signals to distinguish actions based on different goals. For example, subjects correctly identified appropriate actions based on inferences of the physical properties in the environment such as pushing on the edge or spinning door to easily rotate the heavy door (Fischer et al., 2016). These two processes usually happen in parallel and indicate the need for an integrative model about multiple directions of inferences during action observation.

Understanding someone else’s action and its consequences are the foundation of social interaction. Social cognition can be thought of as a causal inference process, such as inferring

agent’s transitory states (goals and intentions) from their eye, head, hand and body movements as well as the environmental context (Overwalle, 2009). On the other hand, the reverse inference is often ignored: predicting other’s behavior from our social knowledge. Understanding someone else’s intentional goals can help with predicting the according actions. These two directions of causal inference: predict action trajectory from goal and predict action outcome from action kinematics, reflect our knowledge about action during social interaction, further emphasizing the role of action in daily social inference.

In my dissertation, I am going to discuss why action understanding can be thought of action prediction in a hierarchical context. Then I am going to introduce two main sources of predictions: affordances and social perception. Lastly, I will describe how we predict action trajectories and action outcomes based on our internal models of the affordances and social perception. This dissertation will include a discussion of the underlying neural mechanisms and supporting neuroimaging evidence. Most importantly, I am going to integrate this evidence into a coherent theory about action understanding in theory of mind network and mirror neuron system.

1.2 Action hierarchy and action prediction

1.2.1 Action Hierarchy

Before talking about the role of prediction in action observation, it is helpful to clarify the definition of action. When we talked about actions, one could refer to either the kinematics of how the actions were performed, the objects of the action acted upon, or a series of actions orchestrated to fulfill the agent’s intention, often aligning with a broader social goals. Thompson et al. (2019) defined action, goal, and intention identification as three levels in a action hierarchy. First, an action is identified as a specific set of kinematic features such

that the "selection of motor programs that will produce the same action in observer's eyes". For example, two actions such as kicking and jumping are differentiated by how the body is shaped and the unique motion features. It is important to note that actions can be understood in terms of their immediate end goal without further referencing the actor's intentions (Rizzolatti et al., 2009). In contrast, goal identification can be generalized across motor features of the observed actions. For example, the goal to grasp a cup can be achieved with different actions: a power grip or a precision grip (McDonough et al., 2020), whereas the actor's intent in grasping is to drink water. Finally, intention identification answers the observer's question about why the actors perform the action. Therefore, inferences about the actor's intentions or internal mental states can facilitate the observer to infer the . Lastly, action understanding is not strictly hierarchical, such that we can understand someone else's intention before we defined their action or goal.

With the action hierarchy defined above, action predictions can be further broken down at different levels. In motor control, the action hierarchy can be described as an overarching goal of an action broken down into several subgoals and achieved with sequences of steps until the most precise unit, the individual muscle activations (Kilner, 2011; Csibra, 2008). In this model, higher and lower level motor modules are connected to each other (Wolpert and Ghahramani, 2000) such that the lower-level inputs, for example motor codes, can propagate upward and be used to estimate higher level goals (Wolpert et al., 2003). In contrast, the constructed higher-level goals propagate downwards to generate motor plans, and therefore the action mirroring in various motor-related brain regions is predictive for the upcoming action (Csibra, 2008). Generating actions to achieve a desired outcome is called emulation in social learning literature (Huang and Charman, 2005) and the predicted motor codes in turn can also be used for goal inferences (Gallese et al., 1996). Here, action prediction during motor control is revealed through estimating higher level goals as well as predicting upcoming actions.

Does a similar process happen during visual action observation such that instead of motor activation, visual features of individual action kinematics can be used to draw inferences about the higher level goals, and the estimation of goals can further generate into visual expectation of individual action? Under this hierarchical context, understanding an action means an observer can predict action outcome and infer the underlying intention, as well as predict an upcoming action.

1.2.2 Action-to-Goal and Goal-to-Action

Observed actions are generally assumed to be goal-directed, even for the actions we cannot produce ourselves. For example, infants expect that sequentially organized actions are directed towards a specific goal state. In Onishi et al. (2007)'s experiment, 15 months old infants used their background knowledge about cups and bottles to predict the pretended goal and the subsequent drinking action even they were incapable of drinking from or pouring water into a cup. Specifically, infants looked longer at the event when there wasn't an imitated pouring action compared to a cup than where there was, indicating that infants organized action events toward a intended goal state (because most people drink from filled up instead of empty one, which the pouring action indicated).

Action predictions, as described by Csibra (2008), involves two types of internal representations that facilitate two predictive processes: anticipating the outcome and the action itself. Goal attribution not only predicts the intended outcome but also allows for the anticipation of actions leading to that goal. This predictive process uses the hypothesized goal as input, enabling an observer to foresee both the final state and the interim actions (Csibra and Gergely, 2007; Gallese and Goldman, 1998).

Observers make predictions as to the action outcome or intention of an individual when observing actions on-line. Observers may not have explicit knowledge of an actor's goals or

intentions prior to watching the action, but nonetheless can make these inferences from the visual inputs about kinematic cues and the larger context (de C Hamilton and Grafton, 2008; Koul et al., 2018). Csibra and Gergely (2007) described this process as “Action-to-Goal” process: the process to “predict a likely effect of an ongoing action”. In the action to goal process, subjects generalize from observed action’s low-level features to likely outcomes and the effect on the environment. This is similar to the forward propagation mentioned above in motor control model, in which lower-level activation propagates upward to the sensory consequences of the action (Wolpert and Ghahramani, 2000).

If the visual inputs other than actions enable the observer to identify the actor’s end goal (e.g. information from the environment), then the observer can use the expected action outcomes to facilitate visually tracking the action. In other words, the top-down signal from the environment about the goal helps with predicting the upcoming action kinematics. This is also called “Goal-to-Action” process: “predicatively tracking of dynamic actions in real time” (Csibra and Gergely, 2007). Specifically, predicting actions are either inferences of possible actions or constructing an action repository for an upcoming action based on available information. That means we can more accurately and quickly predict action from observing the first few frames of actions and with a limited size of action repository Csibra (2008). In the present work, one action event can be goal, action outcome, intention or context depending on the inferential process. For example, to drink water from a cup can be the action performer’s action outcome of grasping a cup as well as intention when talking about action-to-goal: inferring the action’s achievable state from both observer and performer’s view. At the same time, to drink water from a cup can be the context for the observer during goal-to-action: inferring the performer’s action kinematics constrained by his/her current intentions.

Why do we need to predict actions? First, knowing the goal facilitate action observation in the following ways by: 1) limiting alternative action sequence choices during action ob-

ervation, 2) improving predictive tracking of dynamic actions in real time, and 3) helping the acquisition of knowledge related to new means of achieving the goal. Second, in real life, predicting actions helps the observer to: 1) attribute actions to goals and establish their causal relationship 2) lead to faster interventions and dealing with unexpected events 3) learn the physical parameters of objects such as weight, contents, etc., which is inaccessible to the visual system and has to be learned from either performing actions themselves or watching someone else's actions (McDonough et al., 2020; Csibra, 2008).

The simplest example for these predictions was shown in a developmental study where the inferences were based on the observer's existing action-stimulus association knowledge (Woodward, 1998). Early studies have shown that an infant's shifts its eye gaze to the target object before the completion of the movement after establishing the action-stimulus association (Woodward, 1998), indicating that the observer retrieving the goal, which is the target object, is observing unfinished action sequences. This retrieval process can be think of as a simplest prediction process where the observer can predict the goal and upcoming action from matching the observed action sequences to an existing stimulus-action association. Moreover, infants looked longer when the adults took a new reaching trajectory, indicating they've coded action trajectory with the target objects and was predicting action trajectory while observing adult's reaching for a toy.

Other studies also provided evidence for the action predictions. For example, 8 to 9 months old infants track the adult's eye gaze and turn their attention to the target where the adults were attending (Johnson et al., 1998). Eye gaze effect is larger with shorter response time when the performer's action is congruent with the target object locations in a context scene (Perez-Osorio et al., 2015). Also, understanding intentions influences the observer's eye gaze tracking. For example, beliefs about another person's ability to see in turn have strong top-down effects on gaze processing (Teufel et al., 2010). The observed individual's eye gaze has a strong cueing effect on the observer's attention (Nummenmaa and Calder, 2009) and

therefore can help the observer use the actor’s eye gaze to locate the target objects and help observer narrow the actor’s possible end-goals.

1.2.3 Inferences in Action Prediction

While the action prediction output during visuo-motor interaction is the expected sensory response: motor code (Wolpert and Ghahramani, 2000), prediction during action observation is perceptual expectation, or the observer’s generated visual prediction to match with the incoming perceptual input. Prediction from an internal model during action observation is transforming the internal representation to an external perceptual expectation. In predictive coding theory, we perceive to confirm our predictions, and perception is about updating the predictions based on errors (Tamir and Thornton, 2018; Koster-Hale and Saxe, 2013)

Internal models about the world improve our ability to perceive and organize events. In the case of action perception, internal models include knowledge about the environment and the people around us (Bach and Schenke, 2017). The most obvious internal models include our knowledge about the physics of the world, such as gravity and object permanency. This knowledge is related to how we interact with objects, which is known as affordances, and will be explained in more detail below.

How do we make predictions from our internal model? First, our internal model about the world is modeled hierarchically and as a probabilistic mapping between the observed external behaviors and the hidden internal expectations. The world is experienced as the transitional probabilities between actions and goals, or analogously our observations of the world and inferences on other’s mental states (Tamir and Thornton, 2018; Koster-Hale and Saxe, 2013).

Second, we use these probabilistic mappings from our internal knowledge to make behavioral predictions and test the prediction by observing agent’s action choice. It has been suggested

that inferring goal from action, and vice versa, can be seen as a Bayesian inference problem such that we infer the goal from probability (Friston et al., 2011; Baker et al., 2005), which is proportional to the probability of daily experience of certain outcome associated with a certain action. For example, we infer familiar actions from an outcome with higher probability than rare ones. This is also called associative learning. Monkeys learn to associate goals with never-seen-before tool actions by observing the experimenters manipulate the tool to achieve a certain goal (Fogassi et al., 2005).

Finally, actions are predicted hierarchically by breaking down the goals into sub-goals and eventually predicting the elementary motor acts (Csibra and Gergely, 2007). Whereas some models propose the causal attribution process to be a strictly hierarchical structure, such that predicted actions are hierarchically constrained by personal knowledge and then the current environment (Bach and Tipper, 2007), other models allow direct prediction from personal knowledge to motor programs Tamir and Thornton (2018). On the other hand, both Bach and Tipper (2007) and Tamir and Thornton (2018)'s model are organized in a hierarchical fashion such that predictions can be generated at multiple levels, and people learn the probabilistic mapping between observed and inferred through statistical learning (see figure 1.1). Computationally, this has been modeled using observable layer:action, and hidden layers: state and traits. both observed actions and hidden layers includes expectations of actions being rational and efficient (Tamir and Thornton, 2018).

One source of knowledge about likely actions and their outcomes comes from the physical context. Affordance is the action possibilities from the interaction between object properties and the context, originally defined by Gibson (Gibson, 1979; Sakreida et al., 2016). Affordance reflects our internal models of the physical context and provides us with automatic action expectation according to the objects and environment's physical properties.

Another source of knowledge is social in nature and the predictive coding of actions here is transforming the internal assumptions about the external people and environment into

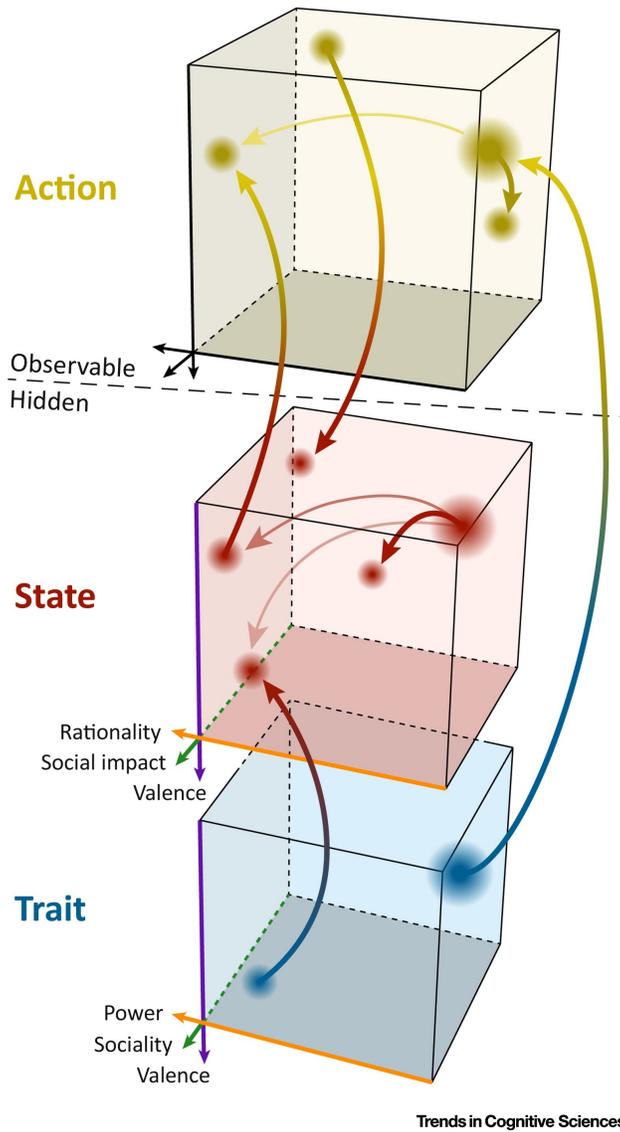


Figure 1.1: Hierarchical predictive model. Action is the observable layer. State and Trait are hidden layers. The arrow indicating probabilistic mapping between two layers. (Tamir and Thornton, 2018)

behavioral predictions (Tamir and Thornton, 2018; Bach and Schenke, 2017). Social perception has been considered as a predictive process (Koster-Hale and Saxe, 2013). We use our knowledge of social dynamics of other people to predict the actions of other people (Tamir and Thornton, 2018). One main theory in social cognition, the theory of mind (ToM), has considered the process of reading someone’s mind as the building of an internal model that makes causal attributions between actions to intentions (Overwalle, 2009). To consider social cognition as an active inference, one main assumption is that the actor actively engages in actions that reflect their goals and intentions (Kelley, 1967). The main function of social cognition during predictive process is to read the agent’s intention to narrow down the agent’s possible target objects or planned behavior.

1.3 Action Observation in the brain: The Action Observation Network (AON)

1.3.1 Neural pathway for action recognition

Studies have identified brain regions recruited during action observation. Action observation and imitation engage a core action observation network (AON), which includes the bilateral inferior frontal gyrus (IFG) or BA45, the posterior middle temporal gyrus (pMTG) and inferior parietal lobule (IPL) (Caspers et al., 2010; Yang et al., 2015). Later, Hafri et al. (2017) found overlapping regions within AON including bilateral IPL, left premotor cortex when subjects generalised to action categories irrespective of the low-level feature input and across different actors, objects, stimuli format, viewpoints, and background.

Biological motion information is mainly processed within STS. Ma et al. (2018) found that different actions can be discriminated based on the activation patterns within three body

sensitive regions: EBA and motion sensitive pSTS and hMT+. Giese et al. (2003) introduce a hierarchical neural model for biological motion recognition including one pathway for body form recognition and one pathway for motion recognition. They are both feed-forward pathways without any top-down signals. In the ventral form pathway (V1 to V4 to IT to STS/FA), biological motion is analyzed by sequences of “snapshots” of body shapes. In the dorsal motion pathway (V1 to MT to STS), biological motion is recognized by analyzing optic-flow patterns, which are abstract moving patterns of direction and speed of a group of moving dots. pSTS is also where form and motion features are bound into action representations of biological motion (Giese et al., 2003).

STS as a social hub

STS, TPJ, and mPFC, the three core regions of ToM network, have been shown in various studies focusing on individual analysis of social information. TPJ were found to be the key regions involved in understanding action goals and inferring an actor’s intentions (Allison et al., 2000; Overwalle and Baetens, 2009). mPFC involved longer social events (Overwalle, 2009), especially during self versus other-referenced tasks. In fact, mPFC was part of the network involved in making causal inference from social intentions, as shown in an computational models (Fleischer et al., 2012).

STS (mostly pSTS) has been shown to respond to different kinds of social features such as bodies, social interaction, speech comparing to non-social feature (Lahnakoski et al., 2012) and this effect was also shown in monkey STS neural activity while watching other monkey’s interactions (Sliwa and Freiwald, 2017). STS activity was also observed when viewing abstract shapes interacting (Tavares et al., 2008), as well as when evaluating relationships between the observed motion and the structure of the surrounding environment (Saxe et al., 2004). Therefore STS is considered a hub of the social brain.

STS picks up social cues such as eye, mouth, hands, and body movements and finds the implied motion by others (Allison et al., 2000). Further, STS integrates human movement perception and represents body parts such as face and limbs with an organized principle (Grosbras et al., 2012; Weiner and Grill-Spector, 2013). STS also infers intention from social cues, such as gaze perception (Nummenmaa and Calder, 2009). Single cells in the superior temporal sulcus (STS) of the macaque brain that are specifically tuned to different gaze directions have been identified (Perrett et al., 1992).

Expressed emotion elicits response in STS. Bilateral STS BOLD activity increased when the adult's expressed emotion that was incongruent with the grasped objects and this effect was larger over development (Wyk et al., 2012, 2009). Also, right pSTS had a causal role in recognizing emotion information from biological motion, which is closely related to the agent's intentions, in a TMS study (Basil et al., 2017). Studies have also shown that STS is involved in the visual analysis of the observer's interpretation of another's actions through biological-motion cues and facial expression to determine the emotional states of others (Pelphrey and Morris, 2006; Pelphrey et al., 2011). Lastly, STS was shown in evaluating interactive stimuli as well as different interactions such as cooperation versus competition (Walbrin et al., 2018; Isik et al., 2017).

pSTS and fusiform gyrus responded to any cues indicative of animacy, which includes goal-directed action, facial features, and biological motion (Shultz and McCarthy, 2014). Interestingly, these regions responded even without human-like feature but only goal-directed machine action in sight. Specifically, goal-directed actions activate pSTS and fusiform gyrus with non-human like machines when the machine either resembles the human or performed goal-directed action with non-human like motion and form (Shultz and McCarthy, 2012), indicating that pSTS generalizes across perceptual features instead of faithfully responds to biological motion.

1.4 Action Prediction Model: Affordances and Generative Model

1.4.1 Predictions based on physical inference of objects and environment: affordance interacting with context

What is affordance?

The likelihood an action is performed to achieve certain goal can also be estimated by teleological reasoning (Csibra and Gergely, 2007) such that the observer expects the actions to be efficient and rational under the current environment (Gergely and Csibra, 2003). Affordance gives us a specific mechanism for this natural expectation, which provides us the ability to form action expectations according to the objects and environment's physical properties. Affordance matching is evident in children's action observation. Children around 12 months old don't faithfully imitate adult's action but rather emulate an action to achieve the same goal only if those motor acts are crucial for achieving the desired goal state (Johnson et al., 1998; Overwalle, 2009). This is achieved when children match the affordances of the goal objects with what they observed (Bach et al., 2014).

Affordance was originally defined as an automatic operation within the visual and motor system in Gibsonian's view (Gibson, 1979). In this view, encoding and retrieving affordance didn't require interpretation of the environment but only retrieving the probabilities of action acted upon the objects.

However, affordance varies with different contexts instead of as a static property. In this view, affordance is classified into three kinds: stable, variable, and canonical, depending on different physical properties of the environment and provides inferences vary in time

length (Borghi and Riggio, 2009; Sakreida et al., 2016). Stable affordance is the information about action probabilities from prior knowledge about objects and not changed in a new environment. An example would be the handle on a coffee cup provides affordance for two action probabilities: grasping and holding. Object shape guides action selection in the early action planning phase (Cuijpers et al., 2004) as well as perception of the mass of the object changes the prior-to-contact grasp kinematics with increased peak grasp aperture and different finger placement on the objects (Eastough and Edwards, 2007).

Variable affordance, on the other hand, is on-line visual processing of goal-directed object interaction and varies with context. Action kinematics are different under different context such as the performer’s intention. One-to-one mapping between sensory stimulus and motor response therefore cannot capture all the kinematic variations. For example, changing the prior intentions with Transcranial Magnetic Stimulation (TMS) stimulation results in different motor selections (Jacquet et al., 2016). Also, motor sequences are remembered better with intention-related cue (e.g. music tone) and also better maintain the mental representation of sensorimotor consequences over long delays (Stöcker and Hoffmann, 2004; Badets and Osiurak, 2015).

Motor intention affects the action kinematics even when manipulating the same objects. In other motor control studies, kinematic structure such as fingertip contact points of reach-to-grasp movements were performed differently according to the performer’s intention of either to pour from a bottle or to move a bottle (Sartori et al., 2011). Other studies also proved that actor’s prior intention during action planning influences the hand shaping (Ansuini et al., 2006) and finger positioning on the object (Ansuini et al., 2008).

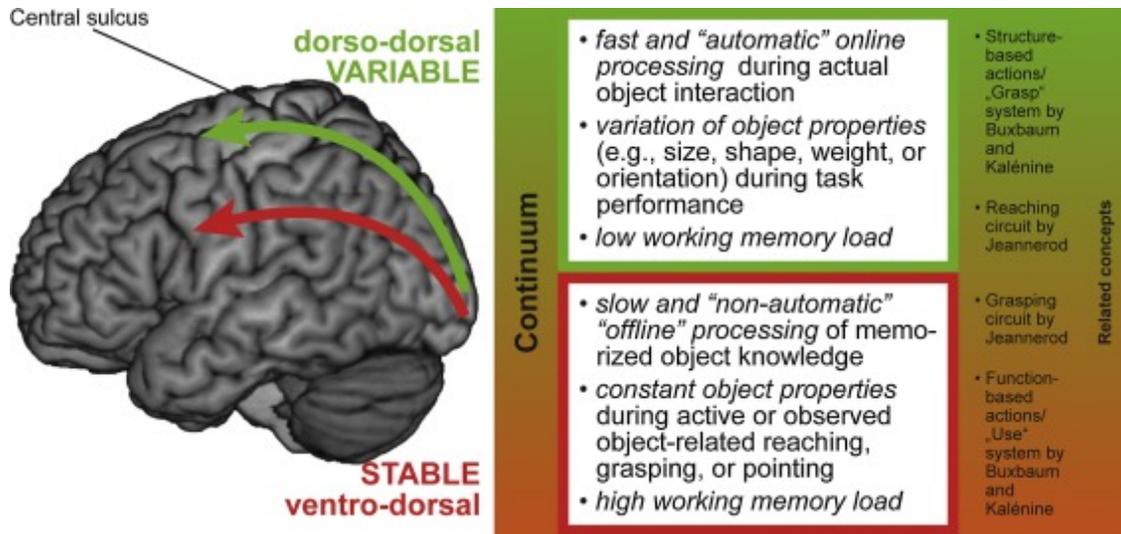


Figure 1.2: Variable and Stable Affordance neural pathway (?)

1.4.2 Neural evidence of affordances interact with context

Retrieve functional knowledge in two streams

How do we retrieve the functional knowledge and infer physical properties from the context? Tools associated functional actions and achievable goals can be retrieved from seeing a picture of a tool and the retrieved action is represented in frontoparietal motor regions, while goals represented in medial temporal gyrus (Chen et al., 2016).

Variable affordance, which is the on-line judgement of an action plan for interacting with an object, involves F2 and F7 via a dorsal-dorsal pathway through the superior parietal lobule (SPL) Sakreida et al. (2016) (See Figure 1.2). This neural pathway involves inferring physical properties and overlaps with brain regions involve action planning and tool use (Sakreida et al., 2016). On the other hand, stable affordances, which have information about action probabilities from prior knowledge about either the object or the environment, are involved in a dorsal-ventral route between visual cortex and F4 and F5 (BA44 in human) (Sakreida et al., 2016).

These two pathways are directly related to the two pathways (Binkofski and Buxbaum, 2013) proposed, which is a bottom-up dorsal dual-route action systems in the human brain for object-directed hand action. In this action system, one pathway recognizes visual properties of motor function during online control of ongoing actions in a dorsal-dorsal pathway (V3 to aIPS to F2/F7), while action understanding happens in ventro-dorsal pathway (MT/MST to IPL to F5/F4). The ventro-dorsal stream's function involves space perception and the recognition of actions made by others.

Two streams interact during object-oriented hand movements. Delayed skilled grasp, a special case of how functional knowledge guides the action, recruits both dorsal and ventral streams because it requires both on-line movement control in the dorsal stream as well as retaining the functional knowledge from visual inputs in the ventral stream (van Polanen and Davare, 2015).

Afforded actions represented in mirror neurons

Ferrari et al. (2005) found that after training monkeys to have visual experience of watching tool usage actions, but not having the motor experience to manipulate the tool depicted in the movie, mirror neurons in their premotor cortex responded to the observation of the tool after 2 months of training. Interestingly, monkeys' mirror neurons responded to a non-executable but effective to achieve the goal action. Also human ventral premotor cortex are equally activated by possible and impossible actions (Costantini et al., 2005). Therefore, the mirror neurons are encoding actions that can achieve the end-goal, which in most cases are afforded action. What's even more interesting is that monkeys' mirror neurons do not respond to actions performed with non-food objects (Gallese et al., 1996). Therefore the mirror neuron activation is closely related to a desirable end-goal state and not by any object, which possibly indicates that mirror neurons activate when the observer actively retrieves the afforded actions.

Also, when the targeted objects were indicative of the corresponding action, or in other words, afforded actions were inferred from the target objects, the activations were observed in mirror neurons in F5 and inferior parietal lobule (IPL) (Jacob and Jeannerod, 2005). In general, matching the afforded action with the perceived action were proved to be in the mirror neuron regions (Bach et al., 2014)

Inferring information from context

Iacoboni (2005) indicates that BOLD activation in the right inferior frontal cortex and ventral premotor cortex (vPMC) were higher when the observer were presented the actions (grasping a cup) with different context of either to clean up or to drink from a cup. Inferring action from context during action observation involves STS, TPJ and anterior fronto-median cortex (Liepelt et al., 2008). Also, inferring physical properties of objects in natural daily scenes and retrieving relevant knowledge involves bilateral dorsal premotor cortex, SPL, and left supramarginal gyrus (Fischer et al., 2016).

Conceptual expectation influence physical movement goal processing and the movement prediction error occurs in IFG and IPL (Ondobaka et al., 2015). If the perceived grip type (precise or power grip) is incongruent with perceived intentions (i.e., reaching-to-grasp a mug full of coffee with a precision grip) from the contexts, response patterns in DLPFC and pMTG are altered differently by CSE stimulation and bias the grip type selection (Amoruso et al., 2018)

Study looking into the communications between neuron populations using tracing in monkey provides a strong evidence this information flow in a ventral connection (Nelissen et al., 2011). They found out that two pathways: one that connects upper STS to ventral premotor cortex (F5) and one that connects lower STS to AIP to premotor area F5a /b transmit information about either agent's intentions or immediate goal (objects) and implied that

visual information encoded in STS is different according to different end goals and then forwards to different part of premotor cortex.

1.5 Prediction in the Brain

1.5.1 Predicted action outcome in inferior frontalparietal cortex

I have mentioned above that an action-to-goal is “predicting the likely effect of an ongoing action” (Csibra and Gergely, 2007) so the observer can infer the action outcome or goal with visual inputs of the first few kinematics of an action sequence.

Action outcomes from processing visual information during object-oriented action are represented in the inferior frontal parietal network. BOLD signals in IPL, SPL, IFG, and the MFG are significantly more active during observing action kinematics toward an action goal (grasp-to-drink vs grasp-to-pour) that’s unknown to the observer before the action starts. More importantly, these regions can successfully classify between the two action outcomes (Koul et al., 2018). When left IFG was being interfered with tDCS, subjects’ performance on predicting the action outcome (the grasping objects) enhanced when excited and impaired when inhibited when watching reaching-to-gasped objects action (Avenanti et al., 2017). Right IFG including BA44 and BA45 is strongly activated when the observer doesn’t know the communicative intention of the observed hand action (e.g. sign language for “rain”) but still trying to use the hand action features to understand the action goal (Möttönen et al., 2016). IFG BOLD activity increases when the subjects answer the differences in action outcomes, irrespective of whether the subject paid attention to the performer’s intention or not (de Lange et al., 2008). Also, right IFG, right IPL, and left postcentral sulcus extending to aIPS responded similarly to the same outcome with different action kinematics (de C Hamilton and Grafton, 2008). Although there was probably still subtle kinematic

differences between repeated action kinematics for different action outcomes. Lastly, STS is sensitive to the action outcome and increase in BOLD signal change when the observer saw failed action by either human or machine arms(Shultz and McCarthy, 2012).

1.5.2 Irrational Action: action prediction error

So far I have talked about two processes: 1) predicting likely outcomes of actions (cognitive) and 2) making inferences about intentions (social) but I haven't talked about evidence for predicting actions when knowing the actor's action goal. During action observation, an agent's actions are assumed to be efficient and rational. If the observer's prediction doesn't match with what the observer's perceived action outcome or trajectory, the observer would be surprised and the perceived action would be categorized as irrational action (at least at first). In the infant study mentioned above (Woodward, 1998), infants looked longer, indicating surprise, when watching an adult performing an action with an inefficient reaching trajectory, indicating they were predicting the action trajectory before the completion of the action.

The top-down signal from other brain regions in which information from the environment is processed with affordance from physical properties provides an observer with limited action choices. The observer considered an action as irrational because the perceived actions were not included in the predicted action choices. In this section, I am going to first look at studies on irrational actions and the brain regions that respond to these irrational actions. Then I will further look at how irrational actions are an important direction to test the prediction models.

mPFC, pSTS or MTG, and IFG were found to be the brain regions that responded to irrational actions. Specifically, observing reaching actions over barriers violating the efficient and rational action assumptions increased response in MTG (Jastorff et al., 2011). Inefficient ball movement animations evoked responses in the observer's pSTS (Deen and Saxe, 2012).

Observing inefficient human ball manipulation over a barrier evoked response in aIPS and after observing the action, the subject’s rationality judgment correlate with response in mPFC (Marsh et al., 2014). When observing an irrational action movie, in which a whole body avatar is moving towards a barrier and the avatar is either performing an unnecessary action or not adjusting the action path over a barrier opening, right supramarginal gyrus, pSTS, and ventral to MTG, lateral occipital cortex increased in BOLD activity (Shultz and McCarthy, 2014).

When watching irrational hand-object movements, IFG and IPL activity increases during incongruent movements with the movement goal, such as the movement goal is to smell but the movement is to move the object to the ears (Ondobaka et al., 2015). Also, observing another person grasping the cup to drink using an inefficient power grip is associated with increased response in lateral occipitotemporal cortex and outstanding goal: moving the cup to ear is associated with increased activation in IFG (de Lange et al., 2008). Lastly, if the perceived grip type (precise or power grip) was incongruent with perceived intentions from the contexts (to drink or to clean), response patterns in DLPFC and pMTG were altered differently by CSE stimulation and bias the grip type selection (Amoruso et al., 2018).

Also, observing actions violated the biological assumptions evoked activation in frontal-parietal temporal cortex. Viewing human performers acting robot-like movements evoked response in bilateral aIPS and left EBA (Saygin et al., 2012; Urgen and Saygin, 2020). Watching stick-figures perform actions at reduced versus normal gravity is associated with general OTC, TPJ, and insula (Maffei et al., 2015).

Brain regions (mPFC, pSTS, IFG, and aIPS) responding to irrational action are the regions overlapping between ToM network, which were hypothesized to be involved in long-term traits and social inference, and the mirror neuron network, which involves evaluating kinematic properties of the target (Rizzolatti and Craighero, 2004). These regions are also where the top-down signal generated predictions from internal knowledge evaluated with sensory

inputs, as shown from irrational action studies above, and the sensory inputs here are the perceived action kinematics and outcomes. Understanding irrational action is the step between basic action understanding and theory of mind (Csibra and Gergely, 2007). Irrational action study findings reflect this by showing the brain regions involved in irrational action stimuli engaged both AON and ToM network. In support to this claim, Marsh et al. (2014) found that mPFC, right IFG, right IPL extending to TPJ, middle occipital gyrus, hippocampus and cingulate were modulated by action rationality.

The encoded prediction error’s functional role: update internal model and this updating shows up as the interaction between AON and MZN

The predictive models mentioned above have also tried to incorporate the findings from irrational action (Koster-Hale and Saxe, 2013; Bach and Schenke, 2017). Action is the predictive model’s output and testable layer. People use the probabilistic mapping from their internal knowledge to make behavioral predictions and test their predictions by observing agent’s action choices. If we consider the increased BOLD activity to irrational action as an activity signaling prediction error, then we can consider how we process irrational action as a test for our prediction compared with perceived input. Specifically, Koster-Hale and Saxe (2013) indicated that the increase in neural responses coded the difference between the expected stimulus value and the perceived stimuli value, which is revealed through the difference between expected action outcome and the perceived actual outcome. Further, Koster-Hale and Saxe (2013) hypothesized that there are two groups of neurons, where one encoded expectation or prediction and the other one encoded prediction error, and the relative activation from these two groups of neuron reflects in the BOLD activity. Therefore, BOLD activity decreased in pSTS when the observer was presented with the same action outcome (Hamilton and Grafton, 2006; Jastorff et al., 2011) and increased when the observer perceived irrational actions (Deen and Saxe, 2012; Wyk et al., 2009).

The prediction errors are further used in these models to revise the internal model (Bach and Schenke, 2017; Tamir and Thornton, 2018; Amoruso et al., 2018). Bach and Schenke (2017) proposed a hierarchical model which includes multiple levels; predictions are generated into a lower level for forthcoming movements, prediction errors occurs when the incoming sensory stimulation mismatches with the predictions. The model updates in a bottom-up fashion with the error backpropagating to the upper level in closest proximity and updating the upper level to minimize the prediction errors (Bach et al., 2014). For example, the prediction errors from pSTS could be further used to update a situation model, assumed to be in TPJ, and try to fit a new environment or new goal to re-evaluate the action fit to the environment first before updating the highest personal knowledge model.

The most likely actions are then selected from the one that minimizes the prediction errors at all hierarchical levels of action presentation (Kilner and Frith, 2007). In (Kilner, 2011)'s top-down generative model, the errors were compared at pSTS with the selected mostly likely action sequences from BA44. TMS study found that AON predicts the concrete levels of action representations (motor commands and perceptual expectation) while DLPFC was shown to bias action selection through top-down signal bias (Amoruso et al., 2018).

1.6 Conclusion and specific aims of the current work

By examining action observation as a predictive process within the action observation network, which encompasses brain regions such as the IFG, pSTS, and IPL, and dissecting the generative model through which observers anticipate action outcomes, our study reveals a significant indirect connection supported by a body of evidence. This sets the stage for a more comprehensive analysis of the neural mechanisms involved in the prediction of actions within the brain.

Here I will investigate the pSTS, IFG, and IPL's functional and statistical structure using the following approaches: To evaluate the coupling between regions of the AON while the observers are attending to different aspects of action, I will measure how the functional connectivity changes depending on the observer's cognitive state. Second, to evaluate whether sensory encoding and representation are altered by prediction of the upcoming visual inputs, I will create an anticipation of a specific action by cueing the observer, then evaluating information in the anticipatory interval and stimulus encoding interval separately. Lastly, to evaluate the hypothesis that prediction differentially strengthens cognitive networks prior to perceptual brain regions, I will evaluate how directed shifts in effective connectivity are modulated by the attentive goals of the observer (action kinematics or spatial goal). **Findings in my dissertation will pave the way to develop a better framework for action processing within AON by including neural mechanisms at different action observation stages and also at the observer's different cognitive states.**

Chapter 2

Configuration of the Action

Observation Network Depends on the Goals of the Observer

1. Introduction

The action observation network (AON) is a large-scale brain network that supports the perceptual encoding and recognition of actions performed by others (Molenberghs et al., 2012). Classically characterized as a frontoparietal system specialized for understanding goal-directed hand actions (Rizzolatti and Matelli, 2003), the complete AON more broadly supports our ability to represent many types of actions, to predict the likely outcome of goal-directed actions, and to make inferences as to the goals of others as derived through that individual's body movements (Thompson et al., 2019).

Despite the long history implicating the action observation network in action understanding, the nature of information and connectivity structure within the system is not yet fully clear.

When observing actions, there are multiple levels of abstraction at which the events can be represented, from the perceived kinematics to predicted outcomes of the actions or the hidden intentional state of the observer (Bach and Schenke, 2017; Thompson et al., 2019). Each of these is linked to distinct nodes within the AON. For example, empirical studies find evidence for mid-level representations in the left anterior intraparietal sulcus (aIPS) and inferior frontal cortex (IFC), such as the identification of unique action goals (de C Hamilton and Grafton, 2008), predicting the likely outcomes of actions (Koul et al., 2018; Möttönen et al., 2016), and representing violations of anticipated outcomes (de Lange et al., 2008; Shultz et al., 2011). In contrast, the posterior superior temporal sulcus (pSTS) is proposed to host lower-level representations that are more perceptually grounded (Masson and Isik, 2021; Pavlova, 2012; Pitcher and Ungerleider, 2021). The pSTS has neural signals that differentiate different action categories (Kable and Chatterjee, 2006), is viewpoint invariant (Grossman et al., 2010) and has activation patterns that are qualitatively modulated by the goals of the observer (Tavares et al., 2008). It is important to note that this distinction between abstracted and perceptual representations is not fully dichotomous, as previous studies have also proposed the pSTS to also contribute to representing the hidden states of others (Pelphrey et al., 2004; Schultz et al., 2004; Osaka et al., 2012).

Given the many possible levels at which actions can be interpreted, an important consideration in characterizing the AON includes the cognitive demands of the task (Kemmerer, 2021; Bach and Schenke, 2017; Vallacher and Wegner, 1987). Evaluating an action with the focus on how it is being achieved (the kinematics or implementation) versus why that action is being performed (the intent) alters patterns of brain activation in the AON and lateral temporal cortex (Spunt et al., 2010, 2016). Whereas evaluating how an action may be implemented strongly activates premotor (PMC), posterior parietal cortex and the left posterior middle temporal gyrus (pMTG), evaluating intent engages a more right lateralized system (Overwalle and Baetens, 2009). Moreover, when observers make a deliberate cognitive shift in the level of abstraction (i.e. from the more concrete how towards the more abstract why),

this shift in cognitive representation is reflected in the BOLD amplitude of the bilateral anterior STS (Spunt et al., 2016). Even shifting focus from the more social and intentional aspects of an action to the spatial properties of the event alters the patterns of activation along the STS (Tavares et al., 2008). These findings highlight the importance of cognitive context in action understanding and extend the AON to include brain systems supporting conceptual and semantic cognitive processes and form the basis for a proposed lateral stream in action understanding (Wurm and Caramazza, 2021; Pitcher and Ungerleider, 2021).

Therefore one goal of this study is to characterize functional connectivity within the AON during action observation, and evaluate how connectivity changes in conjunction with the goals of the observer. In a previous study using multivariate pattern decoding we found that whole-body actions were decoded more accurately in the right pSTS when attending to body kinematics (how an action is achieved) versus the actor’s identity (who is performing the action) or the goals (where are the target objects) (Stehr et al., 2021). Under those same attentional conditions, connectivity was strengthened between the right pSTS and the right inferior frontal cortex (IFC). We interpreted these findings as evidence for a sharpening of perceptual representations in sensory cortex mediated by top-down signals derived from internal models constructed in prefrontal cortex (Geng and Vossel, 2013; Sokolov et al., 2018; Patel et al., 2019; Kilner and Frith, 2007). The current study seeks to identify at a more granular level those regions within the IFC and also AON, including extended connected regions of the lateral temporal lobe that are modulated by shifting attention between the kinematics (the ”how”) or goals (the ”why) during action recognition (Spunt et al., 2010).

A secondary goal of this study is to more carefully characterize the functional systems modulated by top-down goals of the participant during action observation. The IFC is a functionally heterogeneous region, with a gradient of specialization throughout as revealed through a meta-analysis across task domains (Hartwigsen et al., 2018). While the more dorsal and anterior aspects of the IFC are more strongly engaged during cognitive tasks, posterior as-

pects are more closely related to somatomotor networks, and ventral aspects are more closely driven by social cognitive and emotional processing. With specific regards to action recognition, evidence shows the more concrete aspects of actions (i.e. how an action is achieved as conveyed through body kinematics) are represented on the more posterior IFC, whereas action goals (i.e. why an action is being executed) are associated with the more anterior extent (Kilner, 2011). Likewise, posterior parietal cortex has distinct hubs within which concrete somatomotor representations are different from the semantic and executive systems (Nummenmaa and Calder, 2009). Gradients of abstraction in the AON may reflect distinct targets of information pathways specialized for online visuomotor representations of actions versus the more durable ventral conceptual pathway (Rizzolatti and Matelli, 2003; Buxbaum and Kalénine, 2010; Binkofski and Buxbaum, 2013; Wurm et al., 2016).

Moreover, a large-scale effort is underway to reduce the complexity of brain networks to a small set of core functional networks that account for a significant proportion of the variance in brain states, as assessed in the resting state and often applied as network labels during task-related fMRI (Yeo et al., 2011; Schaefer et al., 2018; Kong et al., 2021). Using this approach, previous studies have identified dominant functional systems and topographic organizations within the inferior frontal cortex and posterior parietal lobe (Hartwigsen et al., 2018; Numssen et al., 2021). In this analysis we therefore adopt an atlas parcellation scheme that subdivides large regions of interest in the AON into small atomic parcels, assign functional networks labels to those parcels labels according to standardized functional network organization and draw inferences as to the larger functional systems that are modulated during goal-direction action observation.

2. Methods and Materials

2.1 Participants

Twenty-four healthy adults (8 male, 17 female) ranging in age from 21 to 42 years old (mean=24.7, SD=3.6) with normal or correct-to-normal vision were recruited from the University of California, Irvine campus and surrounding community. All participants gave written informed consent and all experimental procedures were approved by the University of California, Irvine Institutional Review Board. One participant was excluded from the analysis due to excessive motion during scanning.

2.2 Image Acquisition and Preprocessing

Full details on image acquisition and preprocessing can be obtained in Stehr et al. (2021). Briefly, images were acquired on a 3 Tesla Siemens Prisma MRI scanner (Siemens Medical Solutions) equipped with a 32-channel receive-only phased array head coil. T1-weighted images (1x1x1 mm) were reconstructed into native surface-based representations using FreeSurfer’s recon-all algorithm (<http://surfer.nmr.mgh.harvard.edu/>). Functional images were acquired with in-plane resolution = 2x2x2 mm (no gap) using multiband (multiband acceleration factor = 4), interleaved slice acquisition. Functional scans designed to localize the pSTS were acquired with TR = 2000 ms (69 axial slices) and scans designed to capture task-related modulations in functional connectivity were acquired with TR = 1500 ms (68 axial slices). Functional images were slice-time corrected, 3D motion corrected, temporally high-pass filtered (.01Hz cutoff) and field-map distortion corrected in BrainVoyager 20.6 (Brain Innovations, Inc.).

2.3 Stimuli and Experiment

Participants viewed 3 sec action animations depicting one of two avatars (a boy or a man) approaching a shelf, then crouching down or jumping up to reach a target box (Figure 2.1). The vignettes were viewed under one of three task instructions: attend to the actor's actions, attend to the actor's goal or attend to the actor's identity. Each vignette was constructed from 10 viewing angles, ranging from 80 degrees (left) to 280 degrees (right), with a 20-degree increment. The duration of the cue was 1s, followed by a 0.5 sec blank interval between cue and stimuli onset, with a 2.5 sec response period after the movie encoding.

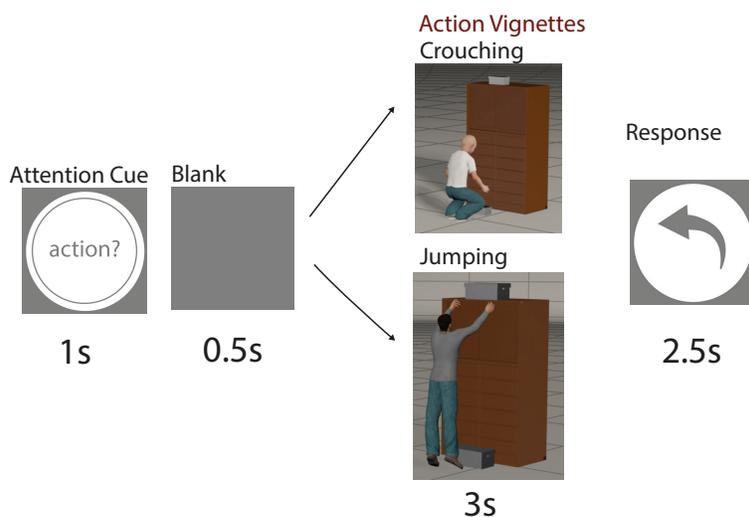


Figure 2.1: **Schematic of trial sequence.** Participants were cued as to which aspect of the vignette to attend (attend to action, goal or identity). The three second vignette depicted an open room as an avatar approaches a shelf, directs their gaze to one of the two boxes, then either jumps or crouches to retrieve an object. Participants were prompted to discriminate the action (jump or crouch), the goal (the box positioned high or low) or the actor (man or boy) in accordance with the attention cue.

Trials were separated by a 3, 4.5 or 6-s ITI, pseudorandomized within each run such that the hemodynamic response associated with each trial could be estimated independently using the least sum of squares (LSS) modeling approach (Mumford et al., 2012; Turner et al., 2012). Participants completed 24 trials per scan, with eight trials per attention condition per scan, and a total of eight scans.

2.4 Regions of Interest

The **right pSTS** region of interest (ROI) served as the seed for functional connectivity analysis because of its importance as the perceptual hub for the AON and the ease with which it can be identified using independent localizers (Pitcher and Ungerleider, 2021; Iacoboni et al., 2001; Grossman et al., 2010). In separate scans, observers viewed 18 sec blocks of point-light biological motion animations (1 sec each with .5 sec intertrial interval), alternating with 18 sec blocks of motion-matched control animations. The pSTS was identified as the region on the superior temporal sulcus with a significantly stronger BOLD response (FDR, $q < 0.005$) for intact vs scrambled biological motion, identified using a group-level random-effects GLM (Figure 2.2a).

IFC, PPC and STS+ were identified using a data-driven approach as the vertices on each subject’s surface that were functionally connected to the seed right hemisphere pSTS during observation of the action vignettes (all attention conditions). Connectivity was computed with the beta-series method (described below) using all trials unlabeled for the attention condition, derived from functional data that was projected onto the smooth white matter surface in native subject space. The individual subject functional connectivity maps were Fisher r-z transformed then projected into a common standardized vertex space, constructed using group cortex-based alignment in BrainVoyager (Frost and Goebel, 2012). The group t-score maps were thresholded at an uncorrected $p > 0.05$ with $z > 1.96$, which was deliberately conservative to be inclusive of all possible connected vertices on the surface.

The group functional connectivity map was then subjected to dimensional reduction into regions of interest using the 360 atom Glasser atlas parcellation, which specifies labels for Brodmann’s areas 44, 45, 45b, and 47 in the IFC (Glasser et al., 2016) (Figure 2.2b). The atlas was applied to native surfaces using Freesurfer’s *mriscalabel* then imported into BrainVoyager using custom library tools (<https://github.com/tarrlab/Freesurfer-to-BrainVoyager>), which

allowed custom selection of the native functional volumes within the atlas parcels. Additionally, atlas parcels in posterior parietal cortex (PPC), inferior frontal, or lateral and anterior temporal cortex (combined within an STS+ ROI) were included if they contained vertices with significant functional connectivity to the pSTS. To avoid any potential bias derived from the alignment procedure, the overlap between functional connectivity maps and the atlas parcellation was conducted on a template pilot subject that was not included in any subsequent analysis.

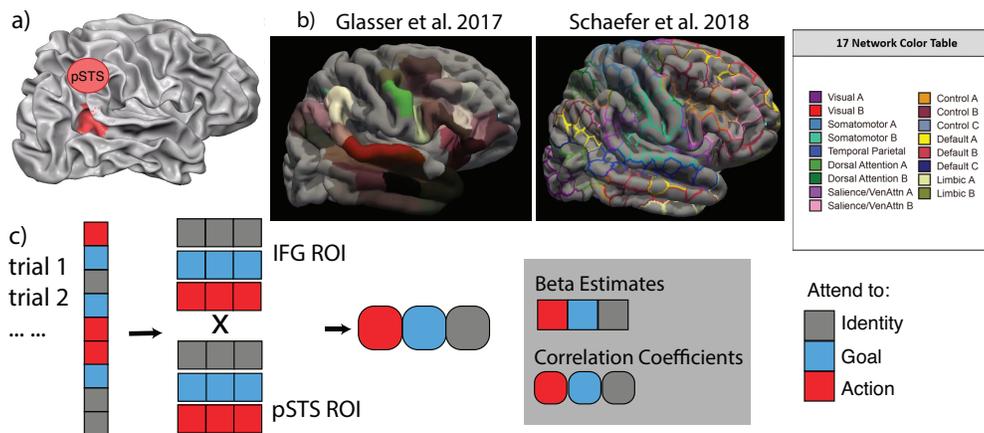


Figure 2.2: Defining ROI and Connectivity Method. a) The pSTS is identified using an independent localizer and a mixed effects model that treated individual subjects as random effects. The GLM analysis was conducted on vertices in a cortex-based aligned surface space. b) Functional connectivity to the seed identified large regions of interest, each of which was further divided into smaller parcels as defined by Glasser et al. (2016). Functional systems labels for each parcel were defined using the modal network system as defined by Schaefer et al. (2018). The 17 networks in that classification were combined into eight for simpler interpretation (i.e. the dorsal attention A and B were combined into a single dorsal attention). c) Functional connectivity between two parcels was calculated using the beta-series approach in which the connectivity was computed as the correlation (circles) between the timeseries of trialwise beta estimates (rectangles). Regions of interest used this approach with trials unlabeled by task condition. The main connectivity analysis further split the trials into unique types based on the task labels.

2.5 Functional Connectivity Analysis

Functional connectivity was conducted using the beta-series method (Rissman et al., 2004) which derives trial-based estimates of BOLD activity across the duration of the scan and is particularly robust when applied in event-related designs with short ITI and stimulus duration, as in the current study (Cisler et al., 2014). Trial estimates of neural activity were calculated using least sum of squares (LSS) design (Mumford et al., 2012; Turner et al., 2012) with all functional scans z-scored and concatenated. Trialwise beta estimates were computed using a fixed effects general linear model that iteratively modeled the predicted neural activity for one trial (a boxcar function for the 3 sec duration of movie watching for that trial, convolved with a hemodynamic response function). The design matrix also captured the predicted brain response for all other trials using a single regressor modeling the expected BOLD response for all the other trials except for the current trial being modeled. The design matrix also included the following nuisance regressors: all six rigid body motion realignment parameters and their Volterra expansion, and global signal as measured from the white matter and ventricles. Trials that contained volumes with instantaneous motion (FD) greater than 0.4 mm were censored (Power et al., 2014).

Functional connectivity between regions of interest were correlations computed at the parcel level separately for three conditions: attending to action, attending to goal and attending to identity (the baseline condition in which attention is directed away from the action features). Correlations were computed between the beta-series of any two parcels (Figure 2.2c), within each subject. The task-specific correlations were then Fisher r-z transformed and a pairwise t-test at the group level tested for significant differences in parcel-parcel connectivity as function of task.

2.5 Functional Network Assignment

Functional network assignment was achieved using the 17-network atlas labels (Yeo et al., 2011), which we consolidated to 8 networks (i.e. Default A, B, C combined into one network: Default). Note that there is no direct mapping between these labels and the Glasser et al. (2016) atlas. We therefore parcelled each native anatomy using the 1000 atom Schaefer system (Schaefer et al., 2018), which yields a high resolution parcellation with the associated functional system labels (Figure 2.2b). Because there are unique boundaries and vertex assignments between the Glasser et al. (2016) and Schaefer et al. (2018) atlases, functional labels were assigned based on the majority network within each parcel computed using custom scripts in Matlab R2017 (Mathworks, Inc.). Additionally, any parcel without a majority network accounting for 60% or more of the vertices was identified as a “multi” system parcel. See Hartwigsen et al. (2018) and Numssen et al. (2021) for a similar data-driven approach.

3. Results

3.1 Functional Connectivity to right pSTS seed

The right pSTS, which has been identified as a proposed input for the AON (Iacoboni et al., 2001), represents the transition from low-level sensory to higher-level cognitive representations (Pitcher and Ungerleider, 2021; Patel et al., 2019) and has the advantage of being easily identified using independent functional localizers (Grossman et al., 2010). We therefore localized the right pSTS using independent scans and identified additional regions of the AON using beta-series functional connectivity method with the right pSTS as a seed. This approach revealed three large bilateral regions of cortex with significant connectivity to the

pSTS during all task manipulations: 1) the inferior frontal cortex (IFC), 2) a region on the lateral and anterior STS (STS+) and 3) the posterior parietal cortex (PPC) (Figure 2.3). Because the regions of interest are large and likely reflect many unique cognitive processes within, we subdivided each ROI into smaller anatomical parcels based on a template atlas applied to individual subjects. In summary, there were 23 IFC parcels, 14 STS+ parcels and 6 PPC parcels identified in each hemisphere.

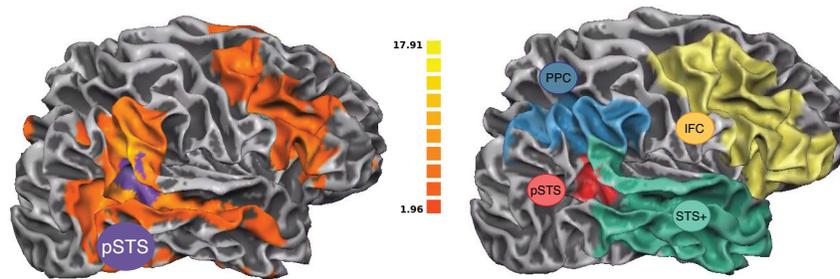


Figure 2.3: **ROI Identification.** Left: Group-level maps of functional connectivity ($z \geq 1.96$, $p < 0.05$ uncorrected) of the whole brain using the right pSTS (purple) as a seed. All trials regardless of attention manipulation were included in this analysis, which served to localize large regions of interest that were further subdivided into parcels. Right: Connectivity maps divided into regions of interest: inferior frontal cortex (IFC), posterior parietal cortex (PPC) and lateral and anterior superior temporal sulcus (STS+).

We found that many of the parcel connectivity strengths within the AON were modulated by the observer’s task such that connectivity was strengthened when attending to actions as compared to the other tasks (Figure 2.4, red bars). A binomial exact test comparing the functional connectivity to the pSTS when attending to action versus identity revealed significantly more parcels with stronger connectivity in the IFC ($p < 0.001$) and the STS+ ($p = 0.02$), but no significant modulations in functional connectivity between the pSTS and parietal cortex ($p = .51$). The comparison of attending to goal versus identity did not result in consistently modulated connectivity in any of the ROIs (binomial exact test, $p > 0.05$ for all)(Figure 2.4, blue bars).

Similarly, in a direct comparison of attending to the action (the how) versus the goal (the why), we found the vast majority of parcels in the IFC and the lateral and anterior STS

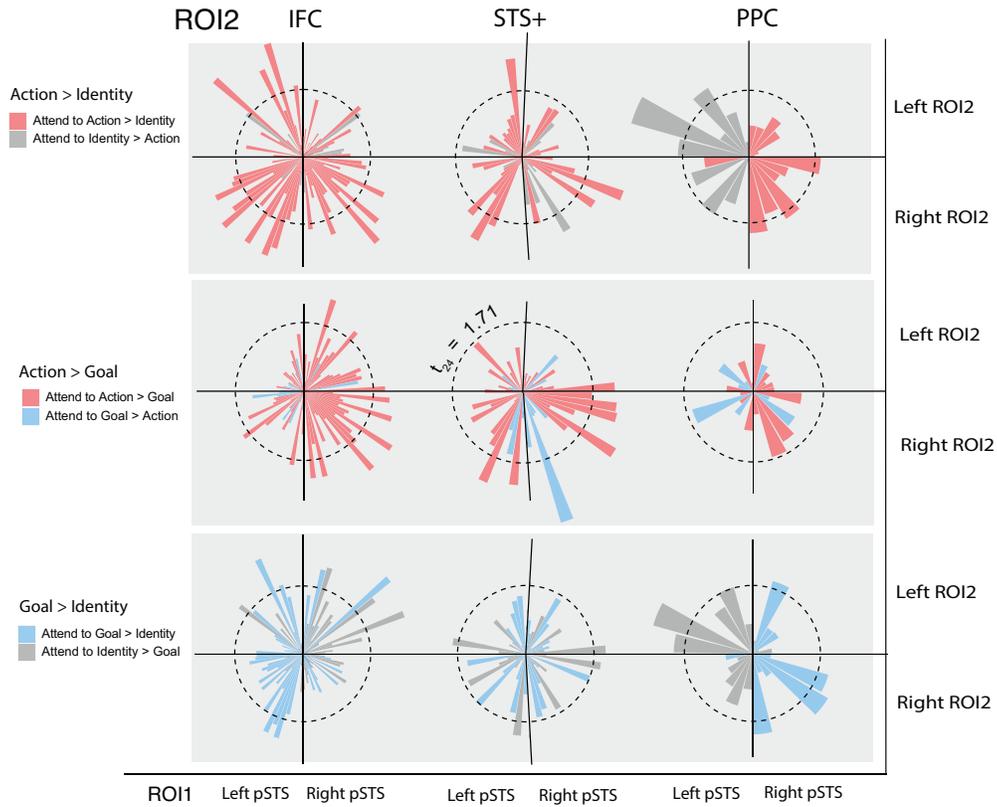


Figure 2.4: **Modulation of Connectivity from pSTS between pairs of tasks (attend to action, goals or identity) in the pSTS (the seed ROI) and the other ROIs.** Each bar indicates the magnitude of the difference in functional connectivity strength in a single parcel, computed across participants using the pair of conditions as indicated on the left. Right and left sides of each polar plot indicate connectivity to the right and left hemisphere pSTS, respectively. Top and bottom quadrants in each polar plot specify left and right hemispheres of the target ROIs. Top row: Attending to action (red) vs identity (gray). Middle row: attending to actions (red) versus goals (blue). Bottom row: attending to goals (blue) versus identity (gray) of the actors. Dashed line indicates the statistical significance threshold ($p < 0.05$).

increased connectivity to the pSTS when the participant attended to the agent’s actions as compared to the agent’s goals (binomial exact test, IFC: $p < 0.001$; STS+: $p < 0.05$; Figure 2.4, middle row). Those parcels with significant modulations in connectivity ($t > 1.71, p < 0.05$) were most frequently found between the right pSTS and right IFC (Bottom right quadrants of the polar plot). All but one parcel in the STS+ (right A4) that reached significance were more strongly connected when attending to actions as compared to goals.

3.2 Functional Systems connected to the pSTS

The IFC and STS+ are large ROIs that can be subdivided into many parcels and exhibit a wide range of functional heterogeneity (Hartwigsen et al., 2018). Therefore to gain deeper insights into these more complex functional regions of interest, we further characterized the parcels exhibiting modulated connectivity using the functional system labels as assigned by the updated Yeo et al. (2011) functional network system (Schaefer et al., 2018).

Among those parcels with significantly modulated functional connectivity between the pSTS and IFC (shown in Figure 2.5a), the dominant functional networks engaged were executive systems, including the frontal-parietal control network, the ventral attention network and the default mode network (Figure 2.5b). The parcels with connectivity between the pSTS and STS+ that were modulated by the task were more diverse, and included fronto-parietal control networks, the default mode network, somatomotor and limbic systems.

3.3 Functional Connectivity between the IFC, PPC and STS+

In addition to examining the connectivity to the seed pSTS, we extended our investigation to explore the modulated connectivity between IFC and PPC, as well as the lateral temporal STS+. Notably, these regions are characterized by large ROIs that can be subdivided into

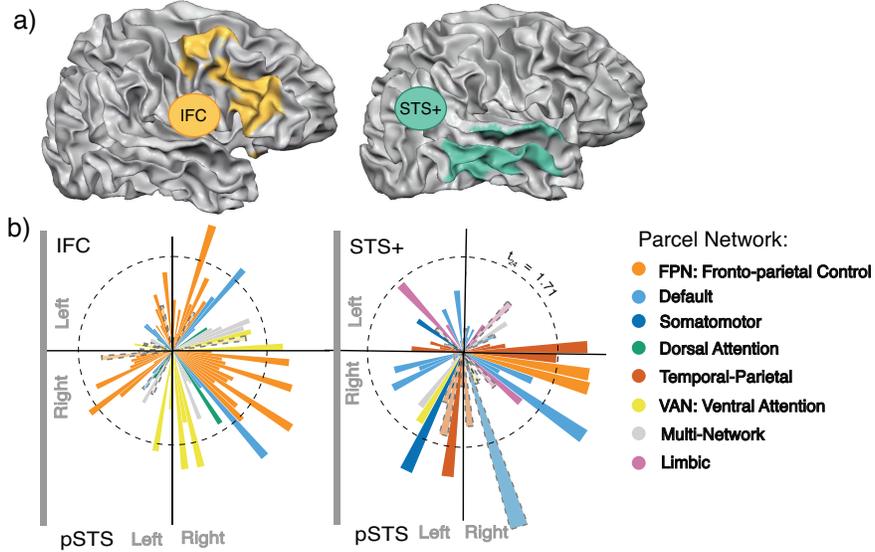


Figure 2.5: **Functional Network of the Modulated Connectivity.** Panel a) IFC and STS+ parcels with significant modulation in the strength of functional connectivity when attending to actions versus goals. Panel b) The magnitude of functional connectivity modulation (as shown in Figure 2.4), colored by functional network assignment as derived from Schaefer et al. (2018). Bars with solid border have magnitude indicating stronger connectivity during actions versus goals. Bars with dashed borders indicates parcels with strong connectivity during goals as compared to actions.

many parcels, and exhibit a much wide range of functional heterogeneity as compared to pSTS (Hartwigsen et al., 2018; Numssen et al., 2021).

To assess the modulated functional connectivity between each parcel from the ROI pairs, we calculated and organized the connectivity results according to the functional system. Figure 2.6 shows the connectivity modulations between the parcels in the IFC and the STS+ and PPC. Among those many parcels, connections that were modulated by attention (action identification greater than goal inferences) were dominated by the executive fronto-parietal control network and ventral attention system in the IFC and right PPC, with some engagement of the temporal parietal network and default mode networks in the STS+. Modulated connectivity between these regions was strongly bilateral in the IFC, but more right lateralized in the STS+ when attending to actions.

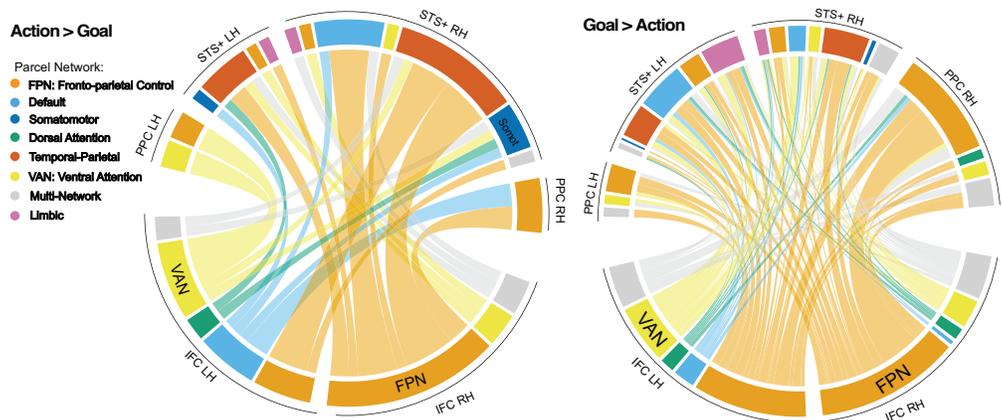


Figure 2.6: **Chordgram** showing the functional system engaged between IFC, STS+, and PPC ROIs when subjects attend to actions more than goals (left) and attend to goals more than actions (right). Each strand represents the proportion of connections between two functional networks that are significantly modulated ($p < 0.1$) by the attentional instruction. In each of the chordgram, colors of the strand represent the functional network as assigned from the IFC (shown on the bottom ring). Target strands are ordered as STS+ (central) to PPC (peripheral).

3.4 The univariate BOLD response

To determine whether the increase in functional connectivity observed during action observation reflects strengthening of connections within the AON or the recruitment of additional neural populations for that task, we also evaluated the univariate response modulations for each of the parcels in our regions of interest, relative to a mean MR signal baseline (Figure 2.7). We found the BOLD responses in three of the regions of interest to have amplitudes that differed when attending to goals versus actions, with all three more strongly activated by goals (right PPC: $t = 3.06, p < 0.01$; left PPC: $t = 1.84, p < 0.05$; right IFC: $t = 2.19, p < 0.05$). It is worth noting that although univariate response in right IFC is significantly stronger during goal inference comparing to action, this difference is mainly driven by two parcels out of 23 parcels, while the univariate response in parietal cortex is more consistent throughout the ROI (See Figure 2.8). There was no effect of task on the univariate BOLD in the left IFC, bilateral pSTS or STS+.

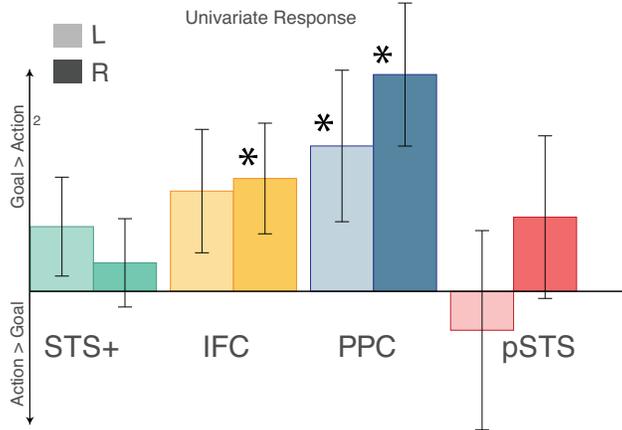


Figure 2.7: **Univariate Analysis Results.** Univariate response modulation within each region of interest, all parcels combined. The BOLD modulation is computed as the difference in beta amplitude on trials with attention directed to goals versus action, and vice versa. Amplitude above zero indicates a stronger BOLD response for trials with attention directed to the goals versus actions in the vignettes. Scores below zero scores indicate stronger BOLD when attending to actions (versus goals). Each bar reflects the average ROI univariate response across all subjects and parcels within the ROI. Lighter colored bars show the univariate response in the left hemisphere and the darker colored bars show the BOLD response in the right hemisphere. Asterisks (*) indicated significant modulation of the BOLD response by the attention condition (in the bilateral PPC and right IFC ROIs).

Table 2.1: Univariate results table showing the repeated measures t test that compares the average BOLD response between the two attention tasks. A: action, G: goal, I: identity.

		A > G		G > I		A > I	
Hemisphere	ROI	p-value	t	p-value	t	p-value	t
LH	STS+	0.13	-1.16	<0.01 *	-3.61	<0.01 *	-5.86
	IFC	0.07	-1.5	0.96	1.88	0.80	0.842
	PPC	0.04 *	-1.84	1	4.01	1.00	2.93
	pSTS	0.45	-0.12	<0.01 *	-3.66	<0.01 *	-3.81
RH	STS+	0.24	-0.709	<0.01 *	-6.34	<0.01 *	-6.6
	IFC	0.02 *	-2.19	0.38	-0.317	0.03 *	-2.02
	PPC	<0.01 *	-3.06	0.92	1.44	0.174	-0.959
	pSTS	0.21	-0.83	0.90	1.29	0.727	0.612

4. Discussion

The goal of this study was to characterize how connectivity within the action observation network changes depending on the cognitive goals of the observer. We identified three large regions of interest, all functionally connected to the pSTS during action observation: the

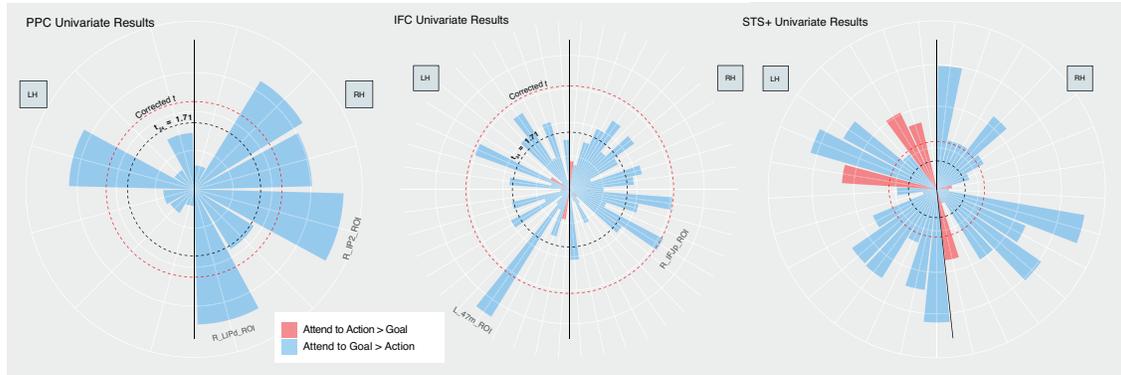


Figure 2.8: Univariate response results in PPC, IFC and STS+ ROI. each bar represent one parcel and the color indicates whether the response is stronger for action(red) or goal(blue).

IFC, PPC and STS+. The PPC and IFC are core components of the action observation network commonly also identified as having mirror neuron properties (Caspers et al., 2010). The lateral temporal lobe, which is included as part of STS+ in our current study, is not traditionally considered part of the AON. However, the lateral STS is associated with increasing levels of abstraction in action understanding (Spunt et al., 2016) and narrative understanding in context of naturalistic viewing, both of which are relevant to this study. The lateral STS+ is included in recent proposals of a conceptually-driven lateral pathway in the temporal lobe supporting semantic knowledge for action understanding, most apparent when actions are directed towards manipulating objects (Pitcher and Ungerleider, 2021; Wurm and Caramazza, 2021; Buxbaum and Kalénine, 2010).

We found that connectivity within this extended AON was modulated by the cognitive state of the observer, consistent with our previous report of sharpened representations in the pSTS when attending to kinematics (Stehr et al., 2021). The present results found that the majority of the parcels within these large regions of interest were more strongly connected when attention was directed to *how* the action was being accomplished (action kinematics) versus *why* the action was taken (actor’s goal). These modulated connections between pSTS and the IFC and STS+ largely reflected the engagement of default mode network (DMN) and frontoparietal control networks, emphasizing the importance coordinated activity of the

AON with executive systems to support the observer’s goals.

We do not interpret our results as evidence for more neural engagement during the different attention conditions. Stronger connectivity to the pSTS was true despite no task-related univariate modulations in the pSTS, left IFC or STS+. Only the IPL and right IFC had univariate activations modulated by task, and in both cases these regions were more strongly activated by attention directed to goals (consistent with Hamilton and Grafton, 2006; de C Hamilton and Grafton, 2008). This finding highlights how coactivation and connectivity reflect two unique mechanisms of engagement, with the former reflecting robustness of the neural response and the latter reflecting coordinated network activity (Bressler and Menon, 2010). Attending to actions strengthens subpathways within the AON that target the pSTS and relevant executive control networks, and this connectivity is independent from the size of the neural population localized to each region of interest (as assessed by the univariate BOLD). Together our findings demonstrate that the cognitive state of the observer is a powerful modulator of the functional architecture of the action observation network.

In the brain, the AON is closely associated with sensorimotor neurons in premotor cortex (PMC) and the anterior intraparietal sulcus (aIPS), both implicated in observing reaching and grasping movements. The AON also includes neurons in the inferior frontal cortex that represent action concepts and goals (Grafton and Tipper, 2012; Orban et al., 2021; Molenberghs et al., 2012; Cook et al., 2014). In addition, the AON includes brain regions that support the recognition of non-grasping effectors (i.e. faces and whole bodies) and actions that are intransitive (i.e. not directed towards objects), which includes the anterior region IFC and dorsal region IPL, in addition to the posterior superior temporal sulcus (pSTS) and the posterior middle temporal gyrus (pMTG)(Caspers et al., 2010; Overwalle, 2009). These more posterior regions lack motor properties and instead have neural populations tuned to body and face kinematics (Pitcher and Ungerleider, 2021) and, in the left hemisphere, represent conceptual action information that generalizes broadly across stimulus formats

(Wurm and Caramazza, 2021). These more perceptual and conceptual representations are proposed to serve as input to the frontoparietal systems such that they facilitate action prediction, recognition and understanding (Rizzolatti et al., 2014; Iacoboni, 2005; Cook et al., 2014).

Here we find that the regions of the IFC more closely linked to action observation are parcels in the fronto-parietal, ventral attention networks and the default mode networks. Rather than motor function, we infer our results to reflect the coactivation of the AON with cognitive-executive functions of the IFC, perhaps serving as a controller for binding relevant subnetworks within AON (Spreng et al., 2010). The nature of the connectivity depends on the subject’s goals and therefore the attended features during the action vignettes, which implies that the coordination of the systems promotes binding of information across established representations (Cole et al., 2013).

Our findings are consistent with the proposed dual pathway models of the action observation system (Rizzolatti and Matelli, 2003; Buxbaum and Kalénine, 2010; Wurm and Caramazza, 2021). In these models, online visuomotor transformations are supported by a dorso-dorsal pathway that is distinct from knowledge-based representations of actions, objects and their affordances. The ventral “conceptual” pathway for action understanding is linked to semantic and conceptual knowledge representations (Binder and Desai, 2011), relating that conceptual knowledge to object affordances (Wurm and Caramazza, 2021), and to social communication more generally (Pitcher and Ungerleider, 2021). In the current study, we found the ventral conceptual pathway to be more strongly engaged by directing attention on the kinematics of the observed actions rather than attending to the actor’s goals in the vignettes.

One important distinction between the proposed models of conceptual action pathways and the results obtained here is in the difference in laterality. While in our current study we found modulated pathways more strongly in the right hemisphere, in dual pathway models the ventral pathway is hypothesized to be left lateralized. For example, the left aIPS is

proposed to serve as a hub for integrating visuomotor and object affordance information when observing hand-object interactions (Orban et al., 2021). Similarly, the left posterior temporal cortex is linked to action representations that generalize across verbal and visual modalities, particularly in the context of grasping actions directed towards tools (Wurm and Caramazza, 2019). In contrast, the results from this study using whole-body action vignettes is more consistent with findings of right-lateralized activation in the pSTS when observing animations of bodies in action versus left-lateralized when reading action verbs (Bedny et al., 2008).

The lateral STS is not typically considered part of the canonical AON, and so we believe that there is an importance of a naturalistic context in our findings. Action animations viewed in more naturalistic settings, such as during movie watching or while listening (or reading) a narrative, are associated with bilateral activation on the lateral STS (for review, see Jääskeläinen et al., 2021). These brain regions are also among those characterized by neural populations with long temporal integration windows such that the information represented within reflects the accumulation of context extracted over long durations rather than moment-to-moment changes within a scene (Hasson et al., 2008; Chen et al., 2016; Yeshurun et al., 2017). These features are consistent with evidence that the bilateral lateral STS has neural signals that capture the essential meaning of narratives that make up episodic scenes (Baldassano et al., 2017). Related work shows that distinct regions of the bilateral lateral temporal lobe weight differentially on the unique multimodal features that carry this episodic information (Derderian et al., 2021), including actions (Lahnakoski et al., 2012). Further studies are required to more closely link the neural mechanisms of episodic narratives with the action observation network.

In conclusion, the current study reveals that goals of the observer alter the network structure of the AON, both in terms of functional connectivity and univariate BOLD responses. We found that the bilateral IFC and pSTS, STS+ and bilateral pSTS had strengthened

functional connectivity when subjects attended to action kinematics, while the magnitude of neural activity in the bilateral parietal cortex and right IFC was larger when attending to goals. These findings reveal two different neural mechanisms in the AON during action observation, co-activated networks versus localized changes in neural activation within regions of interest. Further, the modulated connectivity we observed was strongly right lateralized and largely involved fronto-parietal control network, indicating the importance of cognitive-executive functions in the AON. These results also suggest that the cognitive goals of the observer may strengthen cross-network connectivity between the AON with neural mechanisms that encode naturalistic action viewing in the temporal cortex.

In considering the findings of this study, it is important to acknowledge several limitations that may have influenced the results. First, our research is restricted to the use of only two types of actions, both of which involved whole bodies and were directed towards an object target. This narrow selection of actions potentially impacts the observed connectivity patterns and may limit the generalizability of our findings to a broader range of actions. It is worth noting that not all observed actions possess a specific target or object. For instance, actions like walking do not have a designated target, while social actions such as waving or hugging are directed towards living beings. This distinction between mentalizing and action should be taken into account when interpreting the results, as it may have influenced the connectivity patterns observed in our study (i.e. Overwalle, 2009). Another caveat to consider is the limited range of actors used, particularly the gender representation. The predominance of one gender in avatars portrayed may introduce biases or confounding factors that could influence the observed connectivity effects. Lastly, the use of avatars in our study, may have implications for the generalizability of our findings to real-world contexts. Future research should aim to address these limitations by incorporating a more diverse range of actions, considering a balanced gender representation, and employing stimuli that are more naturalistic in reflecting real-world social interactions

Acknowledgments

This study was supported by the National Science Foundation BCS-1658560 to EG and BCS-1658078 to JP.

Chapter 3

Neural Decoding and the Influence of Predictive Cues on Action Observation Network as Revealed by Multivariate Pattern Analysis

3.1 Introduction

Neuroimaging studies find that attention directed to action goals, rather than the actions themselves, increase univariate activity in IFG and IPL (de C Hamilton and Grafton, 2008; Zhou et al., 2023), and that the spatial activation patterns in these regions have information that classifies anticipated action goals prior to observing the final action outcome (Koul et al., 2018). Likewise, univariate studies show that the brain regions in the action observation network (AON), including pSTS, IFG, and IPL, have increased BOLD responses when the perceived action kinematics are contrary to the expected action trajectories or outcomes

Shultz et al. (2011); ?); Wyk et al. (2009). Together this implies that the viewer is actively constructing a mental model based on available contextual information at different stages of action observation.

One approach for constructing that mental model is action prediction. The action prediction theory is a theoretical framework about an observer's mental model such that the observer generates internal visual predictions of the upcoming actions and action outcomes to later match with the incoming perceptual input (Tamir and Thornton, 2018; Bach and Schenke, 2017). Using this the constructed predictions, the viewer encounters surprise if the perceived actions violate the predictions. These prediction errors are the signals indicating the need for the observer to re-evaluate the expected action into a new environment. The action prediction theory is prevalent in research studying motor response such that the constructed higher-level goals propagate downwards to generate motor code and therefore the action mirroring in various motor-related brain regions is predictive for the upcoming action (Csibra, 2008).

In the action prediction model, the pSTS is hypothesized to be influenced by IFG through a top-down generative process Kilner (2011); Zhou et al. (2023), much in the same way that conceptual expectations influence motor planning and goal processing during physical movements Ondobaka et al. (2015); Amoruso et al. (2018). In one specific implementation of this model, the pSTS is hypothesized to be influenced by IFG through a top-down generative process (Stehr et al., 2021). Indeed, when observers direct their attention to specific actions, functional connectivity between the right hemisphere IFG and pSTS increases (Zhou et al., 2023). This is true despite the univariate response being stronger in the IFG when participants attend to the goals of the action being performed.

To date there have not been studies that systematically evaluate the neurophysiological basis of prediction process separate from action observation. One of the reasons may be because the prediction and perceptual encoding processes often co-occur rapidly therefore

the neural response for each of these two processes may be blended. In this study I examined information available in the neural representations during the mental planning phase, before sensory coding, in different regions of AON.

We independently measured information available in the neural representations during the mental planning phase, before sensory coding and evaluated the information in the prediction interval and compared it to sensory encoding interval's. Recent findings by Monaco et al. (2019) have shown that neural information indicative of upcoming hand actions is encoded before the action itself is observed. Similarly, Hudson et al. (2018) established that when subjects are aware of the validity of predictive cues, the intention-action prediction effect intensifies. Complementary research by Avenanti et al. (2017) demonstrated that modulating the left IFC enhances the prediction of action outcomes, with intentions for forthcoming actions being decoded in the dorsal and ventral visual streams, including the ventral premotor (vPM), parietal, and left lateral STS.

In the current study we gave subjects the cue of the upcoming whole body action performed by an avatar, and then the subjects responded as quickly as possible when the stimuli started. Advance cues have the potential to set the stage for the neural population encoding (Kok et al., 2012) or alternatively could impact post-sensory processes in decision-making (Summerfield and Egner, 2009). On the basis of previous evidence for improved classification accuracy in the pSTS with directed attention (Stehr et al., 2021), we hypothesized that the cue will improve the neural representations during the sensory encoding interval. Here we hypothesized that action classification accuracies can be classified stronger in the perceptual areas such as right pSTS and also hypothesized that IFG and other AON regions might demonstrate predictive coding by showing the action can be classified when cued during the mental planning phase.

3.2 Methods

Subjects and Stimuli. 30 subjects participated in the current study. All subjects were given and signed a written consent form before the experiment. The protocol was approved by the University of California, Irvine Institutional Review Board. Subjects viewed brief (2 seconds) action vignettes depicting one of two types of action: jumping up or crouching down 3.1. These two action categories were chosen because previous studies have successfully associated these actions with unique neural populations in the pSTS and IFG (identified through multivariate decoding or rapid adaptation; Ariani et al. (2015); Grossman et al. (2010); Kable and Chatterjee (2006); Stehr et al. (2021). Prior to viewing the vignette, subjects were cued with the upcoming action, with cues having a 83% reliability (figure 3.1). Trials proceeded with an extended delay interval (2, 3.5 or 5 sec) where an anticipatory BOLD response could be measured independently from the encoding interval Mumford et al. (2012). Subjects then were shown the action vignettes and were asked to report as quickly as possible whether the perceived action matched the cue or not. Trials were preceded by an extended delay interval (3-6 sec anticipatory interval) in which the BOLD response could be measured independently from the encoding interval for each trial. In total participants completed 8 runs with 24 trials in each run (10 validly cued, 2 invalidly cued, 12 uncued trials, all pseudo-randomized), resulting in a total of 192 trials for each participant.

Neuroimaging. Participants were scanned at the Facility for Imaging and Brain Research at the University of California, Irvine, on a 3 Tesla Siemens Prisma MRI scanner (Siemens Medical Solutions) equipped with a 32-channel receive-only phased array head coil. High-resolution anatomical images were collected using a single T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence with 176 sagittal slices and 1 mm isovoxel resolution. Functional images were acquired using a T2*-weighted gradient recalled echoplanar imaging multi-band pulse sequence (cmrrmbep2dbold) from the University of Minnesota Center for Magnetic Resonance Research (CMRR). Scans were ac-

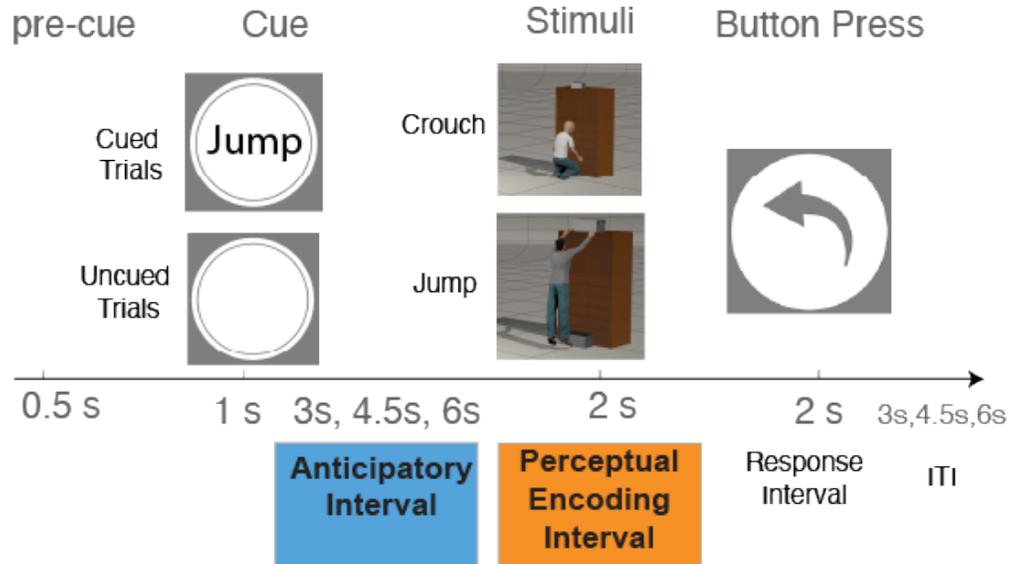


Figure 3.1: **Stimuli and Design** For each trial, participants were first given a cue about the upcoming actions during cued trials, and empty circle for uncued trials. Anticipatory interval was the time window following the cue and lasted from 3 to 6 seconds pseudo-randomized for each trial. After the anticipatory interval, participants then watched a 2s action vignette with 83% of the trials depicted action correspond to the cue given in the current trial. Participants responded with button to the action depicted in the stimuli both during the stimuli and after the stimuli if they haven't responded. At the end of each trial, the intertrial interval was randomly assigned between 3 to 6s.

quired with the resolution = $2 \times 2 \times 2$ mm (TR = 1500 ms TE = 30 ms, multi-slice acquisition factor = 3) and multiband factor = 4. EPI images with opposite phase- encoding direction were also acquired to correct for suscep- tibility distortions in each participant's functional data.

Preprocessing. Anatomical and functional data were preprocessed using fmriprep (citation, see appendix/supplementary materials). In brief, cortical reconstructions were produced using FreeSurfer's recon-all and the BOLD time-series were resampled onto the fsaverage standard surfaces for future analysis.

Regions Of Interest We defined the AON using parcels as identified in our previous study (Zhou et al., 2023) which includes IFG, pSTS, IPL and the lateral STS, which is

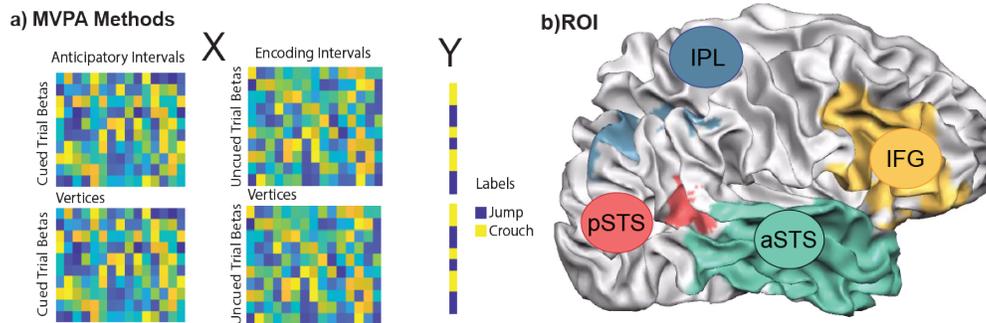


Figure 3.2: **MVPA analysis in AON ROI** A) Multivariate pattern analysis (MVPA) was trained and test to classify the actions depicted in the stimuli for activation patterns during anticipatory interval and perceptual encoding interval separately. B) Regions of Interest in the current study includes bilateral pSTS, bilateral aSTS, bilateral IFG, and bilateral IPL. Each region was further divided into smaller parcels defined by Glasser et al. (2016)

not traditionally included with the AON but is relevant to action semantics and narratives (Pitcher and Ungerleider, 2021; Lahnakoski et al., 2012; Yeshurun et al., 2017) (Figure 3.2 B). These regions were first defined using whole brain functional connectivity from pSTS during the action observation in an independent set of subjects. Since regions varied in size and also these regions are complex, we applied an atlas parcellation (Glasser et al., 2016) to divide and define the ROI. In the current study, we look at the action information both at the parcel level and also at the whole ROI level.

Univariate analysis To elucidate the variations in BOLD activation levels among different cueing conditions, we employed a General Linear Model (GLM) to quantify these activation levels. Each cueing condition—uncued, valid cue, and invalid cue—was modeled by convolving a boxcar functions specifying the duration of the post-cue interval (anticipatory) and the duration of the movie interval (encoding) with a standard hemodynamic response function (hrf). Condition-specific betas were estimated for each subject and region of interest (ROI). Statistical significance of the differences between cueing conditions was evaluated using paired t-tests.

MVPA methods. The time series from each voxel in the ROIs was first z scored across time (see Figure 3.2). The trial-by-trial activation patterns of BOLD response in each brain

regions were derived using the least squares separate (LSS) design (Mumford et al., 2012; Turner et al., 2012). In the design matrix, any trials in which too much head motion occurred with a threshold of $FD > 0.4$ were annotated as outliers and scrubbed from further analysis (Power et al., 2014). Trial betas were mean centered within each run (Lee and McCarthy, 2016) and averaged within each run to have one estimated beta for each action label.

To assess the statistical structure within the Action Observation Network (AON), the primary analysis evaluated the potential benefits of cues during the perceptual encoding interval. In this process, two types of actions were classified during the anticipatory interval for both cued and uncued trials. Support vector machine (SVM) analyses were carried out separately for each subject within regions of interest (ROI) including the posterior superior temporal sulcus (pSTS), inferior parietal lobule (IPL), and inferior frontal gyrus (IFG). The SVM used a linear kernel and a constant cost value of 1 to classify the actions (jump or crouch) depicted in the movie regardless of what subjects answered. The classification was performed using a 8-fold leave-one-run out cross-validation. The resulted cross-validated classification accuracies were made from the held-out test set with one test data for each label. Subsequently, the classification accuracies were first averaged across the 8 folds and all the classification accuracies for each subjects were compared across 30 subjects for each ROI.

Statistical analysis. To evaluate the effect of the cue on classification accuracy, we compared the differences in classification accuracy when subjects were given a cue vs not, within each subject and ROI. Statistical significance of the differences between cued and uncued classification accuracies was assessed using paired t test with a single-tailed significance level of $p < 0.05$ (uncorrected). Also, in order to evaluate whether the classification ability of each ROI, a one sample t-test for the null hypothesis of classification accuracy greater than 0.5 was used.

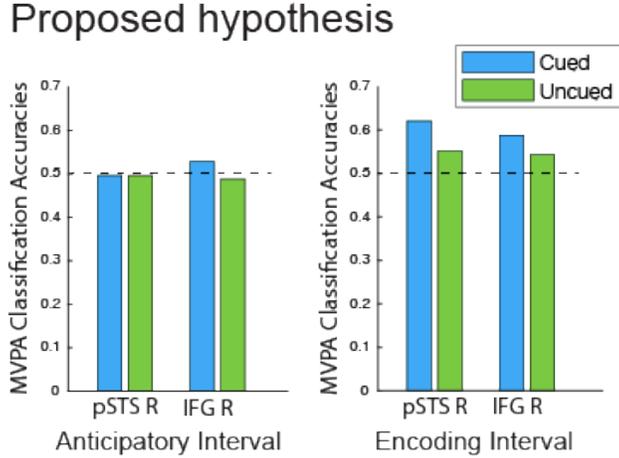


Figure 3.3: **Hypothesis** Hypothesis of the current study. Left: During the anticipatory interval, we expect the sensory driven right pSTS to not be able to classify the actions but right IFG would have the action represented. Right: During the movie interval, we expect that the right pSTS would classify the action and even better when given a valid cue. We also expect improved classification for validly cued trials in the right IFG, but less than the pSTS due to the specificity of action representations in the pSTS.

3.3 Results

Movie encoding and anticipatory interval. Based on previous reports (Stehr et al., 2021), we hypothesized that the classification accuracy for labeling the actions in the cued trials would be superior to that of uncued trials in the pSTS (see Figure 3.3). First we conducted the analysis for each ROI, such that we classify two classes of actions being depicted in the movie during the movie interval. The classification were performed using the parcels within the ROI as features in the classifiers (for example, if right pSTS consists of 2 Glasser parcels, then the classifier was trained on 2 features). The results showed that the improvement in classification accuracy in the left pSTS was significant (repeated measures t test, $t = 2.02$, $p < 0.05$, uncorrected) when subjects were given a valid cue comparing to uncued trials. This effect, however, is in the context of relatively poor classification of the actions overall (Figure 3.4 Right). A t-test on action classification accuracy against chance (with cued and uncued trials pooled) revealed none of the ROI to have classification accuracy significantly above 50% (Figure 3.4 Right).

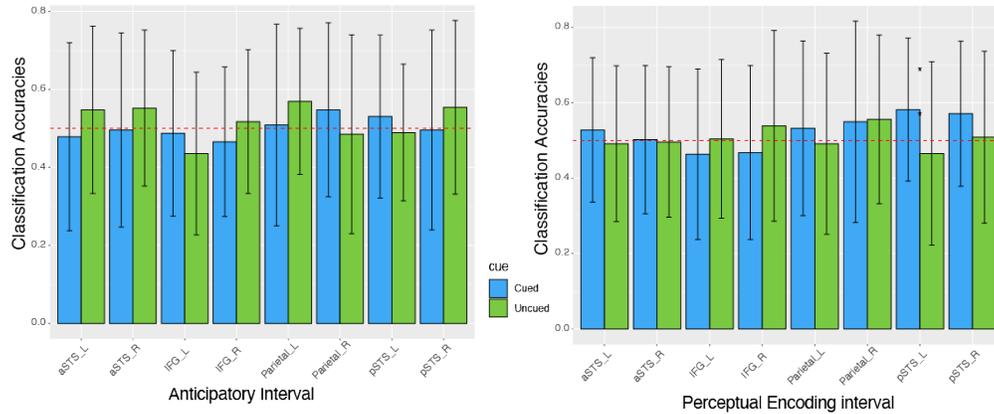


Figure 3.4: **MVPA Results.** Left: Classification accuracies during the anticipatory interval for cued trials, including valid and invalid cues, and for uncued trials. Classifiers trained on each ROI using the parcels within the ROI as features. Right: Classification accuracy during the perceptual encoding interval. Here the left pSTS showed significant difference between cued and uncued trials but note the uncued condition did not classify significantly above chance.

We then evaluated the patterns within each parcel in the ROI and conducted the same classification in the traditional MVPA analysis such that each vertex in the parcel was considered a feature in the SVM classifiers. Statistical significance between the two conditions was assessed with a repeated measured t-test for each parcel and a one sample t-tests evaluated whether overall classification accuracy (using both cued and uncued trials) was above 50% ($p < 0.05$, uncorrected). Figure 3.5 displays the parcels that reached statistical significance for both tests, indicating that these parcels not only had the classification accuracies that meaningful improved by the cue, but also across the cueing condition the classification accuracies were significantly above 50% across all the subjects. Results showed that during anticipatory interval, the parcels AVI and FEF in right IFG, and STV in lateral STS and during perceptual encoding interval one parcel p9-46v in right IFG showed significance.

To evaluate the trends within each ROI, we averaged classification results from each of the parcels within the ROI (Figure 3.6). Here the results revealed both the averaged classification accuracies and also differences between cue and uncued conditions within each subjects. During the perceptual encoding interval, this analysis revealed improved classification during

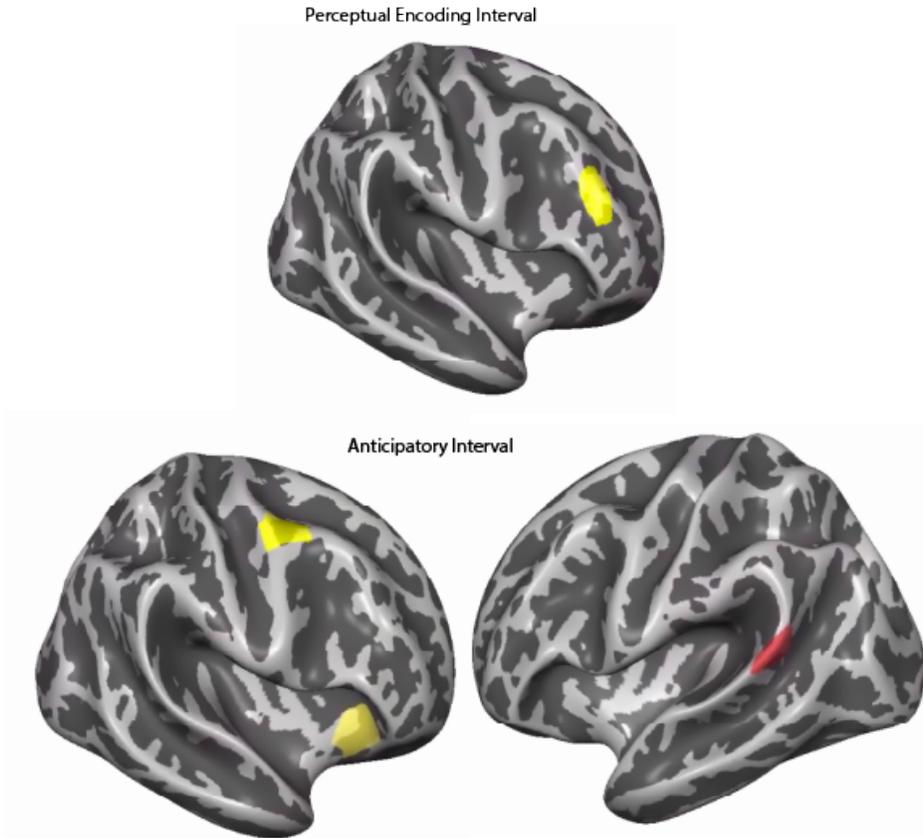


Figure 3.5: **Univariate Analysis Results.** MVPA results for each parcel. Figure shows the parcels in which the the MVPA classifiers trained on the activation patterns both classified the actions above chance across both cueing conditions and also the classification accuracies that were significantly higher during the cued trials than uncued trials by testing the within subject differences using a paired t test ($p < 0.05$, uncorrected). These parcels are p9-46v during the perceptual encoding interval. AVI, FEF in the right IFG, and STV in the lateral left STS during anticipatory interval.

uncued condition versus cued condition in the right pSTS only ($t = 2.04$, $p < 0.05$). During the anticipatory interval, left lateral STS ($t = 1.94$, $p < 0.05$) and right IFG ($t = 2.27$, $p < 0.05$) had improved classification accuracy during cued versus uncued condition.

Independence between anticipatory and encoding interval. We computed the correlation between the BOLD activation patterns (trial-wise beta estimates) during the anticipatory and perceptual encoding intervals, following the poor classification observed during the anticipatory phase. This was done with the aim of investigating whether there is an independent pattern of information processing during the anticipatory and perceptual encoding

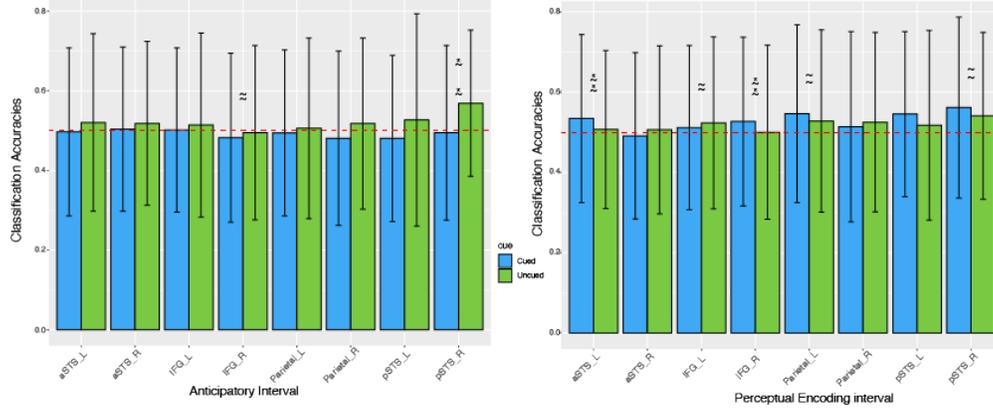


Figure 3.6: **MVPA Results.** Left: Average classification results when the classifiers were trained and tested in each parcel using the vertices within the parcel as the features in the classifiers. The cross-validated classification accuracies were then averaged across the subjects and parcels in each ROI. The asterisks indicate ROIs with significant differences between the classification accuracy for the validly cued and uncued conditions, evaluated using a paired t-test. The tilde indicates ROIs in which classification accuracy (across all conditions) is significantly above chance (0.5). Right: Average classification accuracy across all the parcels with the classifiers trained and tested on each parcel independently and then averaged across all the subjects and parcels.

intervals. The results revealed a median correlation within each parcel of approximately -0.5, suggesting that the beta estimates from the anticipatory interval and the perceptual encoding interval might have significant overlap.

Univariate response. We further conducted the univariate analysis to compare the BOLD response magnitude difference between different cueing conditions (see Figure 3.7). A comparison of the the magnitude of the normalized BOLD activation level indicated the subjects' activation level is higher during perceptual encoding interval across all the AON regions than during anticipatory interval. Among the 8 ROIs, paired t-test showed that the BOLD activation was significantly higher during invalidly cued trials than uncued trials in the left IFG ($t_{28} = 2.85, p < 0.01$) and left Parietal ($t = 1.87, p < 0.1$) (Figure 3.7 right). Repeated measure t-tests showed that BOLD activation was significantly higher during validly cued trials as compared to the uncued trials in left IPL ($t = 2.82, p < 0.01$), right IPL ($t = 2.41, p < 0.05$), higher during invalidly cued trials than uncued trials in left IFG

($t = 2.52, p < 0.05$), and higher during invalidly cued trials than validly cued trials in left pSTS ($t = 2.1, p < 0.05$).

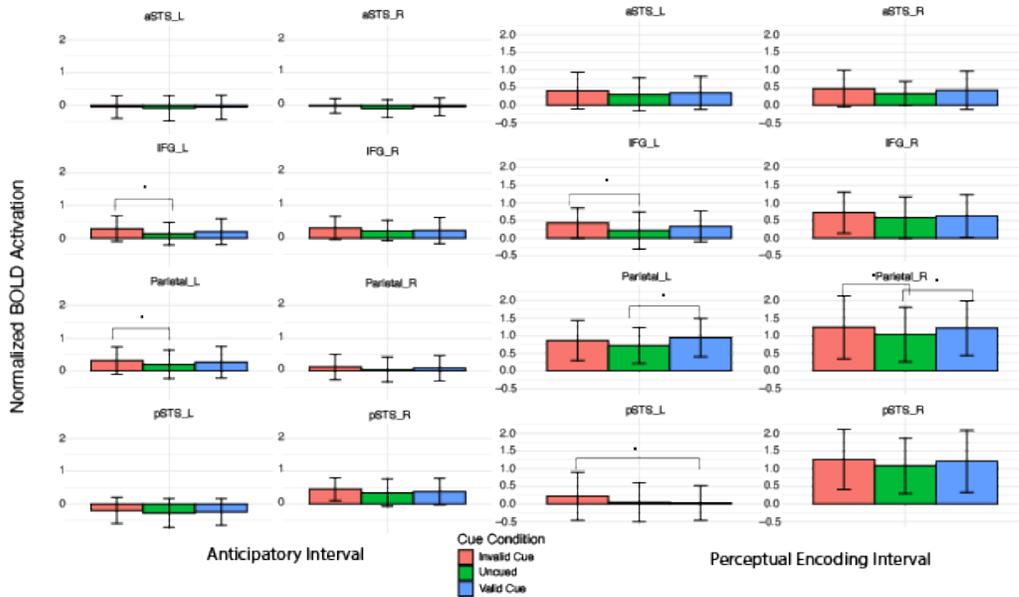


Figure 3.7: **Univariate Analysis Results.** Left: BOLD response during anticipatory interval for different cueing conditions. Right: BOLD response during perceptual encoding interval.

3.4 Discussion

We are interested in predictive coding during action observation such that the and we tried to look into the mental planning stage before sensory encoding stage. In the current study, we used a cue with 83% of validity to prompt subjects as to the upcoming action, and then asked the subjects to make a perceptual judgement of the action being depicted during a whole body movie stimulus. Results showed that in the action observation network, the cue improved classification accuracy in the left pSTS as compared to the than uncued trials during the perceptual encoding interval. Further analysis looking into individual parcels within each ROI found that during the perceptual encoding stage, majority of the regions in the right pSTS increased in classification accuracy during uncued trials as compared to the

cued trials. Only the left lateral STSm and right IFG had a marginal increase in classification accuracy when subjects were given a cue than when they weren't given a cue. These brain regions located at the left upper bank of STS, and the inferior frontal gyrus and premotor area, respectively. Moreover, the univariate results during the perceptual encoding interval showed that the BOLD activation increased during validly cued trials than uncued trials in the bilateral parietal, uncued than invalid trials in the left IFG and valid to invalid trials in the pSTS.

Our preceding studies have explored the encoding interval, focusing on how subjects' attention towards different features of stimuli influences neural activation and functional connectivity (Stehr et al., 2021; Zhou et al., 2023). Our current research aims to dissect these anticipatory and sensory encoding stages within the action observation network (AON). Our results show the cue was effective to have subjects expect the upcoming action, showing by the difference in BOLD response during cued trials than uncued trials during the perceptual encoding interval.

Our study sought evidence for action prediction with the AON, and found some potential evidence in the overall increased of neural response in the left parietal and left pSTS during cued trials than uncued trials. Also, when averaged across all the parcels and subjects, neural activation patterns were more distinguishable in the right IFG and left lateral STS in the time following the cue (the anticipatory interval).

The action prediction theory proposed observer generates a mental model of the anticipated action outcomes. This theory is supported by the pSTS's encoding variability Grossman et al. (2010). Also, studies looking at AON regions during motor action outcome representations found that IFG and Parietal predicting the likely outcomes of actions (Koul et al., 2018; Mottönen et al., 2016). Also found pSTS supporting the contextual updating and social interaction guidance (Geng and Vossel, 2013; Patel et al., 2019).

Additionally, literatures on action prediction error during action observation has shown that BOLD response increased when subjects perceptual inputs violates the expectation in pSTS (Shultz et al 2011) and IFG (de Lange et al 2008), which align with the current study results that shown increased BOLD response during invalid trials. Also enhanced neural representations could possibly explained by effective connectivity studies showed that showed that PMv feedback connection to IPL when violating biological aspect of motion (Urgen and Saygin, 2020).

One limitation of this study was lack of classifier's power, as reflected by the scores while fitting SVM, which was classifiers' distance to decision boundaries. Low classification certainty was noted when using each brain parcel as a feature with SVM scores near e^{-10} , contrasting with higher scores at e^{-1} when using individual vertices as feature in the classifiers. This suggests a potential improvement strategy of utilizing all vertices within each ROI for training and testing to enhance classifier performance and preserve neural pattern information. Another improvement could be made by analyzing the data only for the shorter anticipatory interval (2 or 3.5s) since it might become harder for subjects to maintain the cue over a long period of time. However, this also has the potential difficulty to disentangle the BOLD response between anticipatory and perceptual encoding intervals.

Chapter 4

Diffusion Model Insights into Evidence Accumulation During Action Prediction: The Role of Predictive Cues Brain Responses within the Action Observation Network

4.1 Introduction

Understanding the cognitive processes underlying action observation has been a critical endeavor in cognitive neuroscience. Previous studies, including our own, have revealed that, when cued, information about upcoming actions is present in the neural signal before the

arrival of perceptual inputs, with right IFG and left lateral STS showing weak evidence during cued trials (Koul et al., 2018). This is an example of action prediction, as proposed as generative models by Kilner and Frith (2007); Bach and Schenke (2017), such that the observers form mental models to anticipate future actions and outcomes, matching these internal predictions with incoming perceptual cues, and adjusting them as needed when faced with prediction errors. Prediction is proposed to increase the efficiency with which the anticipated actions are encoded and giving subjects access to action information before directly observing the actions (Friston et al., 2011).

Our interest in subjects' action prediction processes prior to action observation is twofold. Firstly, it acknowledges the existence of an action prediction process, as posited by mental models of action observation (Bach and Schenke, 2017) and internal model of affordance Sakreida et al. (2016), which suggests that observers collect evidence from the context and the environment before observing the action with the predicted kinematics. Secondly, it addresses the challenge of measuring these implicit processes, which are not directly observable through behavior. Thus the predictive nature of this cognitive processing is central to our investigation. To navigate this challenge, we turn to diffusion models, which provide an indirect approximation of the underlying cognition during the decision-making process.

To evaluate for evidence of predictive signals in action cueing, we used diffusion modeling to identify latent parameters reflecting components of the decision-making process (Gold and Shadlen, 2007; Ratcliff and McKoon, 2008). Diffusion model takes into account the stochastic accumulation of evidence over time during decision making leading to a decision boundary with parameters for drift rate (v), initial bias in starting point (z) and time to make a decision (t). These parameters model different processes underlying the decision making process. Hierarchical drift diffusion models (HDDM), in particular, allow us to model posterior distributions at the subject and trial level (?). This allows for the dissection of cognition on a finer scale, revealing how individual differences and trial-by-trial variations

contribute to the observed behaviors. By using HDDM, we can investigate not only the average tendency of decision-making across subjects but also the nuanced fluctuations that occur within subjects across different trials Turner et al. (2013); Nunez et al. (2017).

These models have been recently advanced to integrate neurophysiological data into cognitive modeling, as demonstrated by Turner et al. (2015), who used a hierarchical Bayesian framework to incorporate single-trial brain activity into diffusion model parameters. Past research, including studies by Nunez et al. (2017); Koul et al. (2019), has employed HDDM to demonstrate that trial-by-trial fluctuations in brain regions can predict measurements derived from diffusion models. These studies contribute to a growing understanding of how pre-stimulus brain activity can inform models of response accuracy and timing.

We hypothesize that by examining the bias parameter (z) in diffusion modeling—reflecting the starting location on the evidence accumulation scale—we can shed light on the anticipatory decision-making process. This bias parameter offers a quantifiable measure of the extent to which predictive cues influence decision-making at the onset of perceptual encoding, reflecting the amount of evidence accumulated during action observation.

With this foundation, our study aims to explore the link between predictive cues, as reflected in bias parameters from diffusion models, and neural measurements within the action observation network including brain region pSTS, IFG, IPL and extended brain region lateral STS. We anticipate that this approach will provide a novel perspective on the neural basis of action prediction and the cognitive abstractions it entails.

4.2 Methods

This study analyzes the data from the Chapter 3 study and therefore the detailed experiment design and neuroimaging methods can be found in Chapter 3. In brief, 29 participants

completed the study in which participants were cued (validly or invalidly) or not cued as to the upcoming action. During the cued trials, subjects were first given a written cue of either jump or crouch. For the valid cued trials, the cued actions were consistent with the action being depicted in the stimuli, while during the invalid cued trials, the cued actions are the opposite of the depicted actions. In each scan there were 12 uncued trials where subjects were shown a blank circle in the cue interval, and 10 validly cued trials and 2 invalidly cued trials. These 24 trials were assigned randomly, and following the cue there was a brief anticipatory interval of 2, 3.5 or 5s prior to the onset of the stimuli.

In the current study we measured subject response time for a two-alternative forced choice judgment on the observed action. Subjects were instructed to report as quickly as possible which of two actions was being shown using button press. The response time was the time from the onset of the stimuli until the time subjects made a button response. Responses were labeled as incorrect if the participant either mislabeled the action or did not make a response from the movie onset until after the additional 3s response time window.

Response time analysis. In order to evaluate whether subject used the cue to modify their decision process, we first modeled the response time distributions for the different cueing condition using a Gaussian distribution to model the response time and conducted the statistical test using `gamlss` R package to test if the mean response time was different during validly cued versus uncued trials after fitting the response time with a Gaussian distribution.

Hierarchical diffusion modeling (HDDM). In the current study we used the HDDM package implemented in Python to analyze the response time data. According to the previous literature of parameters about prior information about action and goals (?), we structured the HDDM to include both the bias (z) and non-decision time (t) parameters, which are critical for understanding the starting point of the evidence accumulation and the time consumed by processes other than decision-making, respectively. We also allowed the z initial bias to varied

between different cueing conditions (validly cued, invalidly cued, uncued), providing insights into how different types of cues could influence the starting point of evidence accumulation. We also allowed the model to account for the random effects across subjects and therefore estimate the parameters individually for each subject.

Finally, in order to investigate the underlying brain regions that support the hidden cognition examined by the HDDM model, we used the posterior samples of bias in starting point ($n=2000$, with the first 100 burn to initialize the process), generated with Gibbs sampling and differential evolution with MCMC (Markov Chain Monte Carlo simulation), varied by cueing condition to correlate with various brain measurements including the scores from MVPA classifications from above. To formally test the hypothesis that whether the significant difference in the bias parameters z under different cueing conditions can be linked to the neural measurements, we used linear mixed effect modeling from `lme4` in R. The model was constructed as follows:

```
z estimates = Cue condition + average neural measurements for each subjects +  
cue*neural measurements + random effect of Subjects
```

In this model, we are interested in the interaction fixed effect `cue*neural measurements` after accounting for the individual subject differences in the baseline of bias z .

4.3 Results

We found that the mean reaction time for the uncued trials was significantly longer than the mean reaction time during cued trials ($t = 3.75$, $p < 0.001$). Validly cued trials also had shorter response time than invalidly cue trials ($t = 5.65$, $p < 0.001$). There was no difference in reaction times between uncued and invalid trials ($t = 0.7$, $p = 0.5$).

HDDM modeling. Figure 4.1 shows the posterior distribution of the initial bias (z), fitted using a HDDM for response times in the valid cue, invalid cue, and uncued conditions across all the subjects and trials. Results showed that that cued trials have larger initial bias toward the correct response and lowest for invalid trials (Figure 4.1 left). The posterior test indicated significance such that the probability of the bias in starting point was different between validly cued and uncued trials, validly cued trials and invalidly cued trials were greater than 99%.

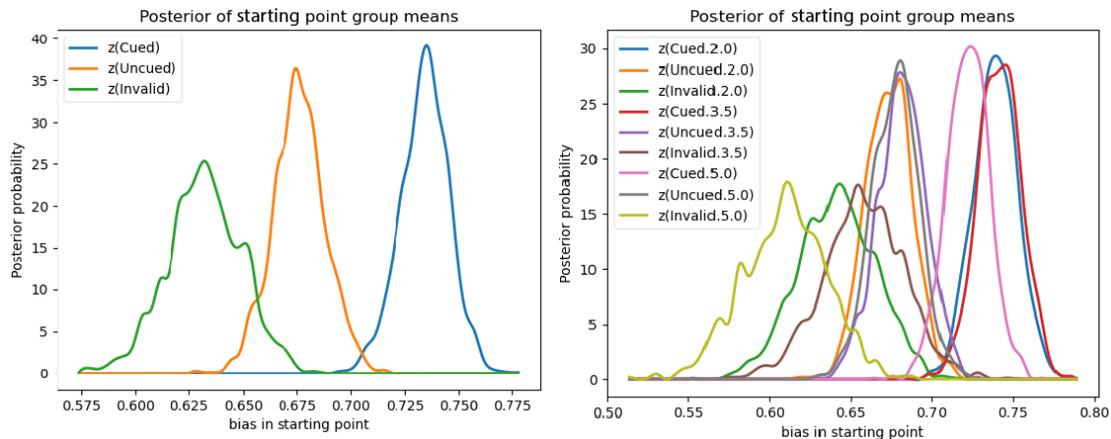


Figure 4.1: **Hierarchical Diffusion Modeling Results.** Left: The posterior distribution of bias in starting point according to each cueing condition. Right: Posterior distributions of bias in starting point according to each cueing condition and also the length of the anticipatory interval.

In a second analysis, we evaluated the effect of the duration of the anticipatory delay (2, 3.5 or 5 sec) on the initial bias. We hypothesized that the longer anticipatory interval might challenge the subjects sustained working memory when remembering the cue, therefore resulting in a lower bias. Alternatively, however, a longer delay could instead give the participants the opportunity to strengthen the cue effect, perhaps by allowing more samples to be drawn from memory during the anticipatory interval (Bornstein Hanks, xx).

In Figure 4.1, the posterior estimates of the bias in the starting point (z) revealed that when the anticipatory interval was extended to 5 seconds, there was a notable reduction in bias for the cued trials with shorter anticipatory intervals (illustrated in pink, blue, and pinkish-

purple). However, the length of the anticipatory interval did not impact the initial starting point in uncued trials. There was a trend for weaker bias for invalidly cued trials, although note that there were fewer of these trial types and thus the posterior distribution has more variance. Together these things are consistent with the hypothesis of decay of the cue during longer anticipatory intervals. There was some indication of

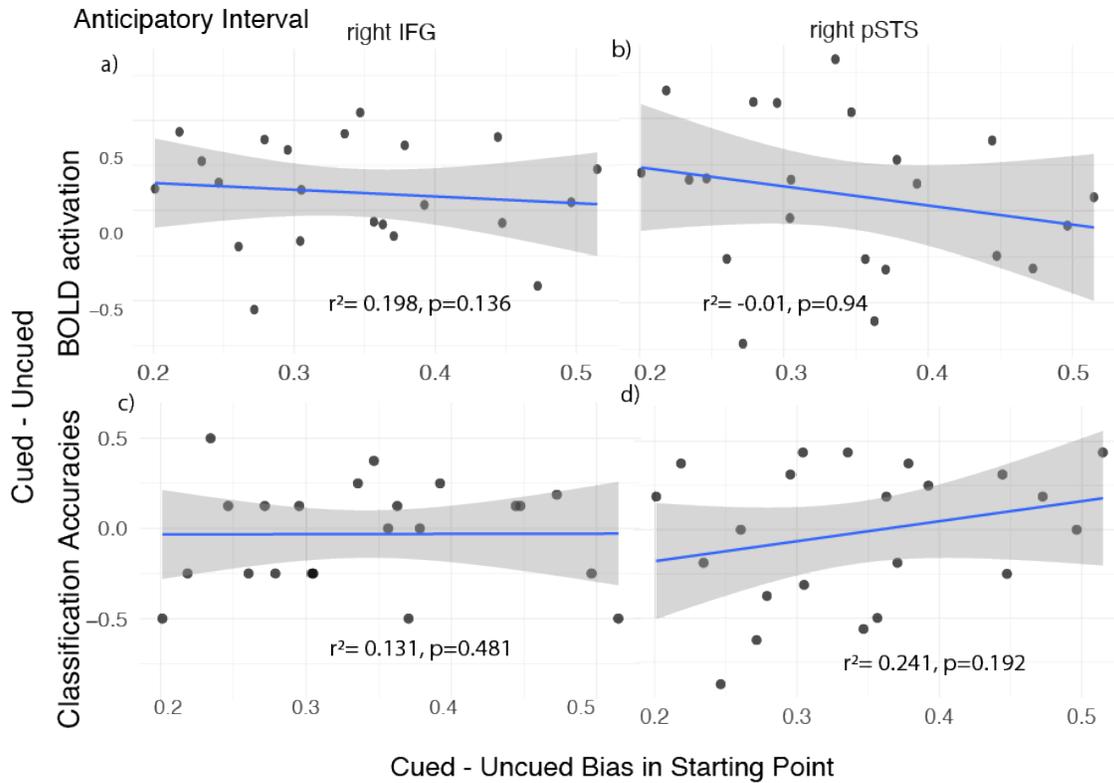


Figure 4.2: **Hierarchical Diffusion Modeling with Neural Measurements.** a) and b) show the distribution plot of difference (Cued - Uncued) in BOLD activation during the anticipatory interval in right IFG and right pSTS against the difference between posterior estimates of bias in starting point fitted from the HDDM models for each subjects. c) and d) show the distribution plot of difference (Cued - Uncued) in MVPA classification accuracies when classifying the action cues during the anticipatory interval in right IFG and right pSTS against the difference between posterior estimates of bias in starting point fitted from the HDDM models for each subjects.

Initial bias and neural metrics. Lastly, we examined whether the neural measurements in the right pSTS and right IFG correlated with the behavioral indicators of bias. We compared the BOLD activation and also the classification accuracy for classifying the action cues given to the subjects with the bias in starting point.

We found no significant correlation in the right IFG and pSTS between the cue's effect on bias (as indicated by the starting point derived from the diffusion model for the difference between cued vs uncued trials in individual subjects) and the effect of the cue on classification accuracy or BOLD response (Figure 4.2). Although there is a plausible trend towards a positive relationship in the difference of right IFG BOLD response during the anticipatory interval ($r^2 = 0.198, p = 0.136$) to the difference in the bias in starting point. Likewise, when evaluating the classification accuracies, we found a positive relationship in the anticipatory interval between the effect of the cue on behavioral bias and classification accuracies in the right pSTS ($r^2 = 0.241, p = 0.192$). These results are suggestive that the cue *may* modulate the initial neural state in the right pSTS and IFG during the anticipatory interval.

4.4 Discussion

In our hypothesis, we proposed that giving individuals a cue as to the upcoming action would serve to facilitate later encoding of that action. We hypothesized that would be apparent in the bias parameter (z), which would serve as a quantitative indicative of the cue's influence on each individual's decision-making. Therefore, the cueing paradigm was designed to intentionally shift subjects' biases in the upcoming action.

Results indicated that subjects were more biased towards the correct answer when the cue was valid, markedly more so than in both uncued and invalidly cued trials, with initial biases being significantly greater in validly cued trials over 99% of the time. Furthermore, upon examining brain responses in right IFG and right pSTS, which based on the previous chapters were proposed to support top-down influence, we found that during the anticipatory interval, a larger BOLD response magnitude corresponded to a higher bias in the cued condition within the right IFG and higher classification accuracies in the pSTS.

The results from this study confirm that prior information can effectively shift subjects' biases during action observation, demonstrating a form of action prediction when viewing whole-body action. The variance in the length of the anticipatory interval further substantiated the cueing effect. The length of the anticipatory interval only altered the bias when subjects were given a cue, with no change in bias for uncued trials. Our findings also suggest that the longer anticipatory intervals might challenge subjects' retention of the cue information, potentially due to the higher cognitive load or energy demands as per the theory of working memory decay.

In the current study, the cue served as an approximation for the action prediction the observer formed when witnessing an action before making a decision on the action's outcome. Subjects were shown a visual word cue to anticipate a whole body action movie stimulus depicting an avatar either jumping up or crouching down. The cue, therefore, provided a manipulable perceptual expectation of the forthcoming kinematics, with 83% of valid cues aiding subjects in making a more accurate initial decision.

Regarding brain measurements, the BOLD magnitude in right IFG during the anticipatory interval increased in those subjects with a larger bias towards the correct answer across subjects when the subjects given a cue. Also, the classification accuracies of action cues in the right pSTS during the anticipatory interval increased when subjects' initial bias in starting point were higher. Together the increased magnitude in BOLD response was suggestive that the cue prepared the neural population for the upcoming action. This preparatory effect in sensory neural populations is consistent with those found in V1 using a cued orientation discrimination task (Kok et al., 2012, 2017).

Our results are consistent with previous findings of effective connectivity in the AON. Those earlier studies demonstrated that when observing biological motion, feedback connectivity between the IFG and IPL is strengthened when predictions are violated (Urgen and Saygin, 2020), that connectivity between pSTS and IFG is strengthened when observing actions in

visually degraded conditions that would benefit from prior information (Sokolov et al., 2018). These studies suggest a top-down feedback information flow between the AON brain regions when the top down expectations were violated. In our current study, the cue increased the bias towards and thereby could be explained by within the AON, higher-level regions transferred prediction information to lower-level regions.

It's worth noting that the analyses relating brain measurements to diffusion modeling bias parameters were limited in that the correlation were not strong. Therefore, future analyses should test hypotheses at a trial-by-trial level proposed by (Nunez et al., 2017). Improvements can be further made by employing neural measurements as prior information to enhance the drift diffusion models (Turner et al., 2013; Nunez et al., 2017). Additional brain measurements, particularly functional connectivity strength between different brain regions during cued trials, may show how strengthened connectivity correlates with increases in bias at the starting point. However, the analyses in the current study still provide a hint of the plausible existence of action prediction in the brain using HDDM and neuroimaging, and future analyses should investigate this relationship more closely.

In summary, we have demonstrated that subjects' behaviors vary according to cueing conditions and that these differences are attributable to the initial bias subjects hold about the forthcoming action. The action prediction evidence accumulated in subjects, with cues effectively aiding in the identification of the upcoming action. This effect could probably relate to the top-down influence between the right IFG and the right pSTS suggesting neurons coding expectations within the action observation network.

Chapter 5

Conclusion

These studies collectively demonstrated how top-down influences from the IFG enhance the pSTS's reconstruction of action dynamics. These results are also consistent with the proposal that predictive cues can bias both behavioral responses and neural activity, providing a deeper understanding of the neural pathways involved in interpreting socially relevant actions.

Bibliography

- Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: Role of the sts region. *Trends in Cognitive Sciences* 4, 267–278. doi:10.1016/S1364-6613(00)01501-1.
- Amoruso, L., Finisguerra, A., Urgesi, C., 2018. Contextualizing action observation in the predictive brain: Causal contributions of prefrontal and middle temporal areas. *NeuroImage* 177, 68–78. URL: <https://doi.org/10.1016/j.neuroimage.2018.05.020>, doi:10.1016/j.neuroimage.2018.05.020.
- Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20, 870–888. URL: <https://www.sciencedirect.com/science/article/pii/S1053811903003367>, doi:10.1016/S1053-8119(03)00336-7.
- Ansuini, C., Giosa, L., Luca, , Gianmarco, T., , A., Castiello, U., 2008. An object for an action, the same object for other actions: effects on hand shaping. *Exp Brain Res* 185, 111–119. doi:10.1007/s00221-007-1136-4.
- Ansuini, C., Santello, M., Massaccesi, S., Castiello, U., 2006. Effects of end-goal on hand shaping. *Journal of Neurophysiology* 95, 2456–2465. doi:10.1152/jn.01107.2005.
- Ariani, G., Wurm, M.F., Lingnau, A., 2015. Decoding internally and externally driven movement plans. *Journal of Neuroscience* 35, 14160–14171. URL: <https://www.jneurosci.org/content/35/42/14160><https://www.jneurosci.org/content/35/42/14160.abstract>, doi:10.1523/JNEUROSCI.0596-15.2015.
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* 12, 26–41. doi:10.1016/J.MEDIA.2007.06.004.
- Avenanti, A., Paracampo, R., Annella, L., Tidoni, E., Aglioti, S.M., 2017. Boosting and decreasing action prediction abilities through excitatory and inhibitory tdcS of inferior frontal cortex. *Cerebral Cortex* , 1–15 URL: <https://academic.oup.com/cercor/article/3052754/Boosting>, doi:10.1093/cercor/bhx041.
- Bach, P., Nicholson, T., Hudsons, M., 2014. The affordance-matching hypothesis: How objects guide action understanding and prediction. *Frontiers in Human Neuroscience* 8, 1–13. doi:10.3389/fnhum.2014.00254.

- Bach, P., Schenke, K.C., 2017. Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass* 11, 1–17. doi:10.1111/spc3.12312.
- Bach, P., Tipper, S.P., 2007. Implicit action encoding influences personal-trait judgments. *Cognition* 102, 151–178. URL: <http://dx.doi.org/10.1016/j.cognition.2005.11.003>, doi:10.1016/j.cognition.2005.11.003.
- Badets, A., Osiurak, F., 2015. A goal-based mechanism for delayed motor intention: considerations from motor skills, tool use and action memory. *Psychological Research* 79, 345–360. doi:10.1007/s00426-014-0581-5.
- Baker, C.L., Tenenbaum, J.B., Saxe, R.R., 2005. Bayesian models of human action understanding. *Advances in Neural Information Processing Systems* , 99–106.
- Baldassano, C., Chen, J., Zadbood, A., Pillow, J.W., Hasson, U., Norman, K.A., 2017. Discovering event structure in continuous narrative perception and memory. *Neuron* 95, 709–721.e5. URL: <http://dx.doi.org/10.1016/j.neuron.2017.06.041>, doi:10.1016/j.neuron.2017.06.041.
- Basil, R.A., Westwater, M.L., Wiener, M., Thompson, J.C., 2017. A causal role of the right superior temporal sulcus in emotion recognition from biological motion. *Open Mind* 2, 26–36. doi:10.1162/opmi_a_00015.
- Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., Saxe, R., 2008. Concepts are more than percepts: The case of action verbs. *Journal of Neuroscience* 28, 11347–11353. URL: <http://www.fil.ion.>, doi:10.1523/JNEUROSCI.3039-08.2008.
- Behzadi, Y., Restom, K., Liao, J., Liu, T.T., 2007. A component based noise correction method (compcor) for bold and perfusion based fmri. *NeuroImage* 37, 90–101. URL: <http://www.sciencedirect.com/science/article/pii/S1053811907003837>, doi:10.1016/j.neuroimage.2007.04.042.
- Binder, J.R., Desai, R.H., 2011. The neurobiology of semantic memory. *Trends in Cognitive Sciences* 15, 527–536. URL: <http://dx.doi.org/10.1016/j.tics.2011.10.001>, doi:10.1016/j.tics.2011.10.001.
- Binkofski, F., Buxbaum, L.J., 2013. Two action systems in the human brain. *Brain and Language* 127, 222–229. URL: <http://dx.doi.org/10.1016/j.bandl.2012.07.007>, doi:10.1016/j.bandl.2012.07.007.
- Borghi, A.M., Riggio, L., 2009. Sentence comprehension and simulation of object temporary, canonical and stable affordances. *Brain Research* 1253, 117–128. URL: <http://dx.doi.org/10.1016/j.brainres.2008.11.064>, doi:10.1016/j.brainres.2008.11.064.
- Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences* 14, 277–290. URL: <http://dx.doi.org/10.1016/j.tics.2010.04.004>, doi:10.1016/j.tics.2010.04.004.

- Buxbaum, L., Kalénine, S., 2010. the two action systems. *Ann N Y Acad Sci* 1191, 201–218. doi:10.1111/j.1749-6632.2010.05447.x.Action.
- de C Hamilton, A.F., Grafton, S.T., 2008. Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex* 18, 1160–1168. doi:10.1093/cercor/bhm150.
- Caspers, S., Zilles, K., Laird, A.R., Eickhoff, S.B., 2010. A meta-analysis of action observation and imitation in the human brain. *NeuroImage* 50, 1148–1167. URL: <http://dx.doi.org/10.1016/j.neuroimage.2009.12.112>, doi:10.1016/j.neuroimage.2009.12.112.
- Chen, Q., Garcea, F.E., Mahon, B.Z., 2016. The representation of object-directed action and function knowledge in the human brain. *Cerebral Cortex* 26, 1609–1618. doi:10.1093/cercor/bhu328.
- Cisler, J.M., Bush, K., Steele, J.S., 2014. A comparison of statistical methods for detecting context-modulated functional connectivity in fmri. *NeuroImage* 84, 1042–1052. URL: <http://dx.doi.org/10.1016/j.neuroimage.2013.09.018>, doi:10.1016/j.neuroimage.2013.09.018.
- Cole, M.W., Reynolds, J.R., Power, J.D., Repovs, G., Anticevic, A., Braver, T.S., 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience* 16, 1348–1355. doi:10.1038/nn.3470.
- Cook, R., Bird, G., Catmur, C., Press, C., Heyes, C., 2014. Mirror neurons: From origin to function. *Behavioral and Brain Sciences* 37, 177–192. doi:10.1017/S0140525X13000903.
- Costantini, M., Galati, G., Ferretti, A., Caulo, M., Tartaro, A., Romani, G.L., Aglioti, S.M., 2005. Neural systems underlying observation of humanly impossible movements: An fmri study. *Cerebral Cortex* 15, 1761–1767. doi:10.1093/cercor/bhi053.
- Cox, R.W., Hyde, J.S., 1997. Software tools for analysis and visualization of fmri data. *NMR in Biomedicine* 10, 171–178. doi:10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L.
- Csibra, G., 2008. Action mirroring and action interpretation: An alternative account. *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII*, 427–435–459 URL: papers2://publication/uuid/F1F14887-90E5-4F6E-A5D4-AE0007A7FA5E.
- Csibra, G., Gergely, G., 2007. 'obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica* 124, 60–78. doi:10.1016/j.actpsy.2006.09.007.
- Cuijpers, R.H., Smeets, J.B., Brenner, E., 2004. On the relation between object shape and grasping kinematics. *Journal of Neurophysiology* 91, 2598–2606. doi:10.1152/jn.00644.2003.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage* 9, 179–194. URL: <http://www.sciencedirect.com/science/article/pii/S1053811998903950>, doi:10.1006/ning.1998.0395.

- Deen, B., Saxe, R., 2012. Neural correlates of social perception: The posterior superior temporal sulcus is modulated by action rationality, but not animacy. *Proceedings of the 33rd Annual Cognitive ...*
- Derderian, K.D., Zhou, X., Chen, L., 2021. Category-specific activations depend on imaging mode, task demand, and stimuli modality: An ale meta-analysis. *Neuropsychologia* 161, 108002. URL: <https://doi.org/10.1016/j.neuropsychologia.2021.108002>, doi:10.1016/j.neuropsychologia.2021.108002.
- Eastough, D., Edwards, M.G., 2007. Movement kinematics in prehension are affected by grasping objects of different mass. *Experimental Brain Research* 176, 193–198. doi:10.1007/s00221-006-0749-3.
- Esteban, O., Blair, R., Markiewicz, C.J., Berleant, S.L., Moodie, C., Ma, F., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Sitek, K.R., Gomez, D.E.P., Lurie, D.J., Ye, Z., Poldrack, R.A., Gorgolewski, K.J., 2018. fmriprep. Software doi:10.5281/zenodo.852659.
- Ferrari, P.F., Rozzi, S., Fogassi, L., 2005. Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *Journal of Cognitive Neuroscience* 17, 212–226. doi:10.1162/0898929053124910.
- Fischer, J., Mikhael, J.G., Tenenbaum, J.B., Kanwisher, N., 2016. Functional neuroanatomy of intuitive physical inference. *Proceedings of the National Academy of Sciences of the United States of America* 113, E5072–E5081. doi:10.1073/pnas.1610344113.
- Fleischer, F., Christensen, A., Caggiano, V., Thier, P., Giese, M.A., 2012. Neural theory for the perception of causal actions. *Psychological Research* 76, 476–493. doi:10.1007/s00426-012-0437-9.
- Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G., 2005. Parietal lobe: From action organization to intention understanding leonardo. *Science* 308, 659–662. URL: www.sciencemag.org/cgi/content/full/308/5722/659/, doi:10.1126/science.1111199.
- Fonov, V.S., Evans, A.C., McKinstry, R.C., Almlí, C.R., Collins, D.L., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* 47, Supplement 1, S102. doi:10.1016/S1053-8119(09)70884-5.
- Friston, K., Mattout, J., Kilner, J., 2011. Action understanding and active inference. *Biological Cybernetics* 104, 137–160. doi:10.1007/s00422-011-0424-z.
- Frost, M.A., Goebel, R., 2012. Measuring structural-functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage* 59, 1369–1381. URL: <https://pubmed.ncbi.nlm.nih.gov/21875671/>, doi:10.1016/j.neuroimage.2011.08.035.
- Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. *Brain* 119, 593–609. doi:10.1093/brain/119.2.593.

- Gallese, V., Goldman, A., 1998. Mirror neurons and the theory of mind rreading. *Trends in Cognitive Sciences* 2, 493–501.
- Geng, J.J., Vossel, S., 2013. Re-evaluating the role of tpj in attentional control: Contextual updating? *Neuroscience and Biobehavioral Reviews* 37, 2608–2620. URL: <http://dx.doi.org/10.1016/j.neubiorev.2013.08.010>, doi:10.1016/j.neubiorev.2013.08.010.
- Gergely, G., Csibra, G., 2003. Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences* 7, 287–292. doi:10.1016/S1364-6613(03)00128-1.
- Gibson, J.J., 1979. *The Ecological Approach to Visual Perception*. Boston, MA, US. doi:10.4324/9781315740218.
- Giese, M.A., Poggio, T., D-Tübingen, 2003. Neural mechanisms for the recognition of biological movements. *Nature Reviews* 4. doi:10.1038/nrn1057.
- Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Essen, D.C.V., 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178. doi:10.1038/nature18933.
- Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. *Annual review of neuroscience* 30, 535–574. URL: <https://pubmed.ncbi.nlm.nih.gov/17600525/>, doi:10.1146/ANNUREV.NEURO.29.051605.113038.
- Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics* 5, 13. doi:10.3389/fninf.2011.00013.
- Gorgolewski, K.J., Esteban, O., Markiewicz, C.J., Ziegler, E., Ellis, D.G., Notter, M.P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., Modat, M., Dewey, B.E., Madison, C., di Oleggio Castello, M., Clark, M.G., Dayan, M., Clark, D., Keshavan, A., Pinsard, B., Gramfort, A., Berleant, S., Nielson, D.M., Bougacha, S., Varoquaux, G., Cipollini, B., Markello, R., Rokem, A., Moloney, B., Halchenko, Y.O., Demian, W., Hanke, M., Horea, C., Kaczmarzyk, J., de Hollander, G., DuPre, E., Gillman, A., Mordom, D., Buchanan, C., Tungaraza, R., Pauli, W.M., Iqbal, S., Sikka, S., Mancini, M., Schwartz, Y., Malone, I.B., Dubois, M., Frohlich, C., Welch, D., Forbes, J., Kent, J., Watanabe, A., Cumba, C., Huntenburg, J.M., Kastman, E., Nichols, B.N., Eshaghi, A., Ginsburg, D., Schaefer, A., Acland, B., Giavasis, S., Kleesiek, J., Erickson, D., Küttner, R., Haselgrove, C., Correa, C., Ghayoor, A., Liem, F., Millman, J., Haehn, D., Lai, J., Zhou, D., Blair, R., Glatard, T., Renfro, M., Liu, S., Kahn, A.E., Pérez-García, F., Triplett, W., Lampe, L., Stadler, J., Kong, X.Z., Hallquist, M., Chetverikov, A., Salvatore, J., Park, A., Poldrack, R., Craddock, R.C., Inati, S., Hinds, O., Cooper, G., Perkins, L.N., Marina, A., Mattfeld, A., Noel, M., Snoek, L., Matsubara, K., Cheung, B., Rothmei, S., Urchs,

- S., Durnez, J., Mertz, F., Geisler, D., Floren, A., Gerhard, S., Sharp, P., Molina-Romero, M., Weinstein, A., Broderick, W., Saase, V., Andberg, S.K., Harms, R., Schlamp, K., Arias, J., Orfanos, D.P., Tarbert, C., Tambini, A., Vega, A.D.L., Nickson, T., Brett, M., Falkiewicz, M., Podranski, K., Linkersdörfer, J., Flandin, G., Ort, E., Shachnev, D., McNamee, D., Davison, A., Varada, J., Schwabacher, I., Pellman, J., Perez-Guevara, M., Khanuja, R., Pannetier, N., McDermottroe, C., Ghosh, S., 2018. Nipype. Software doi:10.5281/zenodo.596855.
- Grafton, S.T., Tipper, C.M., 2012. Decoding intention: A neuroergonomic perspective. *NeuroImage* 59, 14–24. URL: <http://dx.doi.org/10.1016/j.neuroimage.2011.05.064>, doi:10.1016/j.neuroimage.2011.05.064.
- Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* 48, 63–72. doi:10.1016/j.neuroimage.2009.06.060.
- Grosbras, M.H., Beaton, S., Eickhoff, S.B., 2012. Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human Brain Mapping* 33, 431–454. doi:10.1002/hbm.21222.
- Grossman, E.D., Jardine, N.L., Pyles, J.A., 2010. fmr-adaptation reveals invariant coding of biological motion on the human sts. *Frontiers in Human Neuroscience* 4. URL: www.frontiersin.org, doi:10.3389/neuro.09.015.2010.
- Hafri, A., Trueswell, J.C., Epstein, R.A., 2017. Neural representations of observed actions generalize across static and dynamic visual input. *Journal of Neuroscience* 37, 3056–3071. doi:10.1523/JNEUROSCI.2496-16.2017.
- Hamilton, A.F.C., Grafton, S.T., 2006. Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience* 26, 1133–1137. doi:10.1523/JNEUROSCI.4551-05.2006.
- Hartwigsen, G., Neef, N.E., Camilleri, J.A., Margulies, D.S., Eickhoff, S.B., 2018. Functional segregation of the right inferior frontal gyrus: Evidence from coactivation-based parcellation. *Cerebral Cortex* , 1–15 URL: <https://academic.oup.com/cercor/advance-article/doi/10.1093/cercor/bhy049/4975477>, doi:10.1093/cercor/bhy049.
- Hasson, U., Yang, E., Vallines, I., Heeger, D.J., Rubin, N., 2008. A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience* 28, 2539–2550. URL: www.jneurosci.org, doi:10.1523/JNEUROSCI.5487-07.2008.
- Huang, C.T., Charman, T., 2005. Gradations of emulation learning in infants' imitation of actions on objects. *Journal of Experimental Child Psychology* 92, 276–302. doi:10.1016/j.jecp.2005.06.003.
- Hudson, M., Bach, P., Nicholson, T., 2018. You said you would! the predictability of other's behavior from their intentions determines predictive biases in action perception. *Journal of experimental psychology. Human perception and performance* 44, 320–335. URL: <https://pubmed.ncbi.nlm.nih.gov/28557497/>, doi:10.1037/XHP0000451.

- Iacoboni, M., 2005. Understanding others: imitation, language, empathy. *Perspectives on Imitation: From Cognitive Neuroscience to Social Science* .
- Iacoboni, M., Koski, L.M., Brass, M., Bekkering, H., Woods, R.P., Dubeau, M.C., Mazziotta, J.C., Rizzolatti, G., 2001. Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 98, 13995–13999. URL: www.pnas.org/cgi/doi/10.1073/pnas.241474598, doi:10.1073/pnas.241474598.
- Isik, L., Koldewyn, K., Beeler, D., Kanwisher, N., 2017. Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America* 114, E9145–E9152. doi:10.1073/pnas.17144711114.
- Jacob, P., Jeannerod, M., 2005. The motor theory of social cognition: A critique. *Trends in Cognitive Sciences* 9, 21–25. doi:10.1016/j.tics.2004.11.003.
- Jacquet, P.O., Roy, A.C., Chambon, V., Borghi, A.M., Salemme, R., Farnè, A., Reilly, K.T., 2016. Changing ideas about others' intentions: updating prior expectations tunes activity in the human motor system. *Scientific Reports* 6, 1–15. URL: <http://dx.doi.org/10.1038/srep26995>, doi:10.1038/srep26995.
- Jastorff, J., Clavagnier, S., Gergely, G., Orban, G.A., 2011. Neural mechanisms of understanding rational actions: Middle temporal gyrus activation by contextual violation. *Cerebral Cortex* 21, 318–329. doi:10.1093/cercor/bhq098.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17, 825–841. URL: <http://www.sciencedirect.com/science/article/pii/S1053811902911328>, doi:10.1006/nimg.2002.1132.
- Johnson, S., Slaughter, V., Carey, S., 1998. Whose gaze will infants follow? the elicitation of gaze following in 12-month-olds. *Developmental Science* 1, 233–238. doi:10.1111/1467-7687.00036.
- Jääskeläinen, I.P., Sams, M., Glerean, E., Ahveninen, J., 2021. Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage* 224. URL: <https://pubmed.ncbi.nlm.nih.gov/33059053/>, doi:10.1016/J.NEUROIMAGE.2020.117445.
- Kable, J.W., Chatterjee, A., 2006. Specificity of action representations in the lateral occipitotemporal cortex. *Journal of Cognitive Neuroscience* 18, 1498–1517. doi:10.1162/jocn.2006.18.9.1498.
- Kelley, H.H., 1967. Attribution theory in social psychology. *ResearchGate* 14, 192–238. URL: https://psycnet.apa.org/record/1968-13540-001https://www.researchgate.net/publication/235361448_Attribution_Theory_in_Social_Psychology.
- Kemmerer, D., 2021. What modulates the mirror neuron system during action observation?: Multiple factors involving the action, the actor, the observer, the relationship between actor and observer, and the context. *Progress in Neurobiology* 205,

102128. URL: <https://doi.org/10.1016/j.pneurobio.2021.102128>, doi:10.1016/j.pneurobio.2021.102128.
- Kilner, J.M., 2011. More than one pathway to action understanding. *Trends in Cognitive Sciences* 15, 352–357. doi:10.1016/j.tics.2011.06.005.
- Kilner, J.M., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. *Cognitive Processing* 8, 159–166. doi:10.1007/s10339-007-0170-2.Predictive.
- Klein, A., Ghosh, S.S., Bao, F.S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E.C., Keshavan, A., 2017. Mindboggling morphometry of human brains. *PLOS Computational Biology* 13, e1005350. URL: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005350>, doi:10.1371/journal.pcbi.1005350.
- Kok, P., Jehee, J.F., de Lange, F.P., 2012. Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron* 75, 265–270. doi:10.1016/j.neuron.2012.04.034.
- Kok, P., Mostert, P., Lange, F.P.D., 2017. Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America* 114, 10473–10478. URL: <https://pubmed.ncbi.nlm.nih.gov/28900010/>, doi:10.1073/PNAS.1705652114.
- Kong, R., Yang, Q., Gordon, E., Xue, A., Yan, X., Orban, C., Zuo, X.N., Spreng, N., Ge, T., Holmes, A., Eickhoff, S., Yeo, B.T., 2021. Individual-specific areal-level parcellations improve functional connectivity prediction of behavior. *Cerebral Cortex* 31, 4477–4500. doi:10.1093/cercor/bhab101.
- Koster-Hale, J., Saxe, R., 2013. Theory of mind: A neural prediction problem. *Neuron* 79, 836–848. URL: <http://dx.doi.org/10.1016/j.neuron.2013.08.020>, doi:10.1016/j.neuron.2013.08.020.
- Koul, A., Cavallo, A., Cauda, F., Costa, T., Diano, M., Pontil, M., Becchio, C., 2018. Action observation areas represent intentions from subtle kinematic features. *Cerebral Cortex* 28, 2647–2654. doi:10.1093/cercor/bhy098.
- Koul, A., Soriano, M., Tversky, B., Becchio, C., Cavallo, A., 2019. The kinematics that you do not expect: Integrating prior information and kinematics to understand intentions. *Cognition* 182, 213–219. doi:10.1016/J.COGNITION.2018.10.006.
- Lahnakoski, J.M., Glerean, E., Salmi, J., Jääskeläinen, I.P., Sams, M., Hari, R., Nummenmaa, L., 2012. Naturalistic fmri mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Frontiers in Human Neuroscience* 6, 1–14. doi:10.3389/fnhum.2012.00233.
- Lanczos, C., 1964. Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis* 1, 76–85. URL: <http://epubs.siam.org/doi/10.1137/0701007>, doi:10.1137/0701007.

- de Lange, F.P., Spronk, M., Willems, R.M., Toni, I., Bekkering, H., 2008. Complementary systems for understanding action intentions. *Current Biology* 18, 454–457. doi:10.1016/j.cub.2008.02.057.
- Lee, S.M., McCarthy, G., 2016. Functional heterogeneity and convergence in the right temporoparietal junction. *Cerebral Cortex* 26, 1108–1116. doi:10.1093/cercor/bhu292.
- Liepelt, R., Cramon, D.Y.V., Brass, M., 2008. How do we infer others' goals from non-stereotypic actions? the outcome of context-sensitive inferential processing in right inferior parietal and posterior temporal cortex. *NeuroImage* 43, 784–792. URL: <http://dx.doi.org/10.1016/j.neuroimage.2008.08.007>, doi:10.1016/j.neuroimage.2008.08.007.
- Ma, F., Xu, J., Li, X., Wang, P., Wang, B., Liu, B., 2018. Investigating the neural basis of basic human movement perception using multi-voxel pattern analysis. *Experimental Brain Research* 236, 907–918. URL: <http://dx.doi.org/10.1007/s00221-018-5175-9>, doi:10.1007/s00221-018-5175-9.
- Maffei, V., Indovina, I., Macaluso, E., Ivanenko, Y.P., Orban, G.A., Lacquaniti, F., 2015. Visual gravity cues in the interpretation of biological movements: Neural correlates in humans. *NeuroImage* 104, 221–230. URL: <http://dx.doi.org/10.1016/j.neuroimage.2014.10.006>, doi:10.1016/j.neuroimage.2014.10.006.
- Marsh, L.E., Mullett, T.L., Ropar, D., Hamilton, A.F.C., 2014. Responses to irrational actions in action observation and mentalising networks of the human brain. *NeuroImage* 103, 81–90. URL: <http://dx.doi.org/10.1016/j.neuroimage.2014.09.020>, doi:10.1016/j.neuroimage.2014.09.020.
- Masson, H.L., Isik, L., 2021. Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. *NeuroImage* 245, 118741. URL: <https://doi.org/10.1016/j.neuroimage.2021.118741>, doi:10.1016/j.neuroimage.2021.118741.
- McDonough, K.L., Costantini, M., Hudson, M., Ward, E., Bach, P., 2020. Affordance matching predictively shapes the perceptual representation of others' ongoing actions. *Journal of Experimental Psychology. Human Perception and Performance* 46, 847. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7391862/>, doi:10.1037/XHP0000745.
- Molenberghs, P., Cunnington, R., Mattingley, J.B., 2012. Brain regions with mirror properties: A meta-analysis of 125 human fmri studies. doi:10.1016/j.neubiorev.2011.07.004.
- Monaco, S., Malfatti, G., Zendron, A., Pellencin, E., Turella, L., 2019. Predictive coding of action intentions in dorsal and ventral visual stream is based on visual anticipations, memory-based information and motor preparation. *Brain Structure and Function* 224, 3291–3308. URL: <https://doi.org/10.1007/s00429-019-01970-1>, doi:10.1007/s00429-019-01970-1.

- Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage* 59, 2636–2643. URL: <http://dx.doi.org/10.1016/j.neuroimage.2011.08.076>, doi:10.1016/j.neuroimage.2011.08.076.
- Möttönen, R., Farmer, H., Watkins, K.E., 2016. Neural basis of understanding communicative actions: Changes associated with knowing the actor’s intention and the meanings of the actions. *Neuropsychologia* 81, 230–237. doi:10.1016/j.neuropsychologia.2016.01.002.
- Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G., Orban, G.A., 2011. Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience* 31, 3743–3756. URL: <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.4803-10.2011>, doi:10.1523/JNEUROSCI.4803-10.2011.
- Nummenmaa, L., Calder, A.J., 2009. Neural mechanisms of social attention. *Trends in Cognitive Sciences* 13, 135–143. doi:10.1016/j.tics.2008.12.006.
- Numssen, O., Bzdok, D., Hartwigsen, G., 2021. Hemispheric specialization within the inferior parietal lobe across cognitive domains. *eLife* , 1–25.
- Nunez, M.D., Vandekerckhove, J., Srinivasan, R., 2017. How attention influences perceptual decision making: Single-trial eeg correlates of drift-diffusion model parameters. *Journal of Mathematical Psychology* 76, 117–130. doi:10.1016/j.jmp.2016.03.003.
- Ondobaka, S., Lange, F.P.D., Wittmann, M., Frith, C.D., Bekkering, H., 2015. Interplay between conceptual expectations and movement predictions underlies action understanding. *Cerebral Cortex* 25, 2566–2573. doi:10.1093/cercor/bhu056.
- Onishi, K.H., Baillargeon, R., Leslie, A.M., 2007. 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica* doi:10.1016/j.actpsy.2006.09.009.
- Orban, G.A., Lanzilotto, M., Bonini, L., 2021. From observed action identity to social affordances. *Trends in Cognitive Sciences* 25, 493–505. URL: <https://doi.org/10.1016/j.tics.2021.02.012>, doi:10.1016/j.tics.2021.02.012.
- Osaka, N., Ikeda, T., Osaka, M., 2012. Effect of intentional bias on agency attribution of animated motion: An event-related fmri study. *PLoS ONE* 7, 7–12. doi:10.1371/journal.pone.0049053.
- Overwalle, F.V., 2009. Social cognition and the brain: A meta-analysis. *Human Brain Mapping* 30, 829–858. doi:10.1002/hbm.20547.
- Overwalle, F.V., Baetens, K., 2009. Understanding others’ actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage* 48, 564–584. URL: <http://dx.doi.org/10.1016/j.neuroimage.2009.06.009>, doi:10.1016/j.neuroimage.2009.06.009.

- Patel, G.H., Sestieri, C., Corbetta, M., 2019. The evolution of the temporoparietal junction and posterior superior temporal sulcus. *Cortex* 118, 38–50. URL: <https://doi.org/10.1016/j.cortex.2019.01.026>, doi:10.1016/j.cortex.2019.01.026.
- Pavlova, M.A., 2012. Biological motion processing as a hallmark of social cognition. *Cerebral Cortex* 22, 981–995. doi:10.1093/cercor/bhr156.
- Pedregosa, F., Varoquaux, G., Gramfort, A., V., B.M., Thirion, Grisel, O., Blondel, M., P., R.P., Weiss, Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 12, 2825–2830.
- Pelphrey, K., Morris, J., McCarthy, G., 2004. Grasping the intentions of others: The perceived intentionality of an action influences activity in *Journal of Cognitive Neuroscience* , 1706–1716 URL: <http://jocn.mitpress.org/cgi/content/abstract/16/10/1706>.
- Pelphrey, K.A., Morris, J.P., 2006. Brain mechanisms for interpreting the actions of others from biological-motion cues. doi:10.1111/j.0963-7214.2006.00423.x.
- Pelphrey, K.A., Shultz, S., Hudac, C.M., Wyk, B.C.V., 2011. Research review: Constraining heterogeneity: The social brain and its development in autism spectrum disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 52, 631–644. doi:10.1111/j.1469-7610.2010.02349.x.
- Perez-Osorio, J., Müller, H.J., Wiese, E., Wykowska, A., 2015. Gaze following is modulated by expectations regarding others’ action goals. *PLoS ONE* 10, 1–19. doi:10.1371/journal.pone.0143614.
- Perrett, D.I., Hietanen, J.K., Oram, M.W., Benson, P.J., 1992. Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 335, 23–30. doi:10.1098/rstb.1992.0003.
- Pitcher, D., Ungerleider, L.G., 2021. Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences* 25, 100–110. URL: <https://doi.org/10.1016/j.tics.2020.11.006>, doi:10.1016/j.tics.2020.11.006.
- van Polanen, V., Davare, M., 2015. Interactions between dorsal and ventral streams for controlling skilled grasp. *Neuropsychologia* 79, 186–191. URL: <http://dx.doi.org/10.1016/j.neuropsychologia.2015.07.010>, doi:10.1016/j.neuropsychologia.2015.07.010.
- Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fmri. *NeuroImage* 84, 320–341. URL: <http://dx.doi.org/10.1016/j.neuroimage.2013.08.048>, doi:10.1016/j.neuroimage.2013.08.048.

- Ratcliff, R., McKoon, G., 2008. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural computation* 20, 873. URL: [/pmc/articles/PMC2474742/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2474742/) [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2474742/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2474742/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2474742/), doi:10.1162/NECO.2008.12-06-420.
- Rissman, J., Gazzaley, A., D'Esposito, M., 2004. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage* 23, 752–763. doi:10.1016/j.neuroimage.2004.06.035.
- Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., Rozzi, S., 2014. Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological Reviews* 94, 655–706. doi:10.1152/physrev.00009.2013.
- Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. *Annual Review of Neuroscience* 27, 169–192. URL: <https://www.annualreviews.org/doi/abs/10.1146/annurev.neuro.27.070203.144230>, doi:10.1146/annurev.neuro.27.070203.144230.
- Rizzolatti, G., Fabbri-Destro, M., Cattaneo, L., 2009. Mirror neurons and their clinical relevance. doi:10.1038/ncpneuro0990.
- Rizzolatti, G., Matelli, M., 2003. Two different streams form the dorsal visual system: Anatomy and functions. *Experimental Brain Research* 153, 146–157. doi:10.1007/s00221-003-1588-0.
- Sakreida, K., Effnert, I., Thill, S., Menz, M.M., Jirak, D., Eickhoff, C.R., Ziemke, T., Eickhoff, S.B., Borghi, A.M., Binkofski, F., 2016. Affordance processing in segregated parieto-frontal dorsal stream sub-pathways. *Neuroscience and Biobehavioral Reviews* 69, 89–112. URL: <http://dx.doi.org/10.1016/j.neubiorev.2016.07.032>, doi:10.1016/j.neubiorev.2016.07.032.
- Sartori, L., Straulino, E., Castiello, U., 2011. How objects are grasped: The interplay between affordances and end-goals. *PLoS ONE* 6. doi:10.1371/journal.pone.0025203.
- Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughhead, J., Calkins, M.E., Eickhoff, S.B., Hakonarson, H., Gur, R.C., Gur, R.E., Wolf, D.H., 2013. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage* 64, 240–256. URL: <http://linkinghub.elsevier.com/retrieve/pii/S1053811912008609>, doi:10.1016/j.neuroimage.2012.08.052.
- Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., Kanwisher, N., 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42, 1435–1446. doi:10.1016/j.neuropsychologia.2004.04.015.
- Saygin, A.P., Chaminade, T., Ishiguro, H., Driver, J., Frith, C., 2012. The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience* 7, 413–422. doi:10.1093/scan/nsr025.

- Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T., 2018. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral Cortex* 28, 3095–3114. doi:10.1093/cercor/bhx179.
- Schultz, J., Imamizu, H., Kawato, M., Frith, C.D., 2004. Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. *Journal of Cognitive Neuroscience* 16, 1695–1705. doi:10.1162/0898929042947874.
- Shultz, S., Lee, S.M., Pelphrey, K., McCarthy, G., 2011. The posterior superior temporal sulcus is sensitive to the outcome of human and non-human goal-directed actions. *Social Cognitive and Affective Neuroscience* 6, 602–611. doi:10.1093/scan/nsq087.
- Shultz, S., McCarthy, G., 2012. Goal-directed actions activate the face-sensitive posterior superior temporal sulcus and fusiform gyrus in the absence of human-like perceptual cues. *Cerebral Cortex* 22, 1098–1106. URL: <https://academic.oup.com/cercor/article/22/5/1098/279396>, doi:10.1093/cercor/bhr180.
- Shultz, S., McCarthy, G., 2014. Perceived animacy influences the processing of human-like surface features in the fusiform gyrus. *Neuropsychologia* 60, 115–120. URL: <http://dx.doi.org/10.1016/j.neuropsychologia.2014.05.019>, doi:10.1016/j.neuropsychologia.2014.05.019.
- Sliwa, J., Freiwald, W.A., 2017. Neuroscience: A dedicated network for social interaction processing in the primate brain. *Science* 356, 745–749. URL: <http://science.sciencemag.org/>, doi:10.1126/science.aam6383.
- Sokolov, A.A., Zeidman, P., Erb, M., Ryvlin, P., Friston, K.J., Pavlova, M.A., 2018. Structural and effective brain connectivity underlying biological motion detection. *Proceedings of the National Academy of Sciences* 115, E12034–E12042. doi:10.1073/pnas.1812859115.
- Spreng, R.N., Stevens, W.D., Chamberlain, J.P., Gilmore, A.W., Schacter, D.L., 2010. Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage* 53, 303–317. URL: <http://dx.doi.org/10.1016/j.neuroimage.2010.06.016>, doi:10.1016/j.neuroimage.2010.06.016.
- Spunt, R.P., Falk, E.B., Lieberman, M.D., 2010. Dissociable neural systems support retrieval of how and why action knowledge. *Psychological Science* 21, 1593–1598. URL: https://journals.sagepub.com/doi/10.1177/0956797610386618?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed, doi:10.1177/0956797610386618.
- Spunt, R.P., Kemmerer, D., Adolphs, R., 2016. The neural basis of conceptualizing the same action at different levels of abstraction. *Social Cognitive and Affective Neuroscience* 11, 1141–1151. URL: <https://pubmed.ncbi.nlm.nih.gov/26117505/>, doi:10.1093/scan/nsv084.

- Stehr, D.A., Zhou, X., Tisby, M., Hwu, P.T., Pyles, J.A., Grossman, E.D., 2021. Top-down attention guidance shapes action encoding in the psts. *Cerebral Cortex* 31, 3522–3535. doi:10.1093/cercor/bhab029.
- Stöcker, C., Hoffmann, J., 2004. The ideomotor principle and motor sequence acquisition: Tone effects facilitate movement chunking. *Psychological Research* 68, 126–137. doi:10.1007/s00426-003-0150-9.
- Summerfield, C., Egnér, T., 2009. Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences* 13, 403–409. doi:10.1016/j.tics.2009.06.003.
- Tamir, D.I., Thornton, M.A., 2018. Modeling the predictive social mind. *Trends in Cognitive Sciences* 22, 201–212. URL: <http://dx.doi.org/10.1016/j.tics.2017.12.005>, doi:10.1016/j.tics.2017.12.005.
- Tavares, P., Lawrence, A.D., Barnard, P.J., 2008. Paying attention to social meaning: An fmri study. *Cerebral Cortex* 18, 1876–1885. doi:10.1093/cercor/bhm212.
- Teufel, C., Fletcher, P.C., Davis, G., 2010. Seeing other minds: Attributed mental states influence perception. *Trends in Cognitive Sciences* 14, 376–382. URL: <http://dx.doi.org/10.1016/j.tics.2010.05.005>, doi:10.1016/j.tics.2010.05.005.
- Thompson, E.L., Bird, G., Catmur, C., 2019. Conceptualizing and testing action understanding. *Neuroscience and Biobehavioral Reviews* 105, 106–114. doi:10.1016/j.neubiorev.2019.08.002.
- Turner, B.M., Forstmann, B.U., Wagenmakers, E.J., Brown, S.D., Sederberg, P.B., Steyvers, M., 2013. A bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage* 72, 193–206. URL: <https://pubmed.ncbi.nlm.nih.gov/23370060/>, doi:10.1016/J.NEUROIMAGE.2013.01.048.
- Turner, B.M., Maanen, L.V., Forstmann, B.U., 2015. Informing cognitive abstractions through neuroimaging: the neural drift diffusion model. *Psychological review* 122, 312–336. URL: <https://pubmed.ncbi.nlm.nih.gov/25844875/>, doi:10.1037/A0038894.
- Turner, B.O., Mumford, J.A., Poldrack, R.A., Ashby, F.G., 2012. Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. *NeuroImage* 62, 1429–1438. URL: <http://dx.doi.org/10.1016/j.neuroimage.2012.05.057>, doi:10.1016/j.neuroimage.2012.05.057.
- Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging* 29, 1310–1320. doi:10.1109/TMI.2010.2046908.
- Urgen, B.A., Saygin, A.P., 2020. Predictive processing account of action perception: Evidence from effective connectivity in the action observation network. *Cortex* 128, 132–142. URL: <https://doi.org/10.1016/j.cortex.2020.03.014>, doi:10.1016/j.cortex.2020.03.014.

- Vallacher, R.R., Wegner, D.M., 1987. What do people think they're doing? action identification and human behavior. *Psychological Review* 94, 3–15. doi:10.1037/0033-295X.94.1.3.
- Walbrin, J., Downing, P., Koldewyn, K., 2018. Neural responses to visually observed social interactions. *Neuropsychologia* 112, 31–39. URL: <https://doi.org/10.1016/j.neuropsychologia.2018.02.023>, doi:10.1016/j.neuropsychologia.2018.02.023.
- Weiner, K.S., Grill-Spector, K., 2013. Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychological Research* 77, 74–97. doi:10.1007/s00426-011-0392-x.
- Wolpert, D.M., Doya, K., Kawato, M., 2003. A unifying computational framework for motor control and social interaction. doi:10.1098/rstb.2002.1238.
- Wolpert, D.M., Ghahramani, Z., 2000. Computational principles of movement neuroscience. *Nature Neuroscience* 3, 1212–1217. URL: <http://neurosci.nature.com>, doi:10.1038/81497.
- Woodward, A.L., 1998. Infants selectively encode the goal object of an actor's reach. *Cognition* 69, 1–34. doi:10.1016/S0010-0277(98)00058-4.
- Wurm, M.F., Ariani, G., Greenlee, M.W., Lingnau, A., 2016. Decoding concrete and abstract action representations during explicit and implicit conceptual processing. *Cerebral Cortex* 26, 3390–3401. doi:10.1093/cercor/bhv169.
- Wurm, M.F., Caramazza, A., 2019. Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nature Communications* 10. URL: <https://doi.org/10.1038/s41467-018-08084-y>, doi:10.1038/s41467-018-08084-y.
- Wurm, M.F., Caramazza, A., 2021. Two 'what' pathways for action and object recognition. *Trends in Cognitive Sciences* 26, 103–116. URL: <https://doi.org/10.1016/j.tics.2021.10.003>, doi:10.1016/j.tics.2021.10.003.
- Wyk, B.C., Hudac, C.M., Carter, E.J., Sobel, D.M., Pelphrey, K.A., 2009. Action understanding in the superior temporal sulcus region. *Psychological Science* 20, 771–777. doi:10.1111/j.1467-9280.2009.02359.x.
- Wyk, B.C.V., Voos, A., Pelphrey, K.A., 2012. Action representation in the superior temporal sulcus in children and adults: An fmri study. *Developmental Cognitive Neuroscience* 2, 409–416. URL: <http://dx.doi.org/10.1016/j.dcn.2012.04.004>, doi:10.1016/j.dcn.2012.04.004.
- Yang, D.Y., Rosenblau, G., Keifer, C., Pelphrey, K.A., 2015. An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience and Biobehavioral Reviews* 51, 263–275. URL: <http://dx.doi.org/10.1016/j.neubiorev.2015.01.020>, doi:10.1016/j.neubiorev.2015.01.020.

- Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., Fisch, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* 106, 1125–1165. doi:10.1152/jn.00338.2011.
- Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C.J., Hasson, U., 2017. Same story, different story: The neural representation of interpretive frameworks. *Psychological Science* 28, 307–319. URL: www.psychologicalscience.org/PS, doi:10.1177/0956797616682029.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* 20, 45–57. doi:10.1109/42.906424.
- Zhou, X., Stehr, D.A., Pyles, J., Grossman, E.D., 2023. Configuration of the action observation network depends on the goals of the observer. *Neuropsychologia* 191. doi:10.1016/j.neuropsychologia.2023.108704.

Appendix A

Appendix Title

A.1 fmriprep

Appendix B

Appendix

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 21.0.1 (Esteban et al. (2018); Esteban et al. (2018); RRID:SCR_016216), which is based on *Nipype* 1.6.1 (Gorgolewski et al. (2011); Gorgolewski et al. (2018); RRID:SCR_002502).

Preprocessing of B0 inhomogeneity mappings A total of 1 fieldmaps were found available within the input BIDS structure for this particular subject. A *B0*-nonuniformity map (or *fieldmap*) was estimated based on two (or more) echo-planar imaging (EPI) references with `topup` (Andersson et al. (2003); FSL 6.0.5.1:57b01774).

Anatomical data preprocessing A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-

extracted T1w using `fast` (FSL 6.0.5.1:57b01774, RRID:SCR_002823, Zhang et al., 2001). Brain surfaces were reconstructed `recon-all` (FreeSurfer 6.0.1, RRID:SCR_001847, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al., 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing For each of the 11 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 6.0.5.1:57b01774, ?). The estimated *fieldmap* was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. BOLD runs were slice-time corrected to 0.696s (0.5 of slice acquisition range 0s-1.39s) using `3dTshift` from AFNI (Cox and Hyde, 1997, RRID:SCR_005927). The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve and Fischl, 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al.

(2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer’s *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each *CompCor* decomposition, the k components with the largest singular values are retained, such that the retained components’ time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. The BOLD

time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): *fsnative*, *fsaverage*. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.8.1 (Pedregosa et al., 2011, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation.

Copyright Waiver

The above boilerplate text was automatically generated by *fMRIPrep* with the express intention that users should copy and paste this text into their manuscripts *unchanged*. It is released under the CC0 license.