Universität
Münster

Rosari Naveena Selvan

The Predictive Brain in Action:
Multimodal Evidence for Hierarchical Predictive Processing at Action Boundaries

- 2026 -

Psychologie

The Predictive Brain in Action:
Multimodal Evidence for Hierarchical Predictive Processing at Action Boundaries

Inaugural-Dissertation
zur Erlangung des Doktorgrades
im Fachbereich Psychologie und Sportwissenschaft
der Universität Münster

by

Rosari Naveena Selvan

from *Coimbatore, Tamil Nadu, India*

- 2026 -

Doctoral Committee : Prof. Dr. Ricarda Schubotz
Prof. Dr. Florentin Wörgötter
Prof. Dr. Markus Lappe

Dean : Prof. Dr. Ricarda I. Schubotz
First Reviewer : Prof. Dr. Ricarda I. Schubotz
Second Reviewer : Prof. Dr. Markus Lappe

Date of thesis defense : 27th Mar 2026
Date of graduation : 27th Mar 2026

Dedicated to my father ***Selvan***

അച്ഛാ, മൂല്യങ്ങൾക്കും ജീവിതപാഠങ്ങൾക്കും വളരെ നന്ദി

and

my brother ***Naveen***

The one I love the most, always and forever

> காலத்தி னாற்செய்த நன்றி சிறிதெனினும்
> ஞாலத்தின் மாணப் பெரிது.
>
> திருக்குறள் 102, திருவள்ளுவர்
>
> *A favor, though small, bestowed at a time of need transcends the earth.*

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Ricarda Schubotz, for her continuous guidance, encouragement, and valuable feedback throughout this thesis. I am also deeply thankful to my co-supervisor, Prof. Dr. Florentin Wörgötter, for his constant support and the expertise he brought to this project. Furthermore, I would like to thank Prof. Dr. Markus Lappe for his valuable scientific input. I also thank my co-author Minghao Cheng for his invaluable input and support, as well as for generously teaching me all the computer vision techniques required for this project. I am grateful to my co-authors Minija Tamosiunaite, Anoushiravan Zahedi, Sophie Siestrup, Falko Mecklenbrauck, and Katharina Dohr for their collaboration and important contributions to this work.

I gratefully acknowledge Monika Mertens, Jochen Bauer, Birgit Vahrenkamp, and Maximilian Bruchmann for their assistance and technical support throughout the project. My appreciation also extends to the group secretaries, Jutta Linke and Ursula Hahn-Wörgötter, for their help with administrative and bureaucratic matters, as well as to the IT team and Technical Services for their reliable support. I would further like to thank all student assistants Lena Puder, Maria Hofmann, Simon Wieczorek, Kristin Stroop, Lars Schlüter, Leon Exeler, Niklas Dielitzsch, Thea Bröring, Anna Asdonk, Maike Denguth, Pia Niedrée, Yannah Schröder, and Nihan Yildirim for their contributions.

My biggest thanks go to the *World's Best Colleague*, Benjamin Jainta, for standing by me from day one of my PhD and for providing constant emotional support and valuable scientific input. I am also deeply grateful to Viviana Barraza, the *Best Coffee Partner* and PhD friend one could ask for, thank you for listening to all my rants over coffee and cake and introducing me to a bit of pop culture. I thank my Colleagues-turned-Friends Marius Boeltzig, Marlen Roehe, and Jennifer Pomp, whose company, conversations, and constant support kept me going. My special thanks go to my office partner, Marc Dührkop, for his encouragement and the fun we shared during the final phase of my work, as well as for teaching me the offside rule in football. I also thank my colleagues Anas Al-Naji, Nina Liedtke and Lena Innig, and for their companionship and support.

I am especially thankful to my friend, Christin Schwarzer, for her unwavering encouragement, understanding, and support, and for filling this journey with so much love and joy. My gratitude goes to Torge Dellert, with whom a connection forged in the graduate program grew into a meaningful and enduring friendship over the past five years. I would like to thank the members of our *Party Planning Committee*: Franziska Stanke for countless coffee and dinner meetups, long walks, and constant motivation, and Mareike Ehlert and Jan Beck for Friday lunches and for celebrating both successes and failures throughout every step of our PhD journey.

I would like to extend my deepest gratitude to my school *Biology teacher*, Kumar, who was the first to recognize my interest in exploring the brain and encouraged me to participate in science projects that fostered this passion. I also thank my *Zoology professor*, Joseph Reginald, who taught me how to conduct research and believed in my potential to become a researcher. Finally, I am grateful to my Chosen siblings, Bhalaji Anna, who has been a constant source of support for over twenty years since my school days, and Ganga and Ann, who quickly became my sisters, have shared high and low with me and nurtured our shared passion for neuroscience together.

I also thank my friends Visal and Ameena for their emotional support and the strong bond we share, which made this journey easier and filled it with moments of joy. I would also like to thank my friend Kiruba, who came into my life during one of my most difficult periods and helped make it brighter. My deepest thanks go to my pillars of support, my best friends and chosen family Agam, Prachi, Varun, and Mahemaa. Without them, my journey from my post-graduation to the completion of my PhD would not have been possible. Thank you for always being there and for keeping me grounded and sane, no matter what life brought along the way.

Finally, I would like to thank my family – அறிவின் ஆற்றலின் நம்பி எனக்கு கல்வியை வழங்கியதற்கு மிக்க நன்றி அம்மா.

Table of Contents

Summary

This dissertation investigates predictive processing in human action perception and joint action, employing a multimodal approach to examine hierarchical prediction across perception and social interaction. The research comprises three studies, each designed to provide complementary insights and progressively extend the investigation from controlled action observation to interactive coordination.

Studies 1 and 2 are closely interconnected, utilizing the same structured, multi-step action repertoire to probe predictive mechanisms through distinct methodologies. Study 1 employed functional magnetic resonance imaging (fMRI) to explore the neural bases of action prediction at action boundaries. Participants learned probabilistic action sequences and subsequently observed these sequences during scanning. Across all three studies, action boundaries were defined as Touching (T) events, when the hand touched an object, and Untouching (U) events, when the hand released the object. These events marked the onset and offset of an action, respectively. Action boundaries were precisely identified using computer vision algorithms, allowing analyses to be tightly time-locked to critical moments at which predictions were likely generated or violated. Results indicated that different neural processes support the retrieval of potential action outcomes versus the selection of specific predictions. Fronto-parietal networks were prominently involved, underscoring the hierarchical organization of predictive computations in the brain. These findings provide strong evidence for hierarchical predictive coding, demonstrating that the brain anticipates future actions based on learned temporal structure.

Building on these neural insights, Study 2 (eye tracking and pupillometry) examined how prediction errors are reflected in eye movements. Participants observed both previously learned sequences and “illegal” variants in which a single step violated expectations. Using

computer-vision-defined action boundaries, the study measured fixation patterns, saccade velocities, and pupil responses precisely at points of prediction error. Violations were associated with increased pupil dilation, faster saccades, and more exploratory fixations, indicating that the visual system reallocates attention to resolve uncertainty when expectations are disrupted. By employing the same sequences as Study 1, this study provides complementary behavioral evidence that oculomotor responses track predictive processing in a temporally precise manner, aligning closely with neural markers.

Study 3 extended the investigation to interactive contexts using electroencephalography (EEG), focusing on predictive processing during joint action. Participants coordinated actions with a partner under hierarchical constraints, including goals, sub-goals, and movement-level coordination. EEG oscillatory activity was analyzed time-locked to computer-vision-defined action boundaries, same as study 1 and 2, at T and U events revealing that predictive mechanisms guide real-time interpersonal coordination. Participants anticipated their partner's actions and adjusted their own behavior, accordingly, demonstrating that predictive computations operate not only during observation but also in active, socially embedded interactions. This study addresses limitations of the first two studies, including reliance on trained sequences and passive viewing, and highlights the complexity of predictive processes in naturalistic, interactive scenarios.

Across all three studies, the use of computer-vision-defined action boundaries was a central methodological addition, allowing precise temporal alignment of neural, oculomotor, and behavioral measures. The combination of fMRI, eye tracking, and EEG provides complementary perspectives: fMRI identifies the cortical areas supporting hierarchical predictions, eye tracking captures rapid behavioral responses to prediction errors, and EEG reveals the timing and dynamics of predictive computations in social coordination.

Collectively, the three studies show that predictive processing is a fundamental feature of action perception and joint action, expressed hierarchically across neural, behavioral, and oculomotor systems. While Studies 1 and 2 provide detailed insight under tightly controlled conditions, Study 3 increases ecological validity by examining prediction in interactive, real-time coordination. The research design reflects a coherent progression: each study addresses limitations of the others while offering converging evidence for hierarchical predictive coding theories.

In summary, this dissertation demonstrates that humans continually anticipate, monitor, and adjust predictions during structured and interactive actions. By combining multiple measurement techniques and leveraging computer-vision-based temporal markers, the studies provide an integrated view of predictive processing across perception, behavior, and social interaction. These findings lay the groundwork for future research on spontaneous, socially meaningful actions, computational modeling of predictive uncertainty, and hyperscanning approaches to study shared neural dynamics in real-world contexts.

List of Original Publications

This thesis is based on the following original research articles:

Study 1: Selvan, R. N., Cheng, M., Siestrup, S., Mecklenbrauck, F., Jainta, B., Pomp, J., Zahedi, A., Tamosiunaite, M., Wörgötter, F., & Schubotz, R. I. (2024). Updating predictions in a complex repertoire of actions and its neural representation. *NeuroImage*, *296,* 120687. https://doi.org/10.1016/j.neuroimage.2024.120687

Study 2: Selvan, R. N. Zahedi, A., Cheng, M., Tamosiunaite, M., Wörgötter, F., & Schubotz, R. I. Seeing the Unexpected: Oculomotor Signatures of Prediction Error at Action Boundaries. *(Manuscript under Review)*

Study 3: Selvan, R. N. Zahedi, A., Cheng, M., Dohr, K., Tamosiunaite, M., Wörgötter, F., & Schubotz, R. I. Me or Us? Mine or Yours? Here or There? Neural and behavioral responses of prediction in Joint Action. *(Manuscript to be submitted for Peer Review)*

1. Theoretical and Empirical Background

Human action unfolds in a continuous stream of movements that must be rapidly perceived, interpreted, and predicted to support adaptive behavior. From recognizing another person's intentions to coordinating actions in social contexts, the ability to understand action is fundamental to perception, cognition, and social interaction. Over the past decades, research in cognitive neuroscience, supported by methods such as functional neuroimaging, electrophysiology, eye tracking, and computational modeling, has converged on the view that action understanding is not a passive perceptual process, but rather relies on active, predictive mechanisms grounded in sensorimotor systems.

Building on this perspective, the following sections outline a multi-level account of how actions are perceived, predicted, and coordinated, spanning individual cognition, social interaction, and application of computer vision. First, we review theoretical and empirical work on *action observation and event segmentation*, emphasizing how continuous behavior is structured into meaningful units through predictive mechanisms. We then examine *action prediction and prediction error* within the predictive coding framework, highlighting neural and behavioral evidence for hierarchical inference during action observation. Extending from individual to social cognition, we discuss *joint action and shared representations*, integrating theoretical models and electrophysiological findings that reveal how predictive processes are coordinated across interacting agents. Finally, we turn to *computer vision approaches* to action, illustrating how formal models of action representation and prediction such as semantic event chains offer mechanistic insights that bridge neuroscience, psychology, and artificial intelligence. Together, these perspectives converge on a unified view of action understanding as an active, predictive, and hierarchical process.

1.1. Action Observation and Segmentation

Everyday actions though seem to be continuous streams of movement, yet observers reliably perceive them as structured sequences of discrete events. The process by which continuous activity is parsed into meaningful units is known as *event segmentation*, and it plays a central role in action understanding, prediction, and memory (Zacks & Tversky, 2001). A comprehensive theoretical and empirical framework has been developed to explain how observers segment ongoing actions into events, emphasizing the interaction between perceptual input, prediction, and cognitive control (Zacks et al., 2001, 2007).

According to *Event Segmentation Theory (EST)*, event segmentation arises from the brain's attempt to maintain accurate predictions about ongoing activity. Observers continuously generate predictions about upcoming perceptual input based on internal event models. When predictions are accurate, the current event model is maintained. When predictions fail to match incoming sensory input, *prediction error (PE)* is generated, (K. Friston, 2005). Studies have shown that prediction error signals peak at event boundaries, indicating moments of model updating (Zacks et al., 2011).

One of the earlier behavioural studies showed that unit or event segmentation varied quantitatively, whether the subjects encountered an unexpected event. After an unexpected event, the subjects employed more units than the control ones who didn't see it (Newtson, 1973). Similar difference in segmentation has also been observed between the familiar and unfamiliar activities. This shows that event segmentation involves top-down cognitive representations of events, and it is not just the bottom-up perceptual information processing (Zacks & Tversky, 2001). Studies have also shown the placement of event boundaries when viewing naturalistic actions, such watching movies and during action observation (Kurby & Zacks, 2008).

Neuroimaging evidence supports the predictive and model-based account of event segmentation. Functional MRI (fMRI) studies have shown transient increase in multiple brain regions at the event boundaries. This includes the activation of the medial visual cortex, the precuneus and intraparietal sulcus (Kurby & Zacks, 2008; Zacks et al., 2007). The activation in the Default-Mode Network (DMN) decreases at event boundaries while watching movies (Zacks et al., 2010) and listening to music (Burunat et al., 2024). Event boundaries are also associated with hippocampal responses, indicating their importance for episodic memory encoding (Ben-Yakov & Dudai, 2011). Consistent with this view, objects that were presented at the event boundary more effectively were better remembered, suggesting that event boundaries structure memory representations (Swallow et al., 2009). More recently, Pomp et al., (2021) demonstrated that hand touching an object (Touching events) constitute salient segmentation cues, eliciting activation of frontal, hippocampal and insula regions.

The phenomenon of event segmentation is not only observed in adults but also in infants with distinct differences. The working memory capacity (Gathercole et al., 2004) and semantic knowledge (Li et al., 2004) are very much limited in infants compared to adults. However, studies have shown that babies track and use probabilities during observation of continuous events (Roseberry et al., 2011; Stahl et al., 2014). While infants segment events in ways that resemble adults and show memory benefits at event boundaries, the mechanisms underlying retrieval and updating of internal event models during development remain poorly understood.

Within the broader context of action observation, event segmentation provides a critical link between perception and prediction. By partitioning continuous activity into discrete events, observers reduce computational complexity and enable more precise predictions about future action states. Event boundaries thus mark moments of heightened prediction error and model revision, supporting adaptive understanding of dynamic behavior.

1.2. Action Prediction and Prediction Error

The core function of the human cognitive system is the ability to anticipate the actions of others, enabling rapid and adaptive responses in dynamic environments. The *predictive coding framework* provides a principled computational account of how the brain achieves this capacity, positing that perception and action are driven by hierarchical predictions and the minimization of prediction error (Clark, 2013; K. Friston, 2005). The brain minimizes the prediction error in one of the two ways: *perceptual inference* – the process of changing the predictions to match the actual sensory input or *active inference* – the process of changing the environment to match the predictions (Adams et al., 2013; K. Friston & Kiebel, 2009a).

During action observation, humans can not only predict the action but can also infer the intentions of those actions. The major reason that we are able to read the goal of an action is the Mirror Neuron System (MNS) (Fogassi et al., 2005; Frith & Frith, 1999; Rizzolatti & Craighero, 2004). According to Hamilton & Grafton (2012), actions can be described at four levels, namely the *intention level* – gives the intention and long-term goal, the *goal level* – that explains the short-term goals, the *motor signals* – the pattern of muscle activity and the *kinematic level* – the movement in space and time. Observers can infer goals and intentions by mapping observed kinematic information onto internal motor representations. Through this mechanism, observing an action engages premotor and parietal regions that simulate the observed behavior, providing access to its goal structure (Rizzolatti & Craighero, 2004).

Empirical research provides substantial evidence supporting predictive coding accounts of action prediction. During the updating of predictions, fMRI studies have shown the activation of the AON (Sasaki et al., 2018; Urgen et al., 2019; Urgen & Saygin, 2020) suggesting a hierarchical structure in predictive processing. Additionally, the neurocognitive studies have shown increased DMN-AON connectivity while inferring the intentions or beliefs of actions (Koster-Hale & Saxe, 2013; Wurm et al., 2011). Alexander & Brown, (2015,

2018) showed that the Medial Prefrontal Cortex (MPFC), part of the DMN provides input to the lateral prefrontal cortex (LPFC), part of AON suggesting a hierarchical predictive and predictive error processing.

In addition to the neuroimaging methods, eye-tracking has become a powerful tool for probing predictive processing and prediction error in perception and cognition because gaze behavior continuously reflects *where* and *when* observers anticipate future sensory events (Eisenberg et al., 2018; Hayhoe et al., 2012). Studies have looked at different measures such as fixation that reflects increased visual sampling (Brunyé & Gardony, 2017a; Torrents-Rodas et al., 2021), saccade velocity association with visuospatial attention (Go et al., 2022) and pupillometry denotes the arousal in the aspects of prediction error (Mathôt, 2018; Viglione et al., 2023). Across these measures, converging evidence indicates that the brain continuously updates internal models to reduce uncertainty (Yu & Dayan, 2005).

The minimization of prediction error is not only central to perception but also to learning and adaptation. Repeated exposure to consistent action patterns strengthens internal models, reducing the magnitude of prediction errors and enhancing the precision of future predictions. Conversely, unexpected or novel actions generate larger errors, prompting model updating and refinement. This dynamic aligns closely with event segmentation, as prediction errors often coincide with event boundaries, thereby linking action prediction to the structuring of continuous behavior.

Predictive coding thus provides a unifying framework for understanding action anticipation. By continuously generating and updating hierarchical predictions, observers can anticipate movement trajectories, detect expectation violations, and flexibly adapt to dynamic environments. Action prediction is therefore best understood as an active, inferential process grounded in sensorimotor and cognitive systems, rather than as passive sensory processing.

1.3. Joint Action and Shared Representation

Most of the time, in our day to life, it is not only about action observation. Human social interaction frequently involves joint action, in which two or more individuals coordinate their behaviors to achieve shared goals. Unlike isolated action, joint action requires not only predicting one's own movements but also anticipating and integrating the actions of others (Sebanz et al., 2006; Vesper et al., 2010). Predictive mechanisms that support individual action observation and execution are therefore extended and coupled across interacting agents, allowing partners to synchronize behavior, minimize errors, and adapt to dynamic contingencies (K. Friston, 2005; Kilner et al., 2007). According to Sebanz & Knoblich, (2009), successful joint action depends on predicting three key aspects of a partner's behavior: *what* action will be performed, *when* it will occur, and *where* it will take place. Together, these predictive dimensions enable the integration of one's own actions with those of others.

One of the prominent theoretical frameworks include the Predictive Joint Action Model (PJAM), which formalizes joint action as the dynamic coupling of hierarchical predictive models across interacting agents (Pesquita et al., 2018). PJAM proposes three levels of representation: at the highest level, shared *goal representations* specify the intended joint outcome; at the intermediate level, joint *action planning* integrates predictions about one's own and the partner's actions; and at the lowest level, *sensory routing* mechanisms compare predicted and actual sensory consequences of joint behavior. Another model proposed by (Kahl & Kopp, 2018), which distinguishes between a *schema or goal* level (clusters of actions grouped by similarity), a *sequence or subgoal* level (ordered motor acts), and a *sensorimotor or movement* level encompassing visual and motor control. Together, these frameworks converge on the idea that joint action relies on hierarchical prediction entrainment, enabling smooth coordination between multiple agents.

Empirical EEG research on joint action has provided compelling evidence for shared predictive representations and anticipatory neural processing when individuals coordinate their behavior with others. In joint action planning tasks, high-density EEG reveals that anticipatory motor preparation of one's own and a partner's actions is associated with modulations of slow cortical potentials such as the Contingent Negative Variation (CNV), reflecting preparatory processes that align temporal expectations with upcoming coordinated behaviors - CNV amplitude varies with joint action demands and correlates with performance (Fanghella et al., 2025; Kourtis et al., 2019a). Likewise, classic P300 components (including P3a and P3b sub-components) are enhanced or shifted during joint tasks compared to individual actions, indexing higher-order processes of stimulus evaluation and updating of action representations required for predicting co-actors' contributions to the shared goal (Kourtis, Sebanz, Knoblich, et al., 2013).

Beyond time-locked ERP signals, frequency-specific neural oscillations provide additional insight into predictive and interactive dynamics during joint action. Alpha (8–13 Hz) and mu rhythms over sensorimotor regions show task-dependent suppression during joint planning and execution, consistent with sensorimotor simulation and the formation of shared action models, whereas modulation of beta (13–30 Hz) rhythms have been implicated in integration of self and other doing a joint action reflecting shared sensorimotor framework (Bolt & Loehr, 2024; Novembre et al., 2016). Furthermore, theta band (4–8 Hz) interbrain synchrony during cooperative tasks (Wang et al., 2020) and increased midline theta during negative feedback that should result in adjustment of the shared representation (Markiewicz et al., 2024). Together, these oscillatory signatures complement ERP findings by revealing how predictive processes unfold across multiple temporal scales.

Integrating PJAM with hierarchical action models highlights how prediction, error monitoring, and model updating extend from individual to shared cognition. In joint action, prediction errors not only refine one's own motor plans but also update expectations about a partner's behavior, resulting in a coupled, mutually adaptive system. This perspective underscores the continuity between individual action understanding, predictive processing, and social coordination, framing joint action as an inherently predictive, hierarchical, and interactive process.

1.4. Computer Vision Approaches to Action

The study of human action has increasingly benefited from computational modeling and computer vision approaches, which provide formal frameworks for understanding, simulating, and predicting action. These methods bridge cognitive neuroscience, psychology, robotics and artificial intelligence, allowing researchers to formalize theoretical principles such as predictive coding, event segmentation, and joint action into explicit computational architectures. By translating cognitive theories into implementable models, computer vision approaches offer precise, testable accounts of how actions are represented and anticipated.

Computer vision and machine learning techniques enable extraction, classification, and prediction of human actions from visual data. Traditional approaches relied on handcrafted features such as motion trajectories, body pose, and spatiotemporal patterns. More recently, deep learning methods, particularly convolutional neural networks (CNNs) and transformer architectures, have allowed automatic learning of hierarchical representations directly from raw video input (Girdhar et al., 2019; Simonyan & Zisserman, 2014). These systems can recognize complex actions, anticipate future movements, and even model interactions between multiple agents, making them conceptually aligned with theories of joint action and shared predictive models.

A major focus of computational approaches is the prediction of continuous action sequences. Recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and temporal convolutional networks have been used to model the temporal dynamics of action, enabling the anticipation of upcoming movements or goals (Fragkiadaki et al., 2015). These architectures parallel neural accounts of temporal integration and forward modeling, providing a computational analogue of how prediction errors guide online updating of internal models.

A seminal framework by Aksoy et al., (2011) demonstrated that semantic relations between objects and hands at decisive time points can be learned from observed manipulation sequences, using *semantic event chains (SECs)* to encode topological changes in visual scenes that distinguish different manipulations and support generalization to novel contexts. Building on these ideas, researchers have formalized simple ontologies of manipulation actions that explicitly capture the spatial predicates and relational dynamics between hand and object(s), enabling more abstract and interpretable representations of action structure beyond traditional kinematic descriptors (Wörgötter et al., 2013a). Most recently, the *enriched semantic event chain (eSEC)* approach extends SECs by incorporating multiple classes of spatial relations - touching/non-touching, static spatial relations (e.g., above, around), and dynamic relations (e.g., getting closer, moving apart) into a matrix representation that can be used not only for action classification but also for action prediction before execution is completed, even in the absence of contextual or object identity information (Ziaeetabar et al., 2021a). These computational models illustrate how minimal spatial relational information can serve as a basis for *predictive coding* of actions, mirroring cognitive processes hypothesized in humans and providing a bridge between perception, semantic structure, and anticipatory action understanding.

Computer vision approaches have broad applications in robotics, human–computer interaction, and neurocognitive research. By implementing biologically inspired predictive architectures, artificial agents can anticipate human actions, coordinate with humans in real time, and learn from observation. This convergence highlights a reciprocal relationship between neuroscience and artificial intelligence: computational models provide mechanistic hypotheses about action perception and prediction, while empirical findings from neuroscience constrain and refine these models. Together, these approaches underscore the value of computational frameworks for understanding action as a predictive, structured, and temporally unfolding process.

Taken together, the theoretical and empirical evidence reviewed in this chapter highlights that human action perception and coordination are inherently predictive, hierarchical, and structured. Continuous actions are segmented into meaningful units, with event boundaries serving as critical points for updating internal models, generating predictions, and signaling prediction errors. These processes operate across multiple levels from low-level sensorimotor dynamics to higher-order representations of goals and intentions and extend seamlessly from individual action perception to joint action, where predictive models are shared and aligned across interacting agents.

The present thesis addresses remaining gaps by adopting a multi-level, multimodal approach, integrating functional neuroimaging, EEG, eye tracking, pupillometry, and interactive paradigms to examine predictive processing in both individual and social contexts. Through these three studies, this research investigates how action boundaries objectively defined using computer vision function as hubs for generating and updating predictions, how prediction errors are expressed in neural and oculomotor dynamics, and how shared predictive models enable coordinated behavior during joint action. By focusing on the hierarchical organization of prediction, the probabilistic structure of learned actions, and the dynamic coupling of predictive processes between agents, this thesis provides a comprehensive account of the mechanisms supporting efficient action perception, and social coordination. Ultimately, integrating computationally defined boundaries with behavioral, and neural measures advances a unified framework in which action boundaries, prediction errors, and hierarchical predictive models are central to understanding how humans navigate both solitary and interactive behavioral environments.

2. Research Questions and Objectives

Human action perception relies on predictive processing: observers continuously generate expectations about unfolding actions and update these predictions when violations occur. Although actions appear continuous, they are segmented into meaningful units, with action boundaries playing a critical role in prediction updating. Across individual and social contexts, predictive mechanisms operate at multiple hierarchical levels, from low-level sensorimotor features to higher-level representations of goals and intentions. Within this framework, action boundaries mark moments of increased computational demand, as prior predictions are terminated, internal models are evaluated, and new predictions about upcoming actions are generated. Understanding how the brain identifies and utilizes these boundaries is therefore central to explaining how humans efficiently perceive, predict, learn, and adapt to complex action sequences.

Recent research suggests that predictive processing during action perception is expressed across multiple modalities and timescales, including neural activity, eye movements, and interpersonal coordination. Neural evidence indicates that distinct brain networks support action segmentation, prediction, and prediction error processing, while oculomotor behavior provides a sensitive, real-time index of anticipatory processes and surprise. Beyond individual observers, predictive mechanisms also play a crucial role in joint action, where successful coordination depends on aligning predictions across interacting agents at multiple hierarchical levels, from movements to shared goals. However, the mechanisms by which these levels of prediction are dynamically coordinated at action boundaries, and how prediction errors propagate across neural, behavioral, and social domains, have yet to be fully characterized. The present set of studies addresses this gap by adopting a multi-level, multimodal approach to investigate how prediction, prediction errors, and action boundaries shape action prediction and coordination in both individual and interactive contexts.

Research Questions

1. How do neural responses differ at the action boundaries Untouching (U) and Touching (T) during the predictive processing of action repertoire and how does the probabilistic structure of learned action sequences modulate neural activity at these boundaries?
2. How do moment-to-moment oculomotor dynamics characterize the detection and processing of prediction errors (PE) at action boundaries during action observation?
3. How are neural oscillatory dynamics at the action boundaries Untouching (U) and Touching (T) modulated by varying degrees of flexible coupling across different hierarchical levels of action coordination during joint action?

Study 1 examines the neural dynamics of predictive processing during the observation of complex, learned action repertoire using fMRI. By distinguishing action boundaries and manipulating prediction violations to induce Prediction Errors (PEs), the study identifies distinct neural signatures associated with predive processing. The findings demonstrate that action boundaries serve as key hubs for retrieving learned action repertoire and generating predictions about upcoming actions, supporting hierarchical predictive models of action observation.

Study 2 investigates how prediction errors during action observation are expressed in eye movements and pupil dynamics. Using eye tracking and pupillometry, the study reveals a coordinated oculomotor response to Prediction Errors (PEs) during action observation, characterized by increased arousal, faster saccades, and more exploratory fixation behavior. These results suggest that gaze behavior provides a sensitive, time-resolved marker of prediction error processing and internal model updating.

Study 3 extends predictive processing research into a social context by examining joint action coordination in a naturalistic, game-based paradigm. By manipulating shared intention and coordination at multiple hierarchical levels, the study explores how predictive models are aligned between interacting partners. Behavioral, and EEG measures reveal how predictive coupling supports efficient interpersonal coordination, highlighting the role of action boundaries in aligning shared expectations during joint action.

Together, the three studies form a multi-level investigation of predictive processing during action perception and coordination, centered on the functional role of action boundaries. **Study 1** establishes the neural mechanisms by which action boundaries structure prediction, retrieval, and updating during the observation of complex action repertoire. It identifies distinct brain systems involved in ending an action (U), initiating the next (T), and responding to violations (PE) of expected action trajectories. **Study 2** complements these neural findings by revealing the fine-grained oculomotor and autonomic signatures of prediction errors, demonstrating how internal model updating is reflected in moment-to-moment changes in gaze behavior and arousal at precisely those boundary moments. **Study 3** extends these principles from individual action observation to social interaction, showing how predictive hierarchies are flexibly coupled across interacting agents and across multiple levels of coordination, from movements to shared goals that are reflected in the distinct neural oscillations.

By integrating neural, behavioral, and interpersonal measures within a predictive processing framework, the three studies collectively demonstrate that action boundaries constitute critical hubs for prediction, error signaling, and model updating, both within individuals and between interacting partners. This program of research advances a unified account of how the brain anticipates, monitors, and adapts to unfolding action in both solitary and joint contexts, highlighting predictive processing as a fundamental mechanism supporting efficient perception, prediction, learning, and social coordination.

3. Research Articles

3.1. Study 1: Updating predictions in a complex repertoire of actions and its neural representation

Selvan, R. N., Cheng, M., Siestrup, S., Mecklenbrauck, F., Jainta, B., Pomp, J., Zahedi, A., Tamosiunaite, M., Wörgötter, F., & Schubotz, R. I. (2024). Updating predictions in a complex repertoire of actions and its neural representation.

NeuroImage, *296*, 120687.

https://doi.org/10.1016/j.neuroimage.2024.120687

Abstract

Even though actions we observe in everyday life seem to unfold in a continuous manner, they are automatically divided into meaningful chunks, that are single actions or segments, which provide information for the formation and updating of internal predictive models. Specifically, boundaries between actions constitute a hub for predictive processing since the prediction of the current action comes to an end and calls for updating of predictions for the next action. In the current study, we investigated neural processes which characterize such boundaries using a repertoire of complex action sequences with a predefined probabilistic structure. Action sequences consisted of actions that started with the hand touching an object (T) and ended with the hand releasing the object (U). These action boundaries were determined using an automatic computer vision algorithm. Participants trained all action sequences by imitating demo videos. Subsequently, they returned for an fMRI session during which the original action sequences were presented in addition to slightly modified versions thereof. Participants completed a post-fMRI memory test to assess the retention of original action sequences. The exchange of individual actions, and thus a violation of action prediction, resulted in increased activation of the action observation network and the anterior insula. At U events, marking the end of an action, increased brain activation in supplementary motor area, striatum, and lingual gyrus was indicative of the retrieval of the previously encoded action repertoire. As expected, brain activation at U events also reflected the predefined probabilistic branching structure of the action repertoire. At T events, marking the beginning of the next action, midline and hippocampal regions were recruited, reflecting the selected prediction of the unfolding action segment. In conclusion, our findings contribute to a better understanding of the various cerebral processes characterizing prediction during the observation of complex action repertoires.

Keywords: action observation, branching structure, action perception, action prediction hierarchy, internal models, violation of expectation

Introduction

When we observe actions, our brain generates predictions about how this complex, dynamic stimulus will evolve (M. Botvinick & Plaut, 2004; Colder, 2011; Csibra & Gergely, 2007; Kilner et al., 2004, 2007; Schiffer et al., 2013; Stadler et al., 2011). Expected and actual stimulations are compared, and only the difference between the two, the prediction error, is propagated to update future predictions about the probabilistic structure of our reality (Friston, 2005). It is generally assumed that these computations, as complex perception in general, are based on hierarchical predictive processing. According to this hierarchical framework, prediction errors are conveyed in a bottom-up manner from lower cortical areas to higher ones via forward connections, while predictions are delivered top-down by backward connections (Friston, 2010; Friston & Kiebel, 2009). As of today, we still know little about the number of layers and their computational characteristics in this postulated action prediction hierarchy, and in which cerebral networks they are specifically implemented.

One way to conceptualize action prediction is along a part-whole hierarchy, meaning that a goal-directed action sequence consists of multiple segments or action steps (Uithol et al., 2012). Such segmental action structures are not deterministic, since at boundaries between action segments, we can often choose between multiple plausible steps that can be performed next. Studies have shown that subjects can reliably detect segment boundaries (Newtson, 1973) and that the brain's event prediction is also oriented toward these segments (Zacks et al., 2011). This has been observed when watching movies or reading stories, but also when observing actions of others (Kurby & Zacks, 2008). The segment boundaries are of particular interest, because here the prediction comes to a point where a new segment is pending but not yet observable. Accordingly, the prediction error reaches a maximum (Zacks et al., 2011) and triggers prediction updating (Schubotz et al., 2012; Wahlheim et al., 2022; Zacks et al., 2007; Zacks & Sargent, 2010). At the same time, it is reasonable to assume that the predictive model holds several options of connectable segments, which now form the new expectation. These can be compared to the next observed action, resulting in an update of the currently valid prediction. In a recent study (Pomp et al., 2021), it became clear that specific action boundaries, namely start of hand-object contact (T, Touching) and end of hand-object contact (U, Untouching), are important features for the segmentation of observed actions. Additionally, we could not only confirm that prediction error signals peak at the segmental boundaries of actions but also disentangle prediction error and prediction updating as two temporally and neuronally distinct stages of processing (Pomp et al., 2021b).

In the current study, we built on these findings to investigate action prediction processes at boundaries between actions of a structured action repertoire. In a previous study (Pomp et al., 2021), we investigated updating at action boundaries within single, disconnected object manipulations. In contrast, everyday actions mostly consist of action sequences which are hierarchically structured in a way that some action boundaries, namely those with different connection options, involve a change in the level of the action hierarchy (Ondobaka & Bekkering, 2012). In addition, different connection options create a variable degree of uncertainty and predictability. In the present study, we took the view that in an ecologically valid action repertoire, updating should comprise two distinct stages: first, opening of the option space by retrieving all expectable (probable) options, and then, upon exploiting evidence from the ongoing stimulus, restricting to one of them as the most probable option. To investigate this, we generated an action repertoire with a well-defined segmental branching structure; in particular, actions could be followed by one, two, three, or four possible other actions. We videotaped an actress performing all 82 actions of this repertoire and determined action boundaries (i.e., segments) in these action videos using a computer vision-based model. Boundaries were defined as touching (T), i.e., the hand touching an object, and untouching (U), i.e., the hand releasing/ untouching an object. In the present stimulus material, T events defined the beginning of an action segment, whereas U events marked their ending (Aksoy et al., 2011; Wörgötter et al., 2013b; Ziaeetabar et al., 2021b).

Participants first practiced the action repertoire themselves before entering the MRI scanner. There, they were presented with videos of actions that corresponded to the trained repertoire, but also action sequences with a single action that did not match the learned action repertoire to induce prediction errors and, consequently, updating of event models. Specifically, we modeled the brain activation at hand-object T/U events (see Figure 4 for the representation) to disentangle the two suggested stages of model updating in suprasegmental actions: retrieval of all upcoming options at the end of the previous action (hand untouching last object) and selecting the option at the beginning of the next action (hand touching next object). Moreover, we modeled the brain activation time-locked to untouching at the event boundaries by the parametric degree of probabilistic structure (i.e., branching). Since we used branches of equal probability, the level of uncertainty was proportional to the number of branches, meaning that more possible branches led to higher uncertainty, and fewer possible branches resulted in lower uncertainty.

We expected the updating of predictions to activate areas upstream of the action observation network (AON), which is reliably engaged by viewing videos showing individual (separate) object manipulation actions. The AON has been suggested to feature an intrinsic hierarchical processing architecture (Sasaki et al., 2018; Urgen et al., 2019; Urgen & Saygin, 2020). In addition to the goal level represented in the AON, additional layers have been proposed that propagate and prioritize goals of actions, supposedly involving dorsolateral prefrontal cortex and ventromedial networks, respectively (Pezzulo et al., 2018). However, as these latter supra-goal layers have so far been proposed for predicting one's own actions, and, being related to control and motivation, they may not be relevant for passive observers. Therefore, it remains to be seen whether they are also used in predicting observed actions.

Another strong candidate for higher layers of the prediction hierarchy of observed actions is a network of the medial prefrontal cortex, posterior cingulate cortex/precuneus, and temporo-parietal junction. Depending on the research context of prediction hierarchy, this network has been referred to as the default mode network (DMN) whose computations are relevant to the most complex integrative functions of the brain (Pesquita et al., 2018). Studies where participants were asked to infer others’ intentions or beliefs from their actions found an increased DMN-AON connectivity (Koster-Hale & Saxe, 2013; Wurm et al., 2011). Moreover, recently proposed hierarchical prediction and prediction error processing models (Alexander & Brown, 2015, 2018) suggest that the medial prefrontal cortex (a DMN area) provides high-level predictive input to the lateral prefrontal cortex (an AON area). More generally, the DMN including the medial prefrontal cortex is involved in processing complex and hierarchically structured schemas (Baldassano et al., 2018; Masís-Obando et al., 2022; Reagh & Ranganath, 2023; Sommer et al., 2022). Yet, the specific role of these areas in potentially providing higher-level input to action observation, reflecting a repertoire-informed updating of predictions, awaits further testing.

Methods

Participants

Forty-five subjects (30 women, 15 men) between the age of 18 and 30 years (*M* = 23.92; *SD* = 4.32) were recruited for the study. Data from four participants were not included in the final analyses due to excessive movement (> 5 mm) during the fMRI session. Consequently, the final sample included 41 datasets. Participants had normal or corrected to normal vision with no color blindness and were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) with handedness scores varying between 50 and 100 (*M* = 87.02, *SD* = 15.31). All participants gave written informed consent to participate in this study and were compensated with money or course credits. The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee of the University of Münster.

Stimuli

As stimuli material, we used short video clips that showed multistep actions, presented in the third person perspective. Four objects, two cups with green bottom and two-colored cubes (red and blue; Figure 1), were used to execute action sequences, with predefined actions (e.g., put the cup upside down) and probabilistic transitions between them (see Figure 2 for all possible object configurations). Video clips of 41 action sequences were generated, hereafter termed *legal* sequences (Figure 3 A). Each video started with the same initial frame showing two hands of an actress wearing white gloves placed on a table and the four objects, each on a fixed table location. The action sequences varied in length, ranging from four to six action steps and six to twelve seconds in duration.

In addition to the 41 complete action sequences, the individual steps of each action sequence were also recorded separately for the imitation training. The table on which the actions took place was covered with a table cover printed with a QR code-like pattern known as the “AprilTag” markers (Figure 1). These markings are described in more detail below. The recorded videos were cropped to show only the hands of the actress with the objects (see section 2.3 for detailed information).

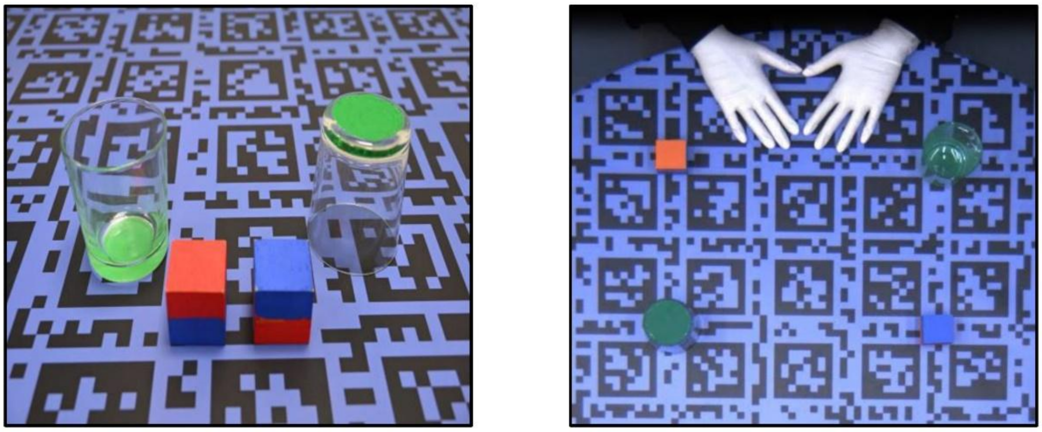

A. Objects used for manipulation B. Start screen in fMRI session

Fig. 1. A. The objects used for the stimuli preparation: two cups and two cubes; B. Starting frame of the videos, showing the actress from the 3pp and the starting distribution of all four objects on the table

Furthermore, another set of 41 action sequences (termed *illegal* hereafter) was recorded, each of which corresponded to a legal sequence, but a single action was modified (Figure 3 B). This modification consisted of doing a different manipulation with the same object as in the legal ones or doing a different manipulation with a different object. The modification happened in one of the actions of the entire action sequence, however, the first step always remained unchanged. Importantly, after the modification, action sequences continued the same way as the legal ones so that only one action differed between legal and illegal sequences. In the present study, illegal sequences were included to (1) allow for a discrimination task during the fMRI session (see section 2.5) and (2) to distinguish well-known prediction error effects due to mismatch of expectancies from those that accumulate at event boundaries, which were the main interest of this study.

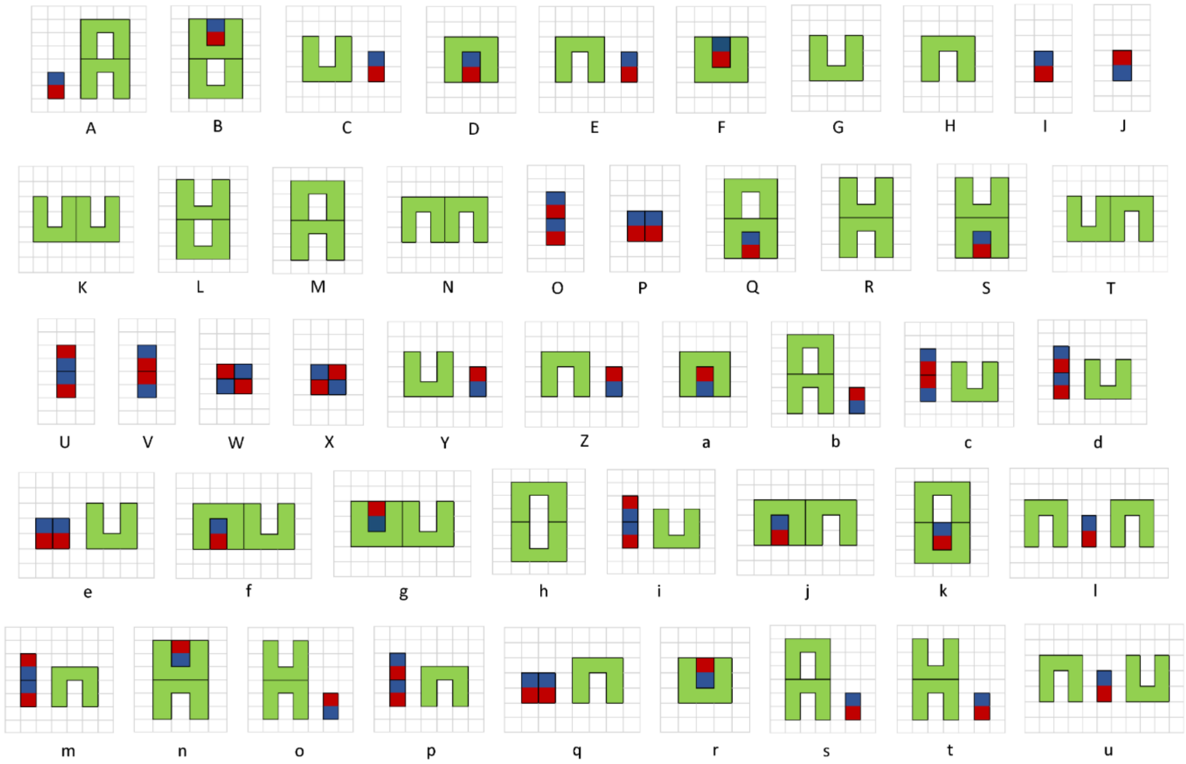


Fig. 2 Schema of the object configurations achieved by actions. Two cups and two cubes were manipulated by stacking up, placing next to, putting in, taking out, turning over, placing over, and taking down. The trained action repertoire consisted only of actions generating object configurations labeled by capital letters A to Z. Lower case letters show configurations only achieved by illegal actions. Note that some actions were also illegal when generating a legal object configuration but in the wrong action sequence

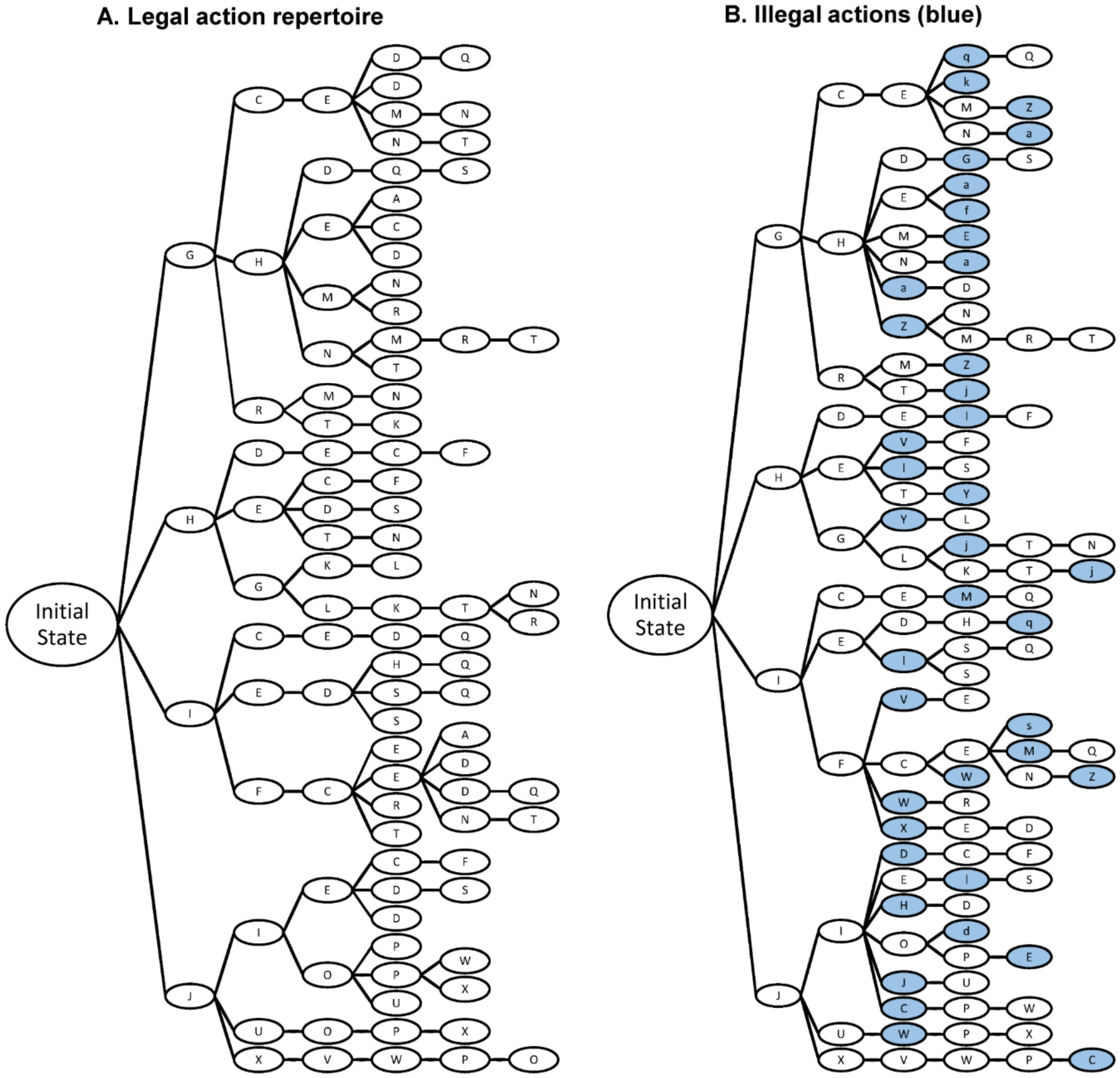


Fig. 3. A. Decision tree of the legal action repertoire; B. Illegal actions; actions that violate the learned repertoire are marked in blue. Letters indicate different actions as defined by their outcome, i.e., a certain configuration of objects (for a detailed description of object configurations, refer to Figure 3). The branching level can be recognized at the transitions between the individual actions; for example, it is three for action D after the sequence I-E-D, but four for action C after the sequence I-F-C.

Action Segmentation and T/U Determination

For the creation of stimulus material and the subsequent segmentation of actions based on touching (T) and untouching (U) events, a computer vision system was used. T was defined as the point in time when the actor’s hand made contact with one of the objects, U was when the actor’s hand released the object. T and U marked the beginning and the end of an action step, respectively (Figure 4).

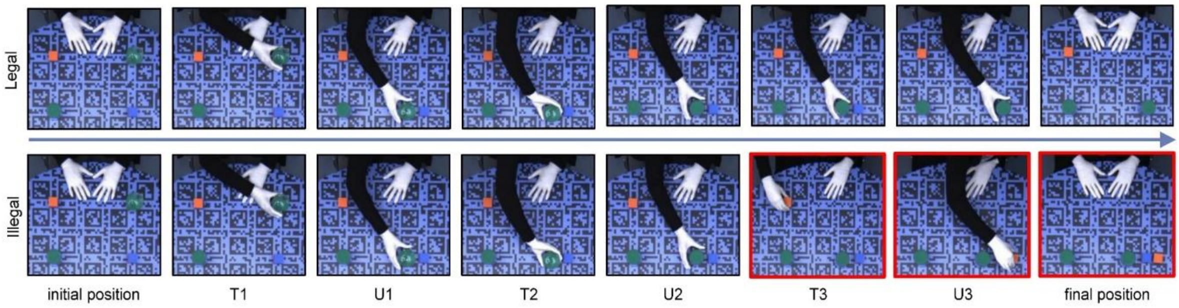


Fig. 4. An example of a legal and the corresponding illegal action sequence with all Touching (T) and Untouching (U) events showing the moment of modification at T3. In the legal action sequence, at T3, the cup is taken to be placed over the blue-side-up cube next to it. In the illegal action sequence, instead of the cup, the red-side-up cube is grasped.

The computer vision system consists of four parts: (1) The multi-camera setup, (2) the touching events detector, (3) the eye-tracker and the corresponding transformation system and (4) the synchronization scheme. These parts will be described in detail in the following.

To record the visual content and track the hand and objects in the video, a five-camera setup was installed around the experiment table. The camera used here was Grasshopper GS3-U3-32S4C-C from Teledyne FLIR (resolution 1024x768; 60.0 frames per second (fps); color space: RGB8; synchronization mode: hardware synchronization via cables). The camera setup was calibrated using a multi-camera calibration tool (Michels, 2022). Images captured from the cameras were saved as video files which consequently went through a pipeline for detection and tracking using YoloV5 (Jocher, 2020). In addition, the “Apriltag” markers (Olson, 2011) were employed to calculate both the eye tracker’s pose (extrinsic parameters of the camera) and the transformation matrix between the eye tracker and the five-camera setup. A tablecloth, with 7 by 7 “Apriltag36h11” markers printed on it, was fixed on the experiment table. Hands and objects were tracked by the multi-camera setup. During the stimuli preparation, eye movements of the actress were also tracked, but these data were not relevant to the current study and are therefore not mentioned further below.

For T/U event detection, a touching sensor was attached to the right index finger. The touching sensor consisted of a force-sensing resistor, a bias resistor with a 5-volt power supply, an OPAMP for buffering and an Arduino Uno for A/D conversion with a sampling rate of 1KHz and resolution of 8-bit. The sampled signal was compared with a threshold value of 127 (half of the max 8-bit binary number) and then denoised with manual human help to exclude false touching and untouching events, such as accidentally touching the table. Importantly, T and U events that were analyzed in the present study were restricted to the touching and untouching of the experimental objects (cups and cubes). Notably, due to the multistep nature of action sequences, untouching events signal transitions between actions and are, therefore, indicative of event boundaries.

Training Procedure

The experimental protocol included four consecutive days. On the first two days, participants went through imitation training sessions. The imitation training consisted of three phases for each action sequence. During the first phase, participants were asked to observe the action sequence presented to them on a monitor (27 inches) using Presentation software 20.3 (Neurobehavioral Systems Inc., Berkley, CA, USA). In the second phase, participants imitated the actions of action sequences while the recorded videos were presented to them. They were instructed to imitate with the same speed and hand movements. In the last phase, participants imitated the complete action sequences along with the video. The distance between the seated participant and the screen measured approximately 80 cm. The same table cover and objects that were used for stimuli preparation were employed for the imitation training as well.

The presentation of the 41 legal videos was organized in four blocks. Videos were assigned to a block based on the object that was manipulated first during the action sequence (upright cup, inverted cup, blue-side-up cube, and red-side-up cube) independent of its concrete manipulation. Within each block, videos were presented in the order of the length of the sequences of four, five, and six steps action sequences. Each training session included a short pause after two blocks, and the whole session lasted for about 75 to 90 minutes. During imitation training, participants were observed by an experimenter, and if the participants made a mistake, they were asked to imitate the action sequence again. The same procedure was followed on the second day of imitation training.

fMRI Session

During the fMRI session, on the third day, participants were presented with the legal and illegal action sequences in randomized order. The session consisted of three blocks, with all 41 legal action sequences presented in three blocks. Forty-one illegal action sequences were distributed across blocks with 13 presented in the first block and 14 each in the second and third block. Each block always started with a legal action sequence. An illegal sequence and its corresponding legal one were never presented consecutively and not more than three illegal sequences were presented in a row. After each block, there was a short break of approximately two to four minutes for eye-tracking calibration (see section 2.7). Additionally, after the presentation of every action sequence, there was a question trial during which participants responded by pressing a button to determine whether they had seen the specific action sequence during the imitation training. The question trial lasted till a button was pressed or a maximum of two seconds (Figure 5 A). Moreover, there were 15 null events, five in each block, during which only a fixation cross was presented for 1500, 2000, or 2500 ms. The entire experiment consisted of 123 trials of legal and 41 trials of illegal action sequences, 164 question trials, and 15 null events, and lasted for about 60 minutes, with each block 16 to 20 minutes long.

Post-fMRI memory test

One day after the fMRI session, participants performed a post-fMRI memory test on the legal sequences that they had observed and imitated during the first two days of the experiment. The task comprised all the 41 legal sequences presented on the monitor step by step. The video of the action sequence paused after each action and two options (one correct and one incorrect) appeared on the screen. The options, those are the possible next individual actions, were presented as videos one after the other, first on the left of the screen and then on the right. The distribution of the correct options was balanced between the right and left side of presentation. The incorrect options consisted of some of the illegal events of the fMRI session and some additionally new illegal events that had not been presented before during the imitation training sessions. The participants were instructed to decide which of the two presented actions would be the correct next action and give the response using two keys on their keyboard. The response time window was two seconds after the presentation of second video option (Figure 5 B). Once the participant gave a response or timed out, the video continued to play from where it paused and thus provided real-time feedback on the response.

At the end of each action sequence, the participants were asked to rate how sure they had been to predict the upcoming steps on a Likert scale of 0 to 3 (0 – no idea, 1 – rather a guess, 2 – rather know, 3 – know for sure) and there was no time-out for this answer. The session lasted for approximately 35 to 45 minutes.

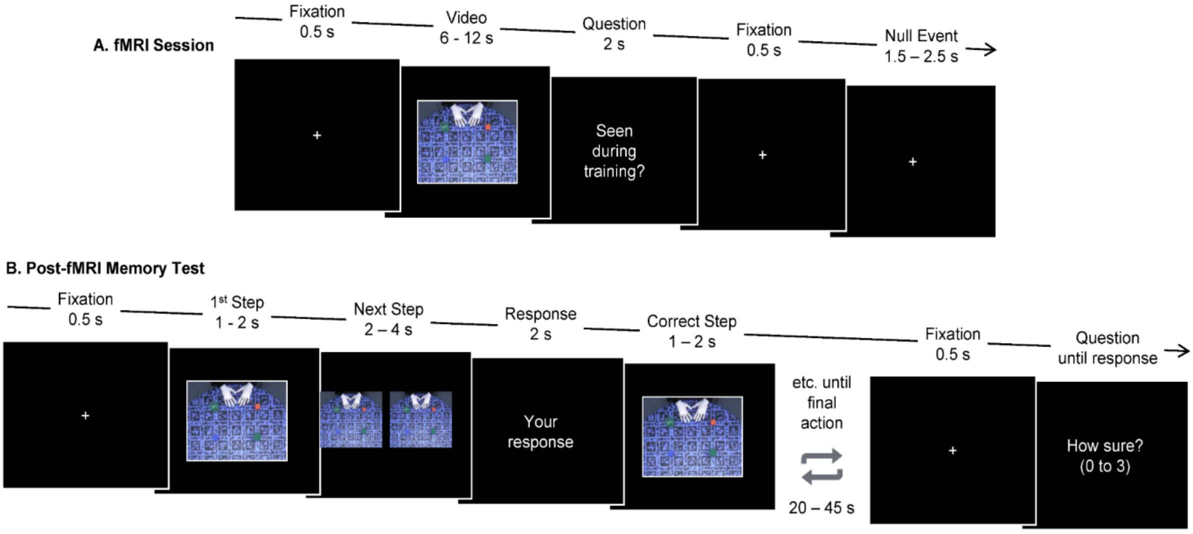


Fig. 5. Schema of the experimental paradigm A. Trial schema for the fMRI session with legal and illegal action sequences B. Trial schema for the post-fMRI memory test.

MRI Data Acquisition

Neuroimaging data were acquired using a 3-Tesla Siemens Magnetom Prisma MR tomograph with a 20-channel head coil. Participants lay supine on the scanner bed and were instructed to remain still. Additionally, the fixation of the head, arms, and hands with form-fitting cushions minimized movements. During fMRI, we also recorded eye-tracking data using the Eyelink 1000 Plus eye tracker, which is not the focus of this manuscript and will not be reported further. Participants were provided with a response box with the right index and middle finger positioned on two buttons. Additionally, participants wore headphones and earplugs to attenuate scanner noise. Stimuli were presented using Presentation software 20.3 and participants watched the videos on the screen in 640 x 512 pixels resolution through a mirror mounted on the head coil. The structural T1 weighted images were acquired using a 3-D magnetization prepared rapid gradient echo sequence (MPRAGE) sequence with 192 slices, repetition time (TR) = 2130 ms, echo time (TE) = 2.28 ms, slice thickness = 1 mm, field of view (FoV) = 256 x 256 mm^2 and flip angle = 8°. Following that, the functional data were obtained in an interleaved order, with the scanning parameters: 33 slices, TR = 2000 ms, TE = 30 ms, slice thickness = 3 mm, slice spacing = 1 mm, FoV = 192 x 192 mm^2, flip angle = 90°.

fMRI Data Preprocessing

Structural and functional images were processed using Statistical Parametric Mapping 12 (SPM12; The Wellcome Centre for Human Neuroimaging, London, UK) implemented in MATLAB (MathWorks R2022a). Functional images were slice time corrected to the middle slice, then movement correction was applied, and images were realigned to the mean image. The structural image was co-registered with the mean functional image. Structural and functional images were then normalized to MNI space (Montreal Neurological Institute, Montreal, QC, Canada). Spatial smoothing was performed using a Gaussian kernel of Full-Width at Half Maximum (FWHM) 8x8x8 mm^3, and a temporal high-pass filter of 128 s was applied.

fMRI Design Specification

Statistical analyses were performed using SPM12. We used a general linear model (GLM) for serially autocorrelated observations (Friston et al., 1995; Worsley & Friston, 1995) and convolved regressors with the canonical hemodynamic response function. On the first level of the analysis, we used a grey matter mask that was comprised of smoothed individual normalized grey matter images (8x8x8 mm^3 FWHM) with a threshold of 0.2 implemented in ImCalc in SPM12.

A total of 12 regressors were included in the GLM. There were four regressors for the experimental conditions, namely one regressor for the legal action sequences (Legal), one for the illegal action sequences (Illegal), one for hand-object touching events (T), and another one for the hand-object untouching events (U). The fMRI experiment comprised three blocks of trials. Per block, all 41 legal action sequences were presented once, summing up to 123 legal action trials in total. The 13 illegal action sequences were presented in the first block, 14 in the second, and 14 in the third block. Out of 123 trials of legal action sequences presented across three blocks, 14 from the first block, 13 from the second, and 13 from the third that corresponded to the illegal action sequences were selected for the GLM. This was done to balance the number trials in each regressors, namely, legal and illegal action sequences. Regressors Legal and Illegal included 41 trials each, while there were 450 trials each for T and U events. Additionally, two regressors for question events (164 trials) and null events (15 trials) were added. The onsets of the illegal action sequences were the point of expectation (or action repertoire) violation, and corresponding time points were determined in the legal sequences. The activity was modeled from the point of expectation violation to the end of the action sequences, T and U of the legal action sequences were modeled as events. Furthermore, a parametric regressor for U events with the number of branches (1, 2, 3, or 4) in decreasing order (i.e., the lowest number for the highest number of branches) was modeled. For the question trials and null events, the onsets were the beginning of each trial and modeled with a duration of zero. All parametric modulators were mean-centered (Mumford, Poline, & Poldrack, 2015). Finally, the six subject-specific rigid-body transformations obtained from realignment (three translations and three rotations) were included as regressors of nuisance.

Whole Brain fMRI Analysis

At the first level, the following t-contrasts were computed: (a) To test the effect of violation of expectation, a contrast between Illegal and Legal events of the action sequences *(Illegal > Legal)* was calculated; (b) To investigate the brain activity involved in the retrieval of information from the internal model formed during the imitation training, the direct contrast *(U > T*) was calculated; (c) To examine the role of the probabilistic structure of the actions in legal action sequences, parametric modeling of number of branches was performed *(Branching)*; (d) To investigate neural changes during the update of the internal model when an action is initiated, the direct contrast *(T > U)* was generated.

For the second level analysis, one sample *t*-tests were performed across participants. A false discovery rate (FDR) threshold of $p < 0.05$ or lower was applied. For the FDR of $p < 0.05$, the results are reported with an additional moderate cluster threshold of 20 voxels. However, when there was no significant activation at FDR of $p < 0.05$, we inspected *t*-maps at $p < 0.001$ (uncorrected) for completeness. Brain activation was visualized with MRIcroGL (version 1.2.20211006, McCausland Center for Brain Imaging, University of South Carolina, USA).

Behavioral Data Analysis

Accuracy and reaction time (RT) data from the fMRI session and the post-fMRI memory test were analyzed using generalized linear modeling conducted in the R programming language (R Core Team, 2019; Version 4.2.1). For further RT analyses, only the trials with correct responses and RTs above 100 milliseconds were included. As the empirical distributions of RTs and accuracy rates were not normal (assessed by the Shapiro-Wilk-Test and the Cullen and Frey graph), for analyses of RTs and accuracy, a log-normal family and a binomial family were used, respectively. Notably, as the output is not normally distributed, it is not possible to use ANOVA or an equivalent general linear model (GLM) using a Gaussian family. For the behavioral results of the fMRI session, the GLM model (Eq. 1) contained the STIMULUS TYPE as a fixed effect; further, two random intercepts were assumed for subjects and trials. For the post-fMRI memory test, the GLM (Eq. 2) consisted of the STIMULUS HISTORY of the to-be-rejected option (Legal, Illegal in the fMRI session) and BRANCHING STRUCTURE (1, 2, 3, 4) as the fixed effects. For the post-hoc analyses, p-values are adjusted using Tukey's correction for multiple comparisons.

fMRI Session: $Outcome \sim Stimulus\ Type + (1|ID) + (1|Trial)$ (Eq. 1)

Post-fMRI memory test: $Outcome \sim Branching\ Structure + Stimulus\ History + (1|ID) + (1|Trial)$ (Eq. 2)

Results

Behavioral Results

fMRI Session

During the fMRI session, participants made 76.5% correct responses, 20.02% incorrect responses, and 3.48% of responses were missed due to time out. The main effects of the STIMULUS TYPE with respect to accuracy $\chi2(1) = 5.6824, p = .0171,\ Cramer's\ V = 0.029$ and RTs $\chi2(1) = 80.07, p < .001, Cramer's\ V = 0.125$ were significant. Tukey-corrected post hoc contrasts showed that the accuracy was significantly higher for illegal events than the legal events $(z.ratio = 2.384, p = .0171)$ and RT for legal events was significantly longer than that of the illegal events $(z.ratio = 8.948, p < .001)$.

Post-fMRI memory test

In the post-fMRI memory test, participants' responses were correct in 89.31% of trials, while 8.33% of incorrect responses were recorded, and 2.36% of responses were missed due to time out. There was no main effect of STIMULUS HISTORY, meaning that actions which were part of other legal sequences and actions which were not derived from legal sequences were rejected with the same probability and equally fast.

The data from the post-fMRI memory test showed that there was a significant main effect of BRANCHING STRUCTURE observed in accuracy $\chi2(3) = 41.591\ , p < .001, Cramer's\ V\ = 0.055$ and RTs $\chi2(3) = 36.05\ , p < .001, Cramer's\ V = 0.052$. Post hoc contrasts showed that the accuracy of one branch (95.7% ± 0.203) was significantly higher compared to two banches (90% ± 0.3) $(z.ratio = 3.332, p = .005)$ and three branches (85.96% ± 0.347) $(z.ratio = 6.205, p < .001)$. There were high accuracy for four branches (93.87% ± 0.24) than three branches (85.96% ± 0.347) $(z.ratio = -3.265, p =\ .006)$.

Further, regarding the RTs, the post hoc tests showed that actions with one branch (M = 429 ± 265.52 ms) were significantly shorter than for those with three branches (M = 442 ± 292.96 ms) $(z.ratio\ = -2.734, p = .031)$, and longer than for actions with four branches (M = 406.5 ± 245.08 ms) $(z.ratio = 3.116,\ p = .009)$. Additionally, the RT for actions with two branches (M = 413 ± 240.84 ms) were significantly shorter than for those with three branches (M = 442 ± 292.96 ms) $(z.ratio = 1.658, p = .005)$ and RTs for actions with three branches (M = 442 ± 292.96 ms) were significantly longer than for those with four branches (M = 406.5 ± 245.08 ms) $(\ z.ratio = 5.654, p < .001)$. The actions with three branches had significantly longer RT than one, two and four branches and the actions with one branch had longer RT than four branches.

fMRI Results

To investigate the effect of violation of expectation generated by the mismatch between predicted and perceived actions, the contrast between illegal and legal events of the action sequences *(Illegal > Legal)* was calculated. The contrast revealed activation in the bilateral supplementary motor area (SMA), supramarginal gyrus (SMG), and middle anterior insula (INS). In the right hemisphere, there was significant activation in the posterior superior temporal sulcus (pSTS) and thalamus (THA) (Figure 6 A, Table 1).

To determine the effects of the event boundary U of the legal action sequences, the first step of the two-step predictive process, where up to four possible courses of action could be expected depending on branching, a direct contrast *(U > T)* was calculated. This contrast yielded bilateral activation in SMA, INS/PUT, and lingual gyrus (LIN) (Figure 6 B, Table 2). The parametric modulator (Branching) at U events, reflecting the effects of the probabilistic branching structure of the action repertoire, revealed activation in the left fusiform gyrus (FFG), bilateral INS, and right anterior cingulate cortex (ACC) (Figure 6 C, Table 2). However, it is important to note that activation found for this effect did not survive correction for multiple comparisons and is therefore described at $p < .001$, uncorrected.

Additionally, to investigate the role of event boundary T of the legal action sequences, the time point when the action begins and which allows the restriction of the prediction to now only one option, the contrast $(T > U)$ was calculated. This contrast showed significant activation in the left caudate (CAU), medial prefrontal cortex (MPFC), posterior cingulate cortex (PCC), and hippocampus (HIP), as well as the bilateral medial cingulate cortex (MCC) (Figure 6 D, Table 2).

As a control, we explored the effect of event boundaries U and T of the illegal action sequences. To this end, two contrasts $(T_i > U_i)$ and $(U_i > T_i)$ were calculated in a second GLM model. The $(U_i > T_i)$ showed activation of the bilateral superior temporal gyrus (STG) and bilateral lingual gyrus (LIN). Furthermore, the contrast $(T_i > U_i)$ did not survive multiple comparison corrections. However, the event boundaries, U and T, between the legal and illegal sequences could not be compared with the caveat of a highly unequal number of trials for the latter.

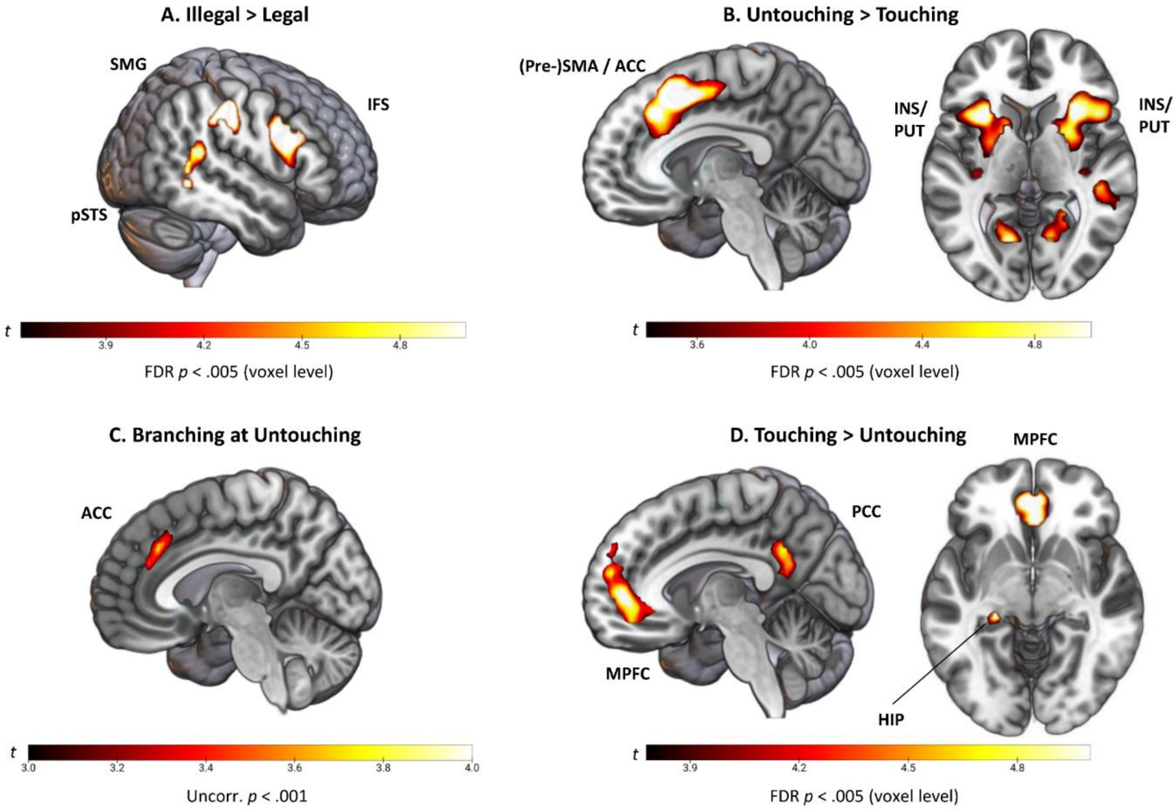


Fig. 6. Brain Activation A. Effect of Violation of Expectation – IFS, inferior frontal sulcus; SMG, supramarginal gyrus; pSTS, posterior superior temporal sulcus; B. Effect of Untouching events – Pre - SMA, Pre - supplementary motor area; INS/PUT, insula/putamen; C. Effect of Branching – ACC, anterior cingulate cortex; D. Effect of Touching events – PCC, posterior cingulate cortex; MPFC, medial prefrontal cortex; HIP, hippocampus.

Table 1: Whole-brain activation for contrasts of *Illegal* > *Legal* action sequences at FDR $p < .005$

Anatomy	**H**	**Cluster extent (voxels)**	**MNI Coordinates**			**t Value**
			x {mm}	**y {mm}**	**z {mm}**	
Pre-Supplementary Motor Area	L	50	-3	17	53	5.14
	R	35	3	20	50	5.46
Supramarginal Gyrus/ anterior Inferior Parietal	L	103	-42	-31	41	7.39

Sulcus						
	R	265	54	-22	41	9.80
Anterior Dorsal Insula	L	75	-33	20	-4	6.26
	R	142	33	23	-1	8.87
Inferior Precentral Sulcus	L	124	-54	5	35	6.16
	R	562	51	11	29	7.80
Inferior Frontal Sulcus	R	l.m.	48	23	23	7.31
Inferior Temporal Gyrus	R	216	48	-55	-7	8.90
Posterior Superior Temporal Sulcus	R	l.m.	54	-40	14	5.71
Thalamus	R	39	9	-16	8	4.59

Only clusters with a minimum extent of 20 voxels are reported. H = Hemisphere; MNI = Montreal Neurological Institute; L = Left; R = Right; l.m. = local maximum.

Table 2: Whole-brain activation for contrasts of Event boundaries *(Touching & Untouching)* and parametric modulator *Branching* at FDR $p < .005$/ uncorrected $p < .001$.

Anatomy	**H**	**Cluster extent (voxels)**	**MNI Coordinates**			**t Value**
			x {mm}	**y {mm}**	**z {mm}**	
Untouching > Touching (FDR at $p < .005$)						
(Pre-)Supplementary Motor Area/ Anterior Cingulate Cortex	L	1871	-3	14	50	7.79
	R	l.m.	12	8	62	7.86
Insula/ Putamen	L	414	-33	23	-4	6.23
	R	76	36	-22	8	4.78
Lingual Gyrus	L	117	-15	-67	-1	6.04
	R	218	24	-49	-1	5.63
Posterior Inferior Temporal Sulcus	R	133	54	-49	-7	5.71
Posterior Superior Temporal Sulcus	R	l.m.	48	-31	-1	5.08

Primary Motor Cortex	L	54	-36	-22	53	5.70
Anterior Middle Frontal Gyrus	L	36	-33	47	17	4.51
Branching (uncorrected at $p < .001$)						
Fusiform Gyrus	L	18	-30	-61	-10	4.99
Insula	L	29	-30	17	-10	4.04
	R	52	42	20	-7	4.76
Anterior Inferior Parietal Sulcus	R	59	48	-25	44	4.21
Middle Frontal Gyrus	R	11	39	50	20	4.00
Anterior Cingulate Cortex	R	40	6	38	23	3.97
Touching > Untouching (FDR at $p < .005$)						
Medial Prefrontal Cortex/ BA10/ Pregenual Anterior Cingulate Cortex	L	598	-3	50	-1	7.33
Dorsal Premotor Cortex	L	80	-24	-7	56	5.68
	R	121	30	-10	59	5.85
Mid Cingulate Sulcus	L	22	-15	-22	44	5.44
	R	23	9	-19	47	4.93
Posterior Cingulate Cortex	L	108	-9	-52	26	6.12
Fusiform Gyrus	L	41	-30	-55	-19	5.54
Hippocampus	L	25	-24	-28	-7	6.09

Only clusters with a minimum extent of 20 voxels for FDR at $p < .005$ and 10 voxels for uncorrected at $p < .001$ are reported. BA = Brodmann Area; H = Hemisphere; MNI = Montreal Neurological Institute; L = Left; R = Right; l.m. = local maximum.

Discussion

Although actions appear to be a continuous stream of activities, they consist of meaningful segments. Studies have shown that the segmentation of actions takes place automatically and plays an important role in learning and memory through the formation of internal models (Eisenberg et al., 2018; Kurby & Zacks, 2008; Richmond & Zacks, 2018). Specifically, the boundaries where one action comes to an end and the next action begins are the essential time points for predictive processing (Schubotz et al., 2012; Swallow et al., 2009). The current study aimed to better understand the processes that happen at these boundaries, using a probabilistic branching structure of actions in a defined action repertoire. The results confirm the assumption of a two-stage updating process, which is also reflected at the neurofunctional level. These two updating steps involve, first, a set of frontal areas that seem to retrieve learned options of upcoming actions, and second, midline and hippocampal regions that reflect the finally selected episodic prediction of the emerging action. We also find that this updating is distinct at the neurofunctional level from, and likely upstream of, the response to a prediction error, caused by an unexpected action at the single action segment level that informs the so-called action observation network. Our findings contribute to a better understanding of action prediction in the context of complex action repertoires that are typical of our everyday lives.

The present study was designed to investigate prediction-updating processes at action boundaries by differentiating U and T events and examining the effects of the statistical properties of the action repertoire. In the first step, however, it was important to present the AON, which is well-reported to be underlying the processing of observed actions as a sanity check. To this end, we modified a single action in some of the presented action sequences, hypothesizing that such modifications induce a prediction error (Exton-McGuinness et al., 2015; Fernández et al., 2016; Jainta et al., 2022; Siestrup et al., 2023; Sinclair & Barense, 2019). In this case, the prediction had to quickly switch from complex expectation models of the suprasegmental structure to simpler expectation models that only call up the set of possible manipulations. As expected, these prediction errors at the level of single actions induced a BOLD response in the AON (Caspers et al., 2010). In addition, activation of the bilateral anterior insula increased in response to altered actions, indicating that participants recognized the modification as an action error (Bossaerts, 2018; Klein et al., 2007, 2013; Ullsperger et al., 2010). This neurophysiological response to error detection also corresponded to the significantly faster classification times for sequences with an altered action. It should be noted here that the prediction error only occurred at the level of a single action segment, while the sequence remained repertoire-compliant up to this point and thereafter. Consequently, no structures further upstream were involved in the processing of the prediction error.

At event boundaries, prediction errors accumulate (Kurby & Zacks, 2008), since the current event model no longer allows for precise prediction, especially when several next actions are possible. In the present study, this situation was implemented by U events in our multistep actions, where the learned action repertoire had to be accessed to form expectations with regard to the next action. At these U events, we found increased activation in SMA, lingual gyrus, putamen, and caudate nucleus. The striatum and the SMA are known to be involved in arranging sequences of action steps in the order in which they are to be executed (Bednark et al., 2015; Graybiel, 1998; Shima & Tanji, 2000; Tanji, 2001). The striatum plays a key role in the coding of action sequences; in particular, it serves to integrate individual actions into coherent, organized behavioral units, a process known as chunking (Favila et al., 2023; Graybiel, 1998). Studies in animals suggest that frontal input to the striatum, in particular the dorsomedial prefrontal cortex including SMA, shapes the representation of action sequences (Ostlund et al., 2009). Interestingly, rats with SMA lesions were still able to learn action sequences but failed when sequences had to be re-organized. Their behavior suggested that they represented the elements of the sequence as distinct behavioral units, corroborating a critical role of the SMA in sequence-level representations. Although the execution of highly trained action sequences probably depends more on the striatum than on the SMA, both areas seem to remain relevant, especially for the initiation of action chunks (Favila et al., 2023). Against the background of these findings, the joint activation of SMA and striatum at the beginning of each action boundary is highly plausible.

Moreover, increased activation detected in the lingual gyrus replicates previous findings where we investigated brain activity at boundaries within single-segment action (Pomp et al., 2021b). At this point in time, there is an increased need to visually evaluate the current positions of the objects (i.e., the local scene), to employ this information during the prediction of the subsequent action in a specific action sequence (Kamps et al., 2016; Ruotolo et al., 2019; Sulpizio et al., 2013a). More generally, increased exploratory vision and visual gain (Shipp, 2016) reflect processes that are typically associated with a prediction error in terms of the predictive coding framework.

In summary, these findings suggest that at the end of an action, when the object is released (U) to enable the initiation of the next action (T), the minimum of predictability is reached and counteracted by the retrieval of previously encoded event models, while at the same time, an intensive use of visual information already begins.

As described above, U events constituted the first of a two-step predictive process in the viewing of multistep actions. At these event boundaries, the brain uses the before-trained probabilistic structure of actions to update predictions of upcoming actions. Our rationale was that this process should depend on the number of possible following actions, as operationalized as the number of branches in the present experiment. The number of branches ranged from one to four, while a situation with only one possible branch represents the lowest level of uncertainty or strongest prediction. With regard to the BOLD effects, the parametric analysis of the branching structure suggested that as the strength of the prediction increased, activity increased in the ACC and the anterior insula. Since this contrast did not survive the correction for multiple comparisons, we refrain from discussing this observation in detail. Also, although the accuracy and RT measures from the post-fMRI test were significantly influenced by the branching structure, the directions of these effects were rather mixed. However, we would like to point out that this observation fits very well with the previous findings that ACC was more active when the internal model was biased towards a specific prediction (Akam et al., 2021; Kennerley et al., 2006; Klein-Flügge et al., 2022; Schiffer et al., 2013) and when participants were able to make predictions on the basis of previously trained sequences compared to observing completely new ones (Jainta et al., 2022; Siestrup et al., 2022, 2023; Siestrup & Schubotz, 2023). Thus, in the present study, higher ACC activation for more restricted expectations might reflect a more straightforward retrieval of available predictive models at U events.

After retrieving information from the internal model at the end of an action (U event), new information that supplies predictive processes are provided when a new action is initiated upon touching the next object (T event). In the case of our task, we assume that the expected options retrieved at U are matched with the now observed hand movements of the actress, leading to a final selection of the best matching option, which is broadcast as the newly expected action. This second step in the prediction updating process was characterized by increased BOLD responses in cortical midline structures MPFC and PCC, fitting our hypothesis regarding the role of DMN areas in updating predictions at the suprasegmental levels of action organization. Notably, the observed brain activation could be related to comparing own predictive models with the unfolding sequence at T events to aid further prediction (Koster-Hale & Saxe, 2013; Wurm et al., 2011). Moreover, recently proposed hierarchical prediction and prediction error processing models (Alexander & Brown, 2015, 2018) suggest that the MPFC provides high-level predictive input to the lateral frontal cortex, thus, DMN areas may provide top-down modulatory input to the AON to update and shape predictions of upcoming object transport and manipulation. Future studies should test this further by addressing causal network interactions using for instance dynamic causal modeling (Stephan et al., 2010). Interestingly, we also found hippocampal activation at T events. The hippocampus is known to be sensitive to event boundaries (Ben-Yakov & Henson, 2018; R.

A. Cooper & Ritchey, 2020). It has an important role in the generation of internal models and predictions (Barron et al., 2020), and operates as a match/mismatch detector between situational input and predictions (Duncan et al., 2009, 2012; Sinclair et al., 2021) Thus, hippocampal activation may reflect the comparison of the previously formed internal model and new information concerning the relevant object.

In summary, these findings suggest that at the beginning of the next action (T events), predictions that have been prepared during finalizing the previous action (U events) are now evaluated in reference to the unfolding observed action to exclude all but one prediction, involving a higher-level network of cortical midline and hippocampal areas.

Conclusion

Actions we observe in our daily life are organized in sequences of individual segments, and this segmental structure is highly relevant for prediction processes. In the present study, we showed that at U events, when the current event model comes to an end, the brain must retrieve previously formed internal models to generate predictions about the upcoming events. The observed neural responses also reflect the level of prediction strength. When new information about the continuation of an action sequence becomes available at T events, the brain evaluates the match between selected predictions and evidence from the observed action. In summary, this study provides valuable new insights into the stepwise predictive processes during the observation of complex manual action repertoires.

Credit authorship contribution statement

Rosari Naveena Selvan: Methodology, Software, Validation, Formal Analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Minghao Cheng:** Methodology, Software, Data curation, Writing – original draft. **Sophie Siestrup:** Software, Resources, Writing – original draft, Writing – review & editing. **Falko Mecklenbrauck:** Methodology, Software. **Benjamin Jainta:** Formal Analysis, Software, Resources. **Jennifer Pomp:** Validation, Visualization, Resources. **Anoushiravan Zahedi:** Formal Analysis, Writing – review & editing. **Minija Tamosiunaite:** Conceptualization, Methodology, Writing – review & editing. **Florentin Wörgötter:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Validation, Funding acquisition. **Ricarda I. Schubotz:**

Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Visualization, Validation, Supervision, Funding acquisition.

Declaration of Interest

The authors declare no competing interest.

Funding

This work was supported by the German Research Foundation (DFG) [WO 388/13-1]. The funder had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report.

Data and Code Availability

Unthresholded statistical maps of all reported and visualized fMRI contrasts in the manuscript have been deposited on NeuroVault (https://neurovault.org/collections/CGYMMSBZ/). The entire stimulus material will be made available via the Action Video Corpus Muenster (AVICOM).

The raw fMRI data and the raw SEC time point extraction data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

We would like to thank Monika Mertens, Christin Schwarzer, Lena Puder, Maria Leonie Hofmann, Simon Wieczorek, Kristin Stroop, Lars Schlüter, Leon Exeler and Niklas Dielitzsch for helping to recruit participants and/or assisting during data collection for the pilot and final studies.

References

Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R. F., Dayan, P., & Costa, R. M. (2021). The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection. *Neuron*, *109*, 149–163. https://doi.org/10.1016/j.neuron.2020.10.013

Aksoy, E. E., Abramov, A., Dörr, J., Ning, K., Dellen, B., & Wörgötter, F. (2011). Learning the semantics of object-action relations by observation. *International Journal of Robotics Research*, *30*(10), 1229–1249. https://doi.org/10.1177/0278364911410459

Alexander, W. H., & Brown, J. W. (2015). Hierarchical Error Representation: A Computational Model of Anterior Cingulate and Dorsolateral Prefrontal Cortex. *Neural Computation*, *27*, 2354–2410. https://doi.org/doi:10.1162/NECO_a_00779

Alexander, W. H., & Brown, J. W. (2018). Frontal cortex function as derived from hierarchical predictive coding. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-21407-9

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *Journal of Neuroscience*, *38*(45), 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. *Progress in Neurobiology*, *192*, 101821. https://doi.org/10.1016/j.pneurobio.2020.101821

Bednark, J. G., Campbell, M. E. J., & Cunnington, R. (2015). Basal ganglia and cortical networks for sequential ordering and rhythm of complex movements. *Frontiers in Human Neuroscience*, *9*(JULY), 1–13. https://doi.org/10.3389/fnhum.2015.00421

Ben-Yakov, A., & Henson, R. N. (2018). The hippocampal film editor: Sensitivity and specificity to event boundaries in continuous experience. *Journal of Neuroscience*, *38*(47), 10057–10068. https://doi.org/10.1523/JNEUROSCI.0524-18.2018

Bossaerts, P. (2018). Formalizing the function of anterior insula in rapid adaptation. *Frontiers in Integrative Neuroscience*, *12*, 1–9. https://doi.org/10.3389/fnint.2018.00061

Botvinick, M., & Plaut, D. C. (2004). Doing Without Schema Hierarchies: A Recurrent Connectionist Approach to Normal and Impaired Routine Sequential Action. *Psychological Review*, *111*(2), 395–429. https://doi.org/10.1037/0033-295X.111.2.395

Caspers, S., Zillers, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–1167. https://doi.org/10.1016/j.neuroimage.2009.12.112.ALE

Colder, B. (2011). Emulation as an integrating principle for cognition. *Frontiers in Human Neuroscience*, *5*(54), 1–12. https://doi.org/10.3389/fnhum.2011.00054

Cooper, R. A., & Ritchey, M. (2020). Progression from feature-specific brain activity to hippocampal binding during episodic encoding. *Journal of Neuroscience*, *40*(8), 1701–1709. https://doi.org/10.1523/JNEUROSCI.1971-19.2019

Csibra, G., & Gergely, G. (2007). "Obsessed with goals": Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*, 60–78. https://doi.org/10.1016/j.actpsy.2006.09.007

Duncan, K., Curtis, C., & Davachi, L. (2009). Distinct memory signatures in the hippocampus: Intentional states distinguish match and mismatch enhancement signals. *Journal of Neuroscience*, *29*(1), 131–139. https://doi.org/10.1523/JNEUROSCI.2998-08.2009

Duncan, K., Ketz, N., Inati, S. J., & Davachi, L. (2012). Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus*, *22*(3), 389–398. https://doi.org/10.1002/hipo.20933

Eisenberg, M. L., Zacks, J. M., & Flores, S. (2018). Dynamic prediction during perception of everyday events. *Cognitive Research: Principles and Implications*, *3*(53), 1–12. https://doi.org/10.1186/s41235-018-0146-z

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories-The role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, *278*, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011

Favila, N., Gurney, K., & Overton, P. G. (2023). Role of the basal ganglia in innate and learned behavioural sequences. *Reviews in the Neurosciences*, *35*(1), 35–55. https://doi.org/10.1515/revneuro-2023-0038

Fernández, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: Reconsolidation and the case of Prediction Error. *Neuroscience and Biobehavioral Reviews*, *68*, 423–441. https://doi.org/10.1016/j.neubiorev.2016.06.004

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P, Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*(4), 189–210. https://doi.org/10.1002/hbm.460020402

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Annals of the New York Academy of Sciences*, *1156*, 97–117. https://doi.org/10.1111/j.1749-6632.2009.04425.x

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, *70*, 119–136. https://doi.org/10.1006/nlme.1998.3843

Jainta, B., Siestrup, S., El-Sourani, N., Trempler, I., Wurm, M. F., Werning, M., Cheng, S., & Schubotz, R. I. (2022). Seeing What I Did (Not): Cerebral and Behavioral Effects of Agency and Perspective on Episodic Memory Re-activation. *Frontiers in Behavioral Neuroscience*, *15*, 1–18. https://doi.org/10.3389/fnbeh.2021.793115

Jocher, G. (2020). *YOLOv5 by Ultralytics*. Github. https://github.com/ultralytics/yolov5

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *NeuroImage*, *132*, 417–424. https://doi.org/10.1016/j.neuroimage.2016.02.062

Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J., & Rushworth, M. F. S. (2006). Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience*, *9*(7), 940–947. https://doi.org/10.1038/nn1724

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S. J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, *7*(12), 1299–1301. https://doi.org/10.1038/nn1355

Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *NeuroImage*, *34*(4), 1774–1781. https://doi.org/10.1016/j.neuroimage.2006.11.014

Klein, T. A., Ullsperger, M., & Danielmeier, C. (2013). Error awareness and the insula: Links to neurological and psychiatric diseases. *Frontiers in Human Neuroscience*, *7*, 1–14. https://doi.org/10.3389/fnhum.2013.00014

Klein-Flügge, M. C., Bongiosnni, A., & Rushworth, M. F. S. (2022). Medial and orbital frontal cortex in decision-making and flexible behavior. *Neuron*, *220*(17), 2743–2770. 10.1016/j.neuron.2022.05.022

Koster-Hale, J., & Saxe, R. (2013). Theory of Mind: A Neural Prediction Problem. *Neuron*, *79*(5), 836–848. https://doi.org/10.1016/j.neuron.2013.08.020

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Masís-Obando, R., Norman, K. A., & Baldassano, C. (2022). Schema representations in distinct brain networks support narrative memory during encoding and retrieval. *ELife*, *11*, 1–25. https://doi.org/10.7554/eLife.70445

Michels, T. (2022). *Multi_Camera_Calibration*. Gitlab. https://gitlab.com/ungetym/Multi_Camera_Calibration

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28–38. https://doi.org/10.1037/h0035584

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Olson, E. (2011). AprilTag: A robust and flexible visual fiducial system. *Proceedings - IEEE International Conference on Robotics and Automation*, 3400–3407. https://doi.org/10.1109/ICRA.2011.5979561

Ondobaka, S., & Bekkering, H. (2012). Hierarchy of idea-guided action and perception-guided movement. *Frontiers in Psychology*, *3*, 1–5. https://doi.org/10.3389/fpsyg.2012.00579

Ostlund, S. B., Winterbauer, N. E., & Balleine, B. W. (2009). Evidence of action sequence chunking in goal-directed instrumental conditioning and its dependence on the dorsomedial prefrontal cortex. *Journal of Neuroscience*, *29*(25), 8280–8287. https://doi.org/10.1523/JNEUROSCI.1176-09.2009

Pesquita, A., Whitwell, R. L., & Enns, J. T. (2018). Predictive joint-action model: A hierarchical predictive approach to human cooperation. *Psychonomic Bulletin and Review*, *25*(5), 1751–1769. https://doi.org/10.3758/s13423-017-1393-6

Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical Active Inference : A Theory of Motivated Control. *Trends in Cognitive Sciences*, *22*(4), 294–306. https://doi.org/10.1016/j.tics.2018.01.009

Pomp, J., Heins, N., Trempler, I., Kulvicius, T., Tamosiunaite, M., Mecklenbrauck, F., Wurm, M. F., Wörgötter, F., & Schubotz, R. I. (2021). Touching events predict human action segmentation in brain and behavior. *NeuroImage*, *243*(March), 118534. https://doi.org/10.1016/j.neuroimage.2021.118534

Reagh, Z. M., & Ranganath, C. (2023). Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. *Nature Communications*, *14*(1279), 1–15. https://doi.org/10.1038/s41467-023-36805-5

Richmond, L. L., & Zacks, J. M. (2018). *Constructing experience: Event models from perception to action*. *21*(12), 962–980. https://doi.org/10.1016/j.tics.2017.08.005.

Ruotolo, F., Ruggiero, G., Raemaekers, M., Iachini, T., van der Ham, I. J. M., Fracasso, A., & Postma, A. (2019). Neural correlates of egocentric and allocentric frames of reference combined with metric and non-metric spatial relations. *Neuroscience*, *409*, 235–252. https://doi.org/10.1016/j.neuroscience.2019.04.021

Sasaki, A. T., Okamoto, Y., Kochiyama, T., Kitada, R., & Sadato, N. (2018). Distinct sensitivities of the lateral prefrontal cortex and extrastriate body area to contingency between executed and observed actions. *Cortex*, *108*, 234–251. https://doi.org/10.1016/j.cortex.2018.08.003

Schiffer, A. M., Ahlheim, C., Ulrichs, K., & Schubotz, R. I. (2013). Neural changes when actions change: Adaptation of strong and weak expectations. *Human Brain Mapping*, *34*(7), 1713–1727. https://doi.org/10.1002/hbm.22023

Schubotz, R. I., Korb, F. M., Schiffer, A. M., Stadler, W., & von Cramon, D. Y. (2012). The fraction of an action is more than a movement: Neural signatures of event segmentation in fMRI. *NeuroImage*, *61*(4), 1195–1205. https://doi.org/10.1016/j.neuroimage.2012.04.008

Shima, K., & Tanji, J. (2000). Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *Journal of Neurophysiology*, *84*(4), 2148–2160. https://doi.org/10.1152/jn.2000.84.4.2148

Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology*, *7*, 1–21. https://doi.org/10.3389/fpsyg.2016.01792

Siestrup, S., Jainta, B., Cheng, S., & Schubotz, R. I. (2023). Solidity Meets Surprise: Cerebral and Behavioral Effects of Learning from Episodic Prediction Errors. *Journal of Cognitive Neuroscience*, *35*(2), 291–313. https://doi.org/10.1162/jocn_a_01948

Siestrup, S., Jainta, B., El-Sourani, N., Trempler, I., Wurm, M. F., Wolf, O. T., Cheng, S., & Schubotz, R. I. (2022). What Happened When? Cerebral Processing of Modified Structure and Content in Episodic Cueing. *Journal of Cognitive Neuroscience*, *34*(7), 1287–1305. https://doi.org/10.1162/jocn_a_01862

Siestrup, S., & Schubotz, R. I. (2023). Minor Changes Change Memories: Functional Magnetic Resonance Imaging and Behavioral Reflections of Episodic Prediction Errors. *Journal of Cognitive Neuroscience*, *35*(11), 1823–1845. https://doi.org/10.1162/jocn_a_02047

Sinclair, A. H., & Barense, M. D. (2019). Prediction Error and Memory Reactivation: How Incomplete Reminders Drive Reconsolidation. *Trends in Neurosciences*, *42*(10), 727–739. https://doi.org/10.1016/j.tins.2019.08.007

Sinclair, A. H., Manalili, G. M., Brunec, I. K., Alison Adcock, R., & Barense, M. D. (2021). Prediction errors disrupt hippocampal representations and update episodic memories. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(51), 1–12. https://doi.org/10.1073/pnas.2117625118

Sommer, T., Hennies, N., Lewis, P. A., & Alink, A. (2022). The assimilation of novel information into schemata and its efficient consolidation. *Journal of Neuroscience*, *42*(30), 5916–5929. https://doi.org/10.1523/JNEUROSCI.2373-21.2022

Stadler, W., Schubotz, R. I., von Cramon, D. Y., Springer, A., Graf, M., & Prinz, W. (2011). Predicting and memorizing observed action: Differential premotor cortex involvement. *Human Brain Mapping*, *32*(5), 677–687. https://doi.org/10.1002/hbm.20949

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage*, *49*(4), 3099–3109. https://doi.org/10.1016/j.neuroimage.2009.11.015

Sulpizio, V., Committeri, G., Lambrey, S., Berthoz, A., & Galati, G. (2013). Selective role of lingual/parahippocampal gyrus and retrosplenial complex in spatial memory across viewpoint changes relative to the environmental reference frame. *Behavioural Brain Research*, *242*(1), 62–75. https://doi.org/10.1016/j.bbr.2012.12.031

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event Boundaries in Perception Affect Memory Encoding and Updating. *Journal of Experimental Psychology: General*, *138*(2), 236–237. https://doi.org/10.1037/a0015631

Tanji, J. (2001). Sequential Organization of Multiple Movements: Involvement of Cortical. *Annual Review of Neuroscience*, *24*, 631–651.

Uithol, S., van Rooij, I., Bekkering, H., & Haselager, P. (2012). Hierarchies in action and motor control. *Journal of Cognitive Neuroscience*, *24*(5), 1077–1086. https://doi.org/10.1162/jocn_a_00204

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure & Function*, *214*, 629–643. https://doi.org/10.1007/s00429-010-0261-1

Urgen, B. A., Pehlivan, S., & Saygin, A. P. (2019). Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. *Neuropsychologia*, *127*, 35–47. https://doi.org/10.1016/j.neuropsychologia.2019.02.006

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. https://doi.org/10.1016/j.cortex.2020.03.014

Wahlheim, C. N., Eisenberg, M. L., Stawarczyk, D., & Zacks, J. M. (2022). Understanding Everyday Events: Predictive-Looking Errors Drive Memory Updating. *Psychological Science*, *33*(5), 765–781. https://doi.org/10.1177/09567976211053596

Wörgötter, F., Aksoy, E. E., Krüger, N., Piater, J., Ude, A., & Tamosiunaite, M. (2013). A simple ontology of manipulation actions based on hand-object relations. *IEEE*

Transactions on Autonomous Mental Development, *5*(2), 117–134. https://doi.org/10.1109/TAMD.2012.2232291

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI Time-Series Revisited - Again. In *NeuroImage* (Issue 2, pp. 173–181). https://doi.org/https://doi.org/10.1006/nimg.1995.1023

Wurm, M. F., von Cramon, D. Y., & Schubotz, R. I. (2011). Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. *Human Brain Mapping*, *32*(12), 2141–2150. https://doi.org/10.1002/hbm.21176

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*(12), 4057–4066. https://doi.org/10.1162/jocn_a_00078

Zacks, J. M., & Sargent, J. Q. (2010). Event Perception: A Theory and Its Application to Clinical Neuroscience. In *Psychology of Learning and Motivation - Advances in Research and Theory* (1st ed.). Elsevier Inc. https://doi.org/10.1016/S0079-7421(10)53007-X

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event Perception: A Mind/Brain Perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273.Event

Ziaeetabar, F., Pomp, J., Pfeiffer, S., El-Sourani, N., Schubotz, R. I., Tamosiunaite, M., & Wörgötter, F. (2021). Using enriched semantic event chains to model human action prediction based on (minimal) spatial information. *PLoS ONE*, *15*, 1–25. https://doi.org/10.1371/journal.pone.0243829

3.2. Study 2: Seeing the Unexpected: Oculomotor Signatures of Prediction Error at Action Boundaries

Selvan, R. N., Zahedi, A., Cheng, M., Tamosiunaite, M., Wörgötter, F., & Schubotz, R. I.

Under Review

Abstract

Eye movements provide a powerful window into how observers generate, maintain, and update predictions during action observation. While prior work shows that gaze anticipates upcoming actions and responds to surprising events, little is known about the moment-to-moment dynamics of gaze when a prediction error occurs within a multi-step action sequence. The present study investigated oculomotor signatures of prediction errors using a controlled yet ecologically valid action-observation paradigm. Participants were trained on probabilistically structured, multi-step action sequences and later observed both learned and "illegal" sequences in which a single action step was substituted at specific point in time during the sequence to induce a prediction error. Eye tracking and pupillometry were analyzed around action boundaries defined by computer vision algorithms, focusing on fixation count, fixation duration, saccade velocity, and pupil dilation. Points of violation elicited a coordinated pattern of gaze responses consistent with high-precision prediction errors: Pupil dilation increased, indicating heightened arousal and noradrenergic engagement, as did saccade velocity, suggesting intensified exploratory sampling under uncertainty. Fixation behavior shifted toward rapid visual exploration, with more incorrect fixations and shorter fixation durations. Together, these findings indicate that prediction errors shape gaze behavior, reflecting attentional reorientation, increased arousal, and updating internal models when expected trajectories break down.

Keywords: *Action observation, Prediction error, Eye tracking, Pupillometry, Oculomotor Dynamics*

Introduction

During action observation, eye movements reveal where observers expect upcoming information to appear, how they monitor unfolding actions, and how they react when events deviate from expectation. Eye tracking has therefore become a central tool for studying how observers generate and update expectations about unfolding events (Eisenberg et al., 2018; Hayhoe et al., 2012; Su et al., 2025). In infancy research, Violation-of-Expectation (VoE) paradigms are one of the extensively used approaches to study infants' understanding of the social environment (Gredebäck & Daum, 2015; Köster et al., 2020). Through such paradigms, studies have been able to reveal statistical learning and predictive phenomena in action observation (Monroy et al., 2017a; Monroy et al., 2017b), with studies reporting indicators of both anticipation and surprise (Henny Yeung et al., 2016; Hunnius & Bekkering, 2010). Similar findings have been reported in adults, with eye movements reflecting predictive locking and changes in gaze behavior when expectations are violated (Bakst & Mcguire, 2021; Gredebäck & Falck-Ytter, 2015a; Henderson, 2017).

Building on this work, fixation count, fixation duration, saccade velocity, and pupil dilation are complementary indicators of perceptual and attentional processes associated with prediction-error signaling. Prior work suggests that the prediction accuracy and ocular fixations are related in a way that when the prediction error increases, the fixation counts also increases reflecting enhanced visual sampling when predictions become unreliable (Torrents-Rodas et al., 2021). Consistently, studies have shown that fixation duration increased when decisions were less certain, indicating that observers spent more time gathering visual information under higher uncertainty (Brunyé & Gardony, 2017b; Hooge & Erkelens, 1996; Jacobs & O'Regan, 1987). Reductions in saccade velocity have been associated with the interruption of ongoing oculomotor routines in the face of unexpected events and the need to reorient visuospatial attention (Go et al., 2022). In addition, gaze responses to unexpected events are reflected not only in oculomotor behavior but also in pupil dilation, which is widely regarded as an index of arousal and noradrenergic engagement during surprising or newsworthy events (Mathôt, 2018; Viglione et al., 2023). Such violation reflects unexpected uncertainty, i.e., events that mismatch the current predictive model and trigger updating (Yu & Dayan, 2005). Accordingly, pupillometry has often been used as a behavioral measure to investigate how the locus coeruleus-norepinephrine system encodes uncertainty (Joshi et al., 2016; Nassar et al., 2013).

Despite substantial evidence that eye movements track both predictions and their violation, we still know little about the moment-to-moment dynamics of gaze when an expected action suddenly unfolds differently than predicted. Most studies provide evidence for anticipation or surprise (Drążyk & Missal, 2022; Go et al., 2022), but rarely capture how gaze evolves across the transition from prediction to prediction error.

Against this background, the present study examined gaze signatures of prediction errors during the observation of multi-step action sequences. We asked how observers' eye movements and pupil responses change when a clearly defined prediction error occurs in an otherwise familiar, probabilistic action sequence. To address this question, participants were trained in a repertoire of multi-step actions composed of fixed action steps. Overlap between actions (e.g., ABCDE and ABCGH) produced characteristic transition probabilities. During a subsequent fMRI/ Eye Tracking (ET) session, they observed the same actions performed by another actor, with a subset of actions introducing a single, time-locked violation of an individual action step, where one expected action step was substituted by a different one. This design allowed us to characterize gaze behavior in the periods immediately before and after a precisely defined prediction error.

The experimental design incorporated several features that extend prior work. First, we examined predictive gaze in multi-step action sequences, capturing a level of complexity that extends beyond the single-step actions commonly used in prior research. Second, our paradigm combines high ecological validity with full experimental control: participants were trained on a well-defined action grammar with probabilistic action steps, yet all transitions and violations remained precisely specified. Third, prediction errors were introduced at time-locked, objectively defined action boundaries, enabling fine-grained analyses of gaze behavior immediately before and after the violation. To this end, we defined two types of action boundaries using computer vision algorithms (Aksoy et al., 2011; Wörgötter et al., 2013a; Ziaeetabar et al., 2021a): the point when the hand touches the object (T), marking the beginning of an action step, and the point when the hand untouches (releases) the object (U), marking the end of an action step.

Participants were first trained by imitation on a set of probabilistically structured action sequences, ensuring that these "legal" sequences were highly familiar and predictable. In the fMRI/ET, they observed another action executing these legal action sequences, but also "illegal" action sequences in which only one single action step was replaced, thereby introducing a clearly defined prediction error at a specific action boundary. After each sequence, participants were asked to indicate whether it had been part of their training, allowing us to assess recognition while recording the gaze behavior. We analyzed fixation count, fixation duration, saccade velocity, and pupil dilation as complementary indicators of perceptual and attentional processes associated with prediction-error signaling. We hypothesized that observing illegal action sequences would elicit distinct oculomotor signatures of prediction error at the point of violation.

Methods

Participants

Forty-five subjects (30 women, 15 men) between the ages of 18 and 30 years were recruited for the study. Data from 20 participants were removed due to inaccurate calibration (n = 8), calibration failure during the experiment (n = 7), or desynchrony between eye-tracking signals and the behavioral task (n = 5). The final sample consisted of 25 datasets (*mean age* = 26.06 ± 5.32). Participants had normal or corrected-to-normal vision, with no color blindness, and were right-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), with handedness scores ranging from 50 to 100 (*M* = 87.02 ± 15.31). All participants provided written informed consent to participate in this study and received either monetary compensation (i.e., 12 euros for days one, two, and three, and 15 euros for the last day) or course credits. The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee of the University of Münster.

Stimuli

Short video clips of multi-step action sequences, viewed from a third-person perspective, were used as stimuli in the study. These action sequences involved the manipulation of four objects, namely two cups with green bases and two cubes (red and blue) on a table. Through these manipulations, 82 structured, probabilistically ordered action sequences (e.g., turning a cup upside down) were created. All action sequences occurred on a table covered with AprilTag markers (i.e., high-contrast QR code-like patterns) to facilitate a computer vision system to detect the exact timing of touching and untouching events. Each action sequence began with a consistent frame showing two hands of an actress wearing white gloves on a table, with the four objects placed in fixed positions. After that, the actress performed between 4 and 6 action steps. Each manipulation started with a hand moving towards an object, touching it, changing its position or direction, and finally untouching the object. Videos were cropped to include only the tabletop, where the actress's hands and objects can be seen clearly (c.f., Selvan et al., 2024).

The stimuli were divided into two sets. The first set of 41 action sequences, hereafter referred to as the *legal* sequences, was used for training participants. In total, there were 20 four-step, 15 five-step, and six six-step legal action sequences, each lasting between six and twelve seconds. The second set of 41 action sequences, hereafter referred to as *illegal* sequences, each corresponds to a legal sequence and contains a single altered action step. This change involved either a different manipulation of the same object or a manipulation involving a different object. The manipulation was any action step other than the first one. Notably, the other action steps of the sequence matched the legal action sequence. These sequences were designed to elicit prediction errors at a pre-determined event boundary. Finally, for imitation training, each action step of the legal sequences was also recorded separately.

Experimental Paradigm

Imitation Training (Day 01 & 02)

The study was conducted over four consecutive days, with the first two days devoted to imitation training. On each of these two days, the imitation training consisted of three phases. In the first phase, participants observed the complete legal action sequence presented on a 27-inch screen via Presentation software (version 20.3, Neurobehavioral Systems Inc., Berkeley, CA, USA). In the second phase, participants were presented with individual action steps of legal action sequences and instructed to imitate the action step as the video played. They were instructed to match the timing and hand movements demonstrated in the video. During the third phase, participants imitated the entire legal action sequence in synchrony with the original full video. During the training, the participants were seated approximately 80 cm from the screen, behind a table with the same tablecloth and the same set of objects used in the videos.

The legal sequences were grouped into four blocks based on the object with which the first action step was initiated. The four possible objects were: an upright cup, an inverted cup, a blue-side-up cube, and a red-side-up cube. Within each block, sequences were presented in increasing order of the number of action steps, beginning with four-step, then five-step, and finally six-step action sequences. Each session lasted 75-90 minutes and included a brief rest period after the second block. Throughout the training, the experimenter supervised performance and, when necessary, provided feedback, prompting participants to repeat sequences performed incorrectly. The same protocol was repeated on the second day.

fMRI - Eye tracking Session (Day 03)

On the third day of the study, participants underwent an fMRI and eye tracking session during which both legal and illegal action sequences were presented in a randomized order. Experimental procedures were written using Presentation® software (Version 20.3, Neurobehavioral Systems, Inc., Berkeley, CA, USA) and presented on a 640x512-pixel-resolution screen via a mirror mounted on the head coil in the fMRI device. The fMRI results are already presented in (Selvan et al., 2024) , and therefore, fMRI settings and results will not be discussed further here. The session was structured into three blocks, each incorporating all 41 legal action sequences, while the 41 illegal action sequences were distributed across the blocks: 13 in the first block and 14 in each of the subsequent ones. The order of videos in each block was pseudo-randomized: (1) the first video in each block was a legal action sequence, (2) no illegal sequence was shown immediately after its legal counterpart, and (3) no more than three illegal sequences were presented consecutively. After viewing each action sequence, a question was displayed on the screen asking whether participants had encountered the sequence during prior imitation training. For responding, participants used an fMRI-compatible keypad (a 4-button Current Designs keypad). Participants could respond within a maximum of two seconds, and as soon as they responded, the trial would end. Additionally, 15 null events were presented throughout the session, five per block, during which only a fixation cross was displayed for 1500, 2000, or 2500 milliseconds. In total, the experimental session included 123 trials of legal action sequences, 41 trials of illegal action sequences, 164 question trials, and 15 null events (Fig. 1). Each block lasted 16 to 20 minutes, and after each block, participants took a short break of approximately 2 to 4 minutes to recalibrate the eye-tracking equipment. The eye-tracking session lasted approximately 60 minutes in total.

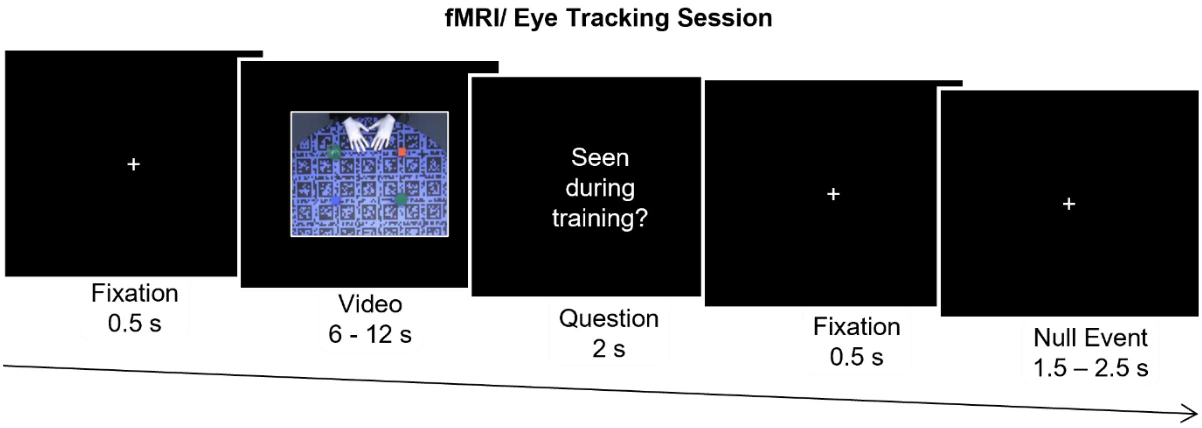


Fig. 1. Trial schema for fMRI/ET session with legal and illegal action sequences

Post-fMRI/Eye tracking Memory Task (Day 04)

On the final day of the experiment, participants completed a post-fMRI/Eye tracking memory task consisting only of the legal action sequences they had observed and imitated during the initial two days of the experiment. This task, written and conducted via Presentation software (version 20.3, Neurobehavioral Systems Inc., Berkeley, CA, USA), involved a step-by-step presentation of all 41 legal sequences. After each action step, the video paused, and participants were shown two video clips: the correct subsequent action step and an incorrect alternative. These options appeared sequentially, first on the left side of the screen, then on the right. The side and order of correct and incorrect options were counterbalanced across trials to avoid response bias.

The incorrect action steps were drawn from two sources: a subset of the illegal sequences presented during the fMRI session and additional novel illegal action steps that participants had not seen before. Participants were instructed to identify the correct next action step of the sequence by pressing one of two designated keys (A - left; L - right) on the keyboard. There was a two-second response window following the presentation of the second option. After a response was registered or once the time window elapsed, participants received immediate feedback regarding their response. Afterward, the legal action sequence resumed from the paused frame.

At the end of each action sequence, participants rated their confidence in predicting the action steps using a 4-point Likert scale (0 – no idea, 1 – rather a guess, 2 – rather know, 3 – know for sure). There was no time constraint for providing these ratings. The entire session lasted approximately 35 to 45 minutes. The results of the memory task are already presented in (Selvan et al., 2024) , and therefore, will not be discussed further here.

Eye Tracking Measurement and Preprocessing

During the third day of the experiment, eye movements and pupil size were recorded monocularly at 500 Hz using an infrared eye-tracking system (MRI-compatible EyeLink1000, SR Research, Mississauga, Ontario) with an average spatial accuracy of 0.25° of visual angle and a pupil size resolution of 0.2% of diameter. The eye-tracker camera was calibrated before each block of trials using 9-point calibration.

For preprocessing and extracting eye-tracking measures, the DataViewer software (SR Research, Ontario, Canada) was employed. To analyze eye movements, the tabletop in the videos was divided into four quadrants (Fig. 2). Fixations lasting more than 100 milliseconds and saccades with a velocity greater than 30˚/s were considered for further analysis. Poor-quality oculomotor recordings (characterized by unreliable pupil tracking and frequent blinks) were excluded from analysis, resulting in a loss of 11.18% of the data.

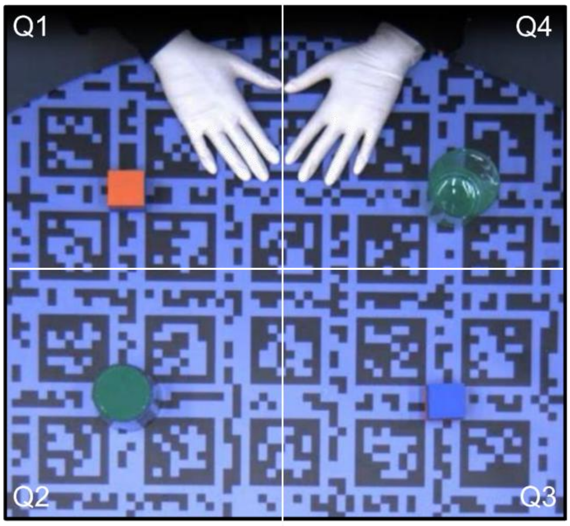


Fig. 2. The action space is divided into four quadrants based on the initial position of study objects as the areas of interest.

Statistical Analysis

Statistical analyses were performed using RStudio (2024.09.1). Both legal and illegal videos were divided into two epochs: before and after the step at which the violation occurred (i.e., the step altered in the illegal sequence relative to the legal sequence). Based on the fixation and saccades extracted from eye tracking data, several measures of interest were calculated for each epoch: total fixation count, total fixation duration, average pupil size, and average saccade velocity. Further, we divided the total fixation count into correct and incorrect fixations. To determine whether a fixation was correct or incorrect, we compared the actual fixation quadrant with the quadrant where the action happened in the video. Additionally, to calculate correct and incorrect fixations, we excluded videos where the action violation step occurred in the same quadrant as the step that it replaced (i.e., in both legal and illegal videos, the same quadrant should have been expected).

All dependent variables were analyzed using a multilevel generalized linear mixed-effect model (MGLM). Pupil size was modeled using a generalized linear mixed-effects model with a Gamma distribution and log link (Eq. 1). Furthermore, to analyze saccade velocity (Eq. 2) and fixation duration (Eq. 1), log-linear mixed-effects models were used. Finally, the fixation count (total, correct, and incorrect) was analyzed using a negative binomial model (Eq. 1), as the data were counts of events over a fixed period of time (i.e., the durations of different epochs).

The fixed-effects and random-intercept structures were identical across pupil dilation, fixation counts (total, correct and incorrect) and fixation duration models (Eq. 1). The fixed effects included Video Type *(Legal vs. Illegal)*, Segment Type *(before vs. after the violation)*, and their interaction. In addition, Branch Number, defined as the number of possible action steps at the step where the violation occurred, was included as a covariate of interest. Steps Number, defined as the number of action steps in the epoch, was also added as a nuisance covariate to regress out the effects of varying lengths of different epochs. For the saccade velocity model, the saccade amplitude was added as covariate (Eq. 2). Random intercepts were specified for Stimulus ID and Subject ID.

$$Outcome \sim Video\_Type * Segment\_Type + Branch\_Num + Steps\_Num + (1 \mid Stimuli\_ID) + (1 \mid Subject\_ID) \quad \textbf{(Equation 1)}$$

$$Saccade\ Velocity \sim Video\_Type * Segment\_Type + Branch\_Num + Steps\_Num + Saccade\_Amplitude + (1 \mid Stimuli\_ID) + (1 \mid Subject\_ID) \quad \textbf{(Equation 2)}$$

For clarifying the interaction between Video Type and Segment Type, we conducted post-hoc contrasts that were corrected for multiple comparisons using the Sidak method.

Results

Average Pupil Dilation

For the pupil dilation, the pupil size before 200 ms of the point of violation was taken as the baseline, and 2000 ms following the point of violation was considered the response window (Fig. 3A). The analysis revealed a marginally significant interaction between Video Type and Segment Type ($\chi^2(1) = 3.70$, $p = 0.054$, *Cramer's* $V = 0.032$). Post hoc contrasts showed that for the illegal action sequences, the change in pupil size was significant after the point of violation ($\widehat{mean} = 0.98 \pm 0.006$, $p < 0.001$, $z = -3.598$), reflecting heightened autonomic arousal or cognitive load in response to an unexpected event. However, there was no significant change for legal action sequences ($\widehat{mean} = 0.993 \pm 0.005$, $p = 0.184$, $z = -1.329$).

There were no significant main effects of Segment Type ($\chi^2(1) = 1.767$, $p = 0.184$, *Cramer's* $V = 0.022$), Video Type ($\chi^2(1) = 0.00$, $p = 0.984$, *Cramer's* $V = 0.000$), Branch Number ($\chi^2(1) = 3.03$, $p = 0.082$, *Cramer's* $V = 0.029$), or Steps Number ($\chi^2(1) = 0.14$, $p = 0.708$, *Cramer's* $V = 0.006$) suggesting that the pupil response was specifically driven by the violation itself rather than other sequence characteristics.

Average Saccade Velocity

In addition to the pupil dilation, the saccade velocity of legal and illegal action sequences (Fig. 3B) was also analyzed to examine whether the prediction error affected the visual scanning behavior of the participants. The results revealed a significant main effect of Segment Type ($\chi^2(1) = 48.143$, $p < 0.001$, *Cramer's* $V = 0.082$). The violation led to a decrease in scanning behavior after the violation ($\widehat{mean} = 67.4 \pm 1.55$) compared to before ($\widehat{mean} = 73.2 \pm 1.68$), reflecting uncertainty about the next action step. Notably, there was a marginally significant interaction between Video Type and Segment Type ($\chi^2(1) = 3.829$, $p = 0.05$, *Cramer's* $V = 0.022$). Post hoc contrasts indicated that after the point of violation, the saccade velocity for illegal sequences was significantly higher than legal ones ($\widehat{mean} = 0.978 \pm 0.009$, $p = 0.017$, $z = -2.384$), while there was no difference before the violation ($\widehat{mean} = -0.999 \pm 0.009$, $p = 0.95$, $z = -0.052$).

Furthermore, Saccade Amplitude ($\chi^2(1) = 5193.951$, $p < 0.001$, *Cramer's* $V = 0.847$) had a significant effect. This result suggests that as the amplitude of the saccades increased, the saccade velocity also increased after the point of violation at the critical action boundary. Further, there were no significant main effects observed for Video Type ($\chi^2(1) = 1.381$, $p =$

0.24, *Cramer's* $V = 0.014$), Steps Number ($\chi^2(1) = 0.19$, $p = 0.662$, *Cramer's* $V = 0.005$), or Branch Number ($\chi^2(1) = 1.008$, $p = 0.315$, *Cramer's* $V = 0.012$).

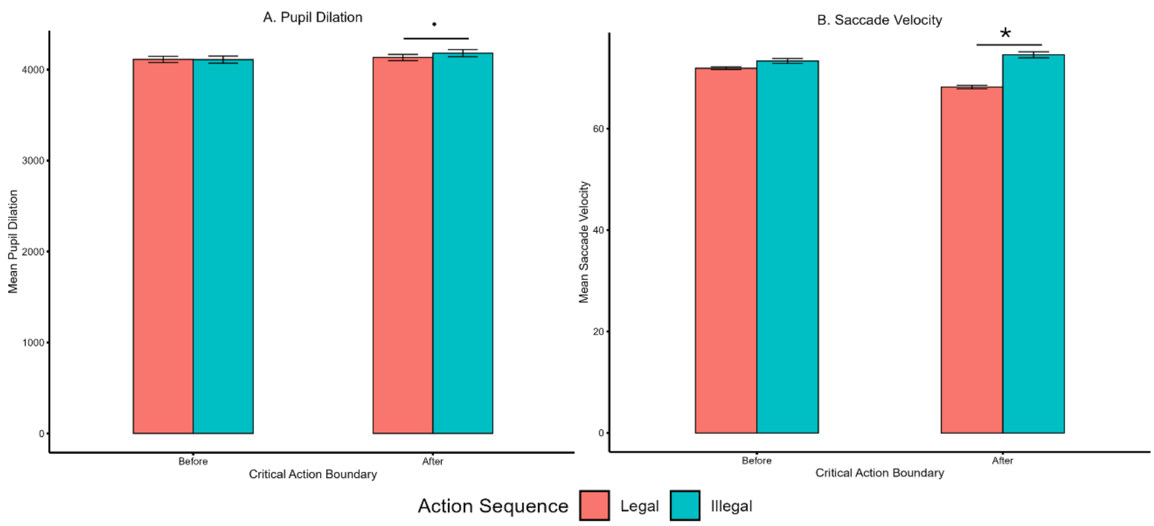


Fig. 3. Oculomotor measures before and after the critical action boundary of the legal and illegal action sequences. A) Pupil Dilation; B) Saccade Velocity

Fixation Count

Total Fixations

Following the saccade velocity, to assess whether prediction errors affected fixations, we analyzed the total fixation count of legal and illegal action sequences (Fig. 4A) for different Segment Types using a negative binomial MGLM. The results revealed a significant main effect of Segment Type ($\chi^2(1) = 54.201$, $p < .001$, *Cramer's* $V = 0.086$) showing that the number of total fixations was higher after the point of violation ($\widehat{mean} = 5.97 \pm 0.256$) compared to before ($\widehat{mean} = 5.21 \pm 0.222$). This result can serve as a manipulation check, indicating that a violation of the predictions affected the dependent variable as expected. Interestingly, we also found a significant interaction between Video Type and Segment Type ($\chi^2(1) = 10.46$, $p < .001$, *Cramer's* $V = 0.038$). Post hoc contrasts indicated that although there were fewer fixations in illegal than in legal videos before the violation ($\widehat{mean} = -0.099 \pm 0.013$, $p < .001$, $z = -7.362$), after the prediction error, the trend was reversed ($\widehat{mean} = -0.174 \pm 0.022$, $p < .001$, $z = -8.054$).

Also, the nuisance covariate, the Step Number, was strongly significant ($\chi^2(1) = 3644.52$, p < .001, *Cramer's* $V = 0.708$). This effect can be expected, as longer episodes should contain more steps, providing more opportunities to gaze at different objects and locations. Finally, neither Video Type ($\chi^2(1) = 2.98$, $p = .085$, *Cramer's V* = 0.02), nor Branch Number ($\chi^2(1) = 2.19$, $p = .139$, *Cramer's V* = 0.017), showed significant main effects.

Total Correct Fixations

After seeing the effects of prediction errors on total fixation count, we analyzed correct and incorrect fixations separately to identify the source of variance between legal and illegal videos.

For correct fixation counts, legal and illegal action sequences (Fig. 4B), we observed a significant main effect of Segment Type ($\chi^2(1) = 331.64$, $p < 0.001$, *Cramer's V* = 0. 232), showing that the number of correct fixations was higher after the point of violation ($\widehat{mean}$ = 3.95 ± 0.203) compared to before ($\widehat{mean}$ = 2.98 ± 0.153). This result indicates that training was effective in inducing predictions in participants that were translated into correct fixations. Notably, a significant interaction was found between Video Type and Segment Type ($\chi^2(1) = 10.21$, $p = 0.001$, *Cramer's V* = 0.041). Post hoc contrasts revealed that although before the violation, legal and illegal action sequences did not differ significantly ($\widehat{mean}$ = 0.048 ± 0.035, $p = 0.164$, $z = 1.392$), after the violation, correct fixations were lower in *illegal* action sequences compared to legal ones ($\widehat{mean}$ = 0.152 ± 0.036, $p < 0.001$, $z = 4.188$).

Additionally, Step Number also had a significant effect ($\chi^2(1) = 2120.20$, $p < 0.001$, *Cramer's V* = 0.586), which, similar to total fixation count, was expected. Finally, no significant main effects were observed for Video Type ($\chi^2(1) = 1.94$, $p = 0.164$, *Cramer's V* = 0.018) or Branch Number ($\chi^2(1) = 2.47$, $p = 0.116$, *Cramer's V* = 0.02).

Total Incorrect Fixations

Finally, after checking the correct and total fixations, we focused on the incorrect fixations of legal and illegal action sequences (Fig. 4C). The analysis of incorrect fixations revealed a significant main effect of Segment type ($\chi^2(1) = 5.48$, $p = 0.019$, *Cramer's V* = 0.032), revealing more incorrect fixations after the violation point ($\widehat{mean}$= 2.75 ± 0.162) compared to before ($\widehat{mean}$= 2.73 ± 0.1559). Notably, a significant interaction was found between Video Type and Segment Type ($\chi^2(1) = 9.58$, $p = 0.002$, *Cramer's V* = 0.042). Post-hoc contrasts revealed that in legal action sequences, incorrect fixations significantly decreased after the violation point ($\widehat{mean}$ = 0.053 ± 0.023, $p = .019$, $z = 2.341$). In contrast, in illegal videos, incorrect fixations showed a marginal increase after the violation ($\widehat{mean}$ = –0.066 ± 0.035, $p = .059$, $z. = –1.885$).

Additionally, a strong effect of Step Number was observed ($\chi^2(1) = 440.16$, $p < 0.001$, *Cramer's V* = 0.286), which was expected, similar to the effects observed in the analysis of correct and total fixations. No significant main effects were observed for Video Type ($\chi^2(1) = 2.39$, $p = 0.122$, *Cramer's V* = 0.021) or Branch Number ($\chi^2(1) = 1.89$, $p = 0.169$, *Cramer's V* = 0. 019).

Total Fixation Duration

Following the fixation counts, the analysis of fixation duration of legal and illegal action sequences (Fig. 4D) also revealed a significant main effect of Segment Type ($\chi^2(1) = 75.42$, $p < 0.001$, *Cramer's V* = 0.102) with duration increased after the point of violation ($\widehat{mean} = 593 \pm 31.2$) compared to before ($\widehat{mean} = 483 \pm 25.4$). Notably, there was a significant interaction between the Video Type and Segment Type ($\chi^2(1) = 19.20$, $p < 0.001$, *Cramer's V* = 0.051). Post hoc contrasts revealed that after the point of violation, fixation durations were significantly shorter for illegal compared to legal ones ($\widehat{mean} = 1.118 \pm 0.026$, $p < 0.001$, $z = 4.832$). However, before the point of violation, the fixation duration was not significantly different between legal and illegal action sequences ($\widehat{mean} = 1.012 \pm 0.023$, $p = 0.593$, $z = 0.535$). This result complements fixation count results: there are more fixations following the point of violation, while the duration of each fixation decreases.

Moreover, there was a main effect of Steps Number ($\chi^2(1) = 374.33$, $p < 0.001$, *Cramer's V* = 0.102), with longer fixations observed when violations occurred at later steps, suggesting heightened prediction error processing as action sequences progressed. No significant main effects were observed for Video Type ($\chi^2(1) = 0.29$, $p = 0.593$, *Cramer's V* = 0.006) or Branch Number ($\chi^2(1) = 0.91$, $p = 0.341$, *Cramer's V* = 0.111).

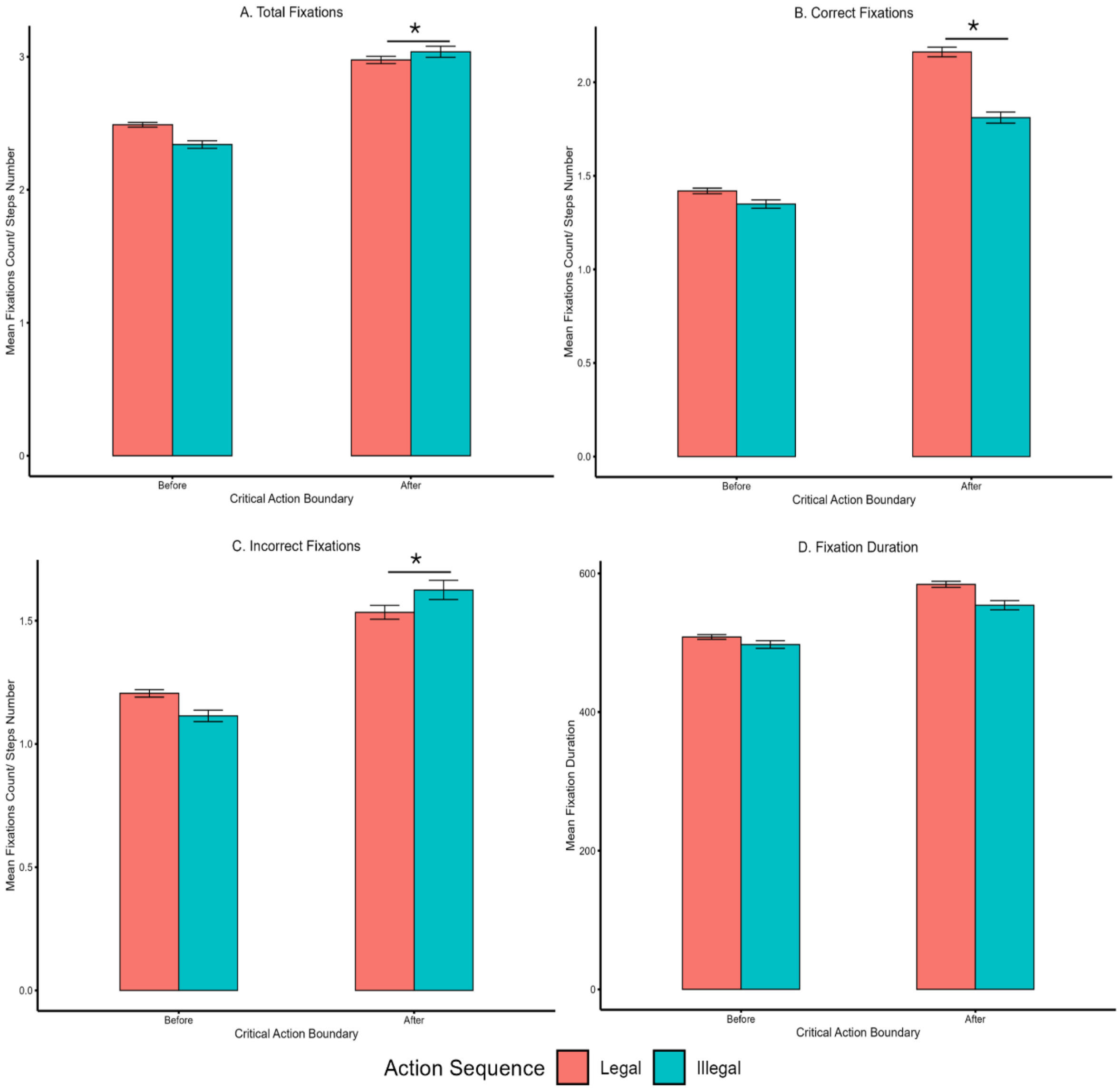


Fig. 4. Oculomotor measures before and after the critical action boundary of the legal and illegal action sequences. A) Total Fixations Count; B) Correct Fixations Count; C) Incorrect Fixations Count; D) Fixation Duration

Discussion

Eye measures have been an important tool to study how people generate and update expectations about ongoing events. Previous studies have shed light on various eye measures that include fixations, saccades and pupillometry to understand oculomotor signatures of prediction. Specifically, fixations indicate where observers look for goal-relevant information to appear, saccades reveal the dynamics of predictive monitoring and exploratory sampling, and pupil dilation provides a physiological index of the arousal and cognitive effort associated with detecting surprising events. The current study examined gaze signatures of prediction errors during the observation of complex, multi-step action sequences. By capturing moment-to-moment gaze behavior around the point of violation in this structured action repertoire, our work provides fine-grained evidence that has so far been missing on how observers visually track unexpected deviations in ongoing action sequences. Taken together, our findings show that prediction errors evoke coordinated changes across fixation patterns, saccadic metrics, and pupillary responses, underscoring the active role of the oculomotor system in adapting to unexpected events within unfolding actions. At the critical action boundary, the pupil dilated for illegal action sequences compared to legal ones. Following the point of violation, the saccades were faster, while the number of correct fixations decreased and incorrect fixations increased in illegal action sequences. Additionally, the fixation durations were shorter for illegal than for legal action sequences.

At the point of violation of the illegal action sequences, pupil dilation occurred. This aligns with the finding that pupils dilate in response to heightened arousal or cognitive load (Mathôt, 2018). In our experiment, we hypothesized that pupil dilation would selectively increase in response to violations in the illegal action sequences, reflecting a prediction error. As pupil diameter indexes locus coeruleus–noradrenergic activity (Joshi et al., 2016), this increase under high prediction error can be interpreted as enhanced precision weighting (Preuschoff et al., 2011) and cognitive effort involved in the reorganization of internal models (Filipowicz et al., 2020). Thus, the arousal-related nature of pupil dilation provides physiological evidence that illegal action sequences generated prediction errors of sufficient magnitude and precision to signal them as a salient, newsworthy event.

Another distinct gaze behavior observed following the prediction error at the critical action boundary was the saccade velocity: Illegal action sequences elicited higher saccade velocity compared to legal ones. We interpret this increase as an arousal-driven response to the violation, consistent with the concomitant rise in pupil diameter, which likewise indicates heightened noradrenergic engagement. Elevated arousal was found to accelerate saccadic movements by increasing movement urgency and amplitude; DiGirolamo et al., (2016) showed that suddenly induced arousal physiologically increases saccade amplitude and peak saccade velocity, indicating that arousal can directly amplify saccade velocity. This arousal-related speeding stands in contrast to evidence from studies reporting decreases in saccade velocity under prediction error. In Drążyk & Missal, (2022), saccades slowed as spatial uncertainty increased: participants maintained central fixation while 1-4 potential target locations were cued, and higher uncertainty reliably reduced peak saccade velocity. Similarly, Go et al., (2022) observed reduced saccade velocity when participants encountered unexpected shifts in the learned probability distribution of single-target locations. In both studies, prediction errors arose from uncertainty about where the target would appear, a mechanism that tends to promote more cautious oculomotor responses rather than arousal-driven urgency. Under this interpretation, the continuous nature of our action sequences may also be relevant: in contrast to the abruptly appearing static targets used in the prior studies, both our task and the dynamic pursuit–saccade paradigm in DiGirolamo et al., (2016) involved violations or arousal cues embedded within ongoing movement, a context that may elicit more sustained arousal responses and thereby contribute to the observed increase in saccade velocity.

In the fixation-frequency analysis, we examined total, incorrect, and correct fixations occurring after the critical action boundary, for both legal and illegal action sequences. Although the total number of fixations was not significantly different between conditions, illegal action sequences elicited significantly more incorrect fixations, whereas legal sequences produced significantly more correct fixations. From a predictive-processing perspective, the violation introduced uncertainty that prompted the system to evaluate the sensory input against the stored internal representation (Aggelopoulos, 2015), leading participants to check the environment more extensively. In line with this, the study by (Torrents-Rodas et al., 2021) demonstrated that prediction errors broaden the attentional focus and increase sensitivity to irrelevant cues when solving a discrimination task. Correspondingly, participants in our study may have directed attention towards the three non-target quadrants reflecting uncertainty about which action steps could follow, as the violation disrupted the internal model established during imitation training. This increased sensitivity to irrelevant spatial alternatives, coupled with a stronger drive to reduce ambiguity, likely amplified visual search behavior. Consistent with the concept of attentional exploration described by Beesley et al., (2015), the violation in our study altered the configuration of objects, and this change persisted even after the correct action sequence resumed. Because the prediction error left a visible trace in the scene, observers continued to explore the modified layout, resulting in more incorrect fixations in the illegal action sequences.

Finally, the uncertainty in the action space introduced by the violation of expectations in illegal action sequences led to a decrease in the average fixation durations after the critical action boundary, compared to the legal action sequences. Previous findings have shown that under uncertainty, the fixation duration can increase to gather visual information to make decisions (Brunyé & Gardony, 2017). In contrast, the present findings showed a reduction in fixation duration alongside an increase in fixation frequency, indicating a shift toward rapid sampling and exploratory scanning. It is important to note that even though the action sequences consisted of only one actor and four study objects, the training phase established a meaningful action space that participants actively monitored. When a violation occurred, this space effectively reopened: suppressed alternatives became relevant again, prompting participants to scan the scene more rapidly to determine which action might plausibly follow. Thus, the increased exploratory scanning is not only a visible trace of the violation but also a functional response to the probabilistic action repertoire they had internalized.

In summary, illegal action sequences consistently elicited stronger uncertainty responses, including intensified visual search reflected in increased incorrect fixations, rapid attentional shifts evidenced by decreased fixation duration, and heightened arousal indexed by increased saccade velocity and pupil dilation. Taken together, these converging eye-movement signatures show that prediction errors in multi-step action sequences trigger a coordinated reorientation of perceptual processing. Violations do not merely disrupt expectations but reshape how observers sample the visual scene, broadening attentional deployment, accelerating exploratory movements, and increasing arousal. This pattern suggests that the oculomotor system plays an active role in minimizing uncertainty and revising internal models (Parr & Friston, 2019) when anticipated action trajectories are no longer valid.

In addition to these insights, the current study also offers a basis for future extensions. Unlike prior work focused on isolated or single-step actions, our multi-step design makes it possible to examine how predictive processing unfolds and adjusts across an extended, hierarchically organized action repertoire. The combination of probabilistic action sequences and clearly defined action boundaries provides a controlled yet naturalistic context in which future studies can relate gaze behavior to evolving expectations and prediction errors. This framework can thus support more fine-grained analyses of how fixations, saccades, and pupillary responses track predictive mechanisms across complex actions.

CRediT authorship contribution statement

Rosari Naveena Selvan: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Anoushiravan Zahedi:** Writing – review & editing, Formal analysis, Validation, Methodology. **Minghao Cheng:** Writing – original draft, Software, Methodology, Data curation. **Minija Tamosiunaite** – Writing – review & editing, Resources, Methodology.

Florentin Wörgötter: Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Ricarda I. Schubotz:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Data availability

The entire stimulus material is made available via the Action Video Corpus Muenster (AVICOM, https://www.uni-muenster.de/IVV5PSY/AvicomSrv/). The raw eye tracking data and the raw SEC time point extraction data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the German Research Foundation (DFG) [WO 388/13–1]. The funder had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report.

Acknowledgments

We extend our sincere gratitude to Falko Mecklenbrauck and Maximilian Bruchmann for their technical support in initiating the study. We would like to thank Monika Mertens, Christin Schwarzer, Lena Puder, Maria Leonie Hofmann, Simon Wieczorek, Kristin Stroop, Lars Schlüter, Leon Exeler and Niklas Dielitzsch for helping to recruit participants and/or assisting during data collection for the pilot and final studies.

References

Aggelopoulos, N. C. (2015). Perceptual inference. *Neuroscience and Biobehavioral Reviews*, *55*, 375–392. https://doi.org/10.1016/j.neubiorev.2015.05.001

Aksoy, E. E., Abramov, A., Dörr, J., Ning, K., Dellen, B., & Wörgötter, F. (2011). Learning the semantics of object-action relations by observation. *International Journal of Robotics Research*, *30*(10), 1229–1249. https://doi.org/10.1177/0278364911410459

Bakst, L., & Mcguire, J. T. (2021). Eye movements reflect adaptive predictions and predictive precision. *Journal of Experimental Psychology: General,* *150*(5), 915–929. https://doi.org/10.1037/xge0000977.Eye

Beesley, T., Nguyen, K. P., Pearson, D., & Le Pelley, M. E. (2015). Uncertainty and predictiveness determine attention to cues during human associative learning. *Quarterly Journal of Experimental Psychology*, *68*(11), 2175–2199. https://doi.org/10.1080/17470218.2015.1009919

Brunyé, T. T., & Gardony, A. L. (2017). Eye tracking measures of uncertainty during perceptual decision making. *International Journal of Psychophysiology*, *120*, 60–68. https://doi.org/10.1016/j.ijpsycho.2017.07.008

DiGirolamo, G. J., Patel, N., & Blaukopf, C. L. (2016). Arousal facilitates involuntary eye movements. *Experimental Brain Research*, *234*(7), 1967–1976. https://doi.org/10.1007/s00221-016-4599-3

Drążyk, D., & Missal, M. (2022). The oculomotor signature of expected surprise. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-06403-4

Eisenberg, M. L., Zacks, J. M., & Flores, S. (2018). Dynamic prediction during perception of everyday events. *Cognitive Research: Principles and Implications*, *3*(53), 1–12. https://doi.org/10.1186/s41235-018-0146-z

Filipowicz, A. L. S., Glaze, C. M., Kable, J. W., & Gold, J. (2020). Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *ELife*, *9*, 1–23. https://doi.org/10.7554/eLife.57872

Go, H., Danckert, J., & Anderson, B. (2022). Saccadic eye movement metrics reflect surprise and mental model updating. *Attention, Perception, and Psychophysics*, *84*(5), 1553–1565. https://doi.org/10.3758/s13414-022-02512-4

Gredebäck, G., & Daum, M. M. (2015). The Microstructure of Action Perception in Infancy: Decomposing the Temporal Structure of Social Information Processing. *Child Development Perspectives*, *9*(2), 79–83. https://doi.org/10.1111/cdep.12109

Gredebäck, G., & Falck-Ytter, T. (2015). Eye Movements During Action Observation. *Perspectives on Psychological Science*, *10*(5), 591–598. https://doi.org/10.1177/1745691615589103

Hayhoe, M. M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research*, *217*(1), 125–136. https://doi.org/10.1007/s00221-011-2979-2

Henderson, J. M. (2017). Gaze Control as Prediction. *Trends in Cognitive Sciences*, *21*(1), 15–23. https://doi.org/10.1016/j.tics.2016.11.003

Henny Yeung, H., Denison, S., & Johnson, S. P. (2016). Infants' looking to surprising events: When eye-tracking reveals more than looking time. *PLoS ONE*, *11*(12), 1–14. https://doi.org/10.1371/journal.pone.0164277

Hooge, I. T. C., & Erkelens, C. J. (1996). Control of fixation duration in a simple search task. *Perception and Psychophysics*, *58*(7), 969–976. https://doi.org/10.3758/BF03206825

Hunnius, S., & Bekkering, H. (2010). The Early Development of Object Knowledge: A Study of Infants' Visual Anticipations During Action Observation. *Developmental Psychology*, *46*(2), 446–454. https://doi.org/10.1037/a0016543

Jacobs, A. M., & O'Regan, J. K. (1987). Spatial and/or temporal adjustments of scanning behavior to visibility changes. *Acta Psychologica*, *65*(2), 133–146. https://doi.org/10.1016/0001-6918(87)90023-0

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, *89*(1), 221–234. https://doi.org/10.1016/j.neuron.2015.11.028

Köster, M., Kayhan, E., Langeloh, M., & Hoehl, S. (2020). Making Sense of the World: Infant Learning From a Predictive Processing Perspective. *Perspectives on Psychological Science*, *15*(3), 562–571. https://doi.org/10.1177/1745691619895071

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/joc.18

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017a). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology*, *157*, 14–28. https://doi.org/10.1016/j.jecp.2016.12.004

Monroy, C. D., Gerson, S., & Hunnius, S. (2017b). Infants' motor proficiency and statistical learning for actions. *Frontiers in Psychology*, *8*, 1–11. https://doi.org/10.3389/fpsyg.2017.02174

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2013). Rational regulation of learning dynamics by pupil–linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. https://doi.org/10.1038/nn.3130.Rational

Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics*, *113*(5–6), 495–513. https://doi.org/10.1007/s00422-019-00805-w

Preuschoff, K., ’t Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline’s role in decision making. *Frontiers in Neuroscience*, *5*, 1–12. https://doi.org/10.3389/fnins.2011.00115

Selvan, R. N., Cheng, M., Siestrup, S., Mecklenbrauck, F., Jainta, B., Pomp, J., Zahedi, A., Tamosiunaite, M., Wörgötter, F., & Schubotz, R. I. (2024). Updating predictions in a complex repertoire of actions and its neural representation. *NeuroImage*, *296,* 120687. https://doi.org/10.1016/j.neuroimage.2024.120687

Su, X., Bezdek, M. A., Nguyen, T. T., Hall, C. S., Zacks, J. M., Su, X., Bezdek, M. A., Nguyen, T. T., Hall, C. S., & Zacks, J. M. (2025). Predictive Looking and Predictive Looking Errors in Everyday Activities. Advance online publication. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001851

Torrents-Rodas, D., Koenig, S., Uengoer, M., & Lachnit, H. (2021). A rise in prediction error increases attention to irrelevant cues. *Biological Psychology*, *159*, 108007. https://doi.org/10.1016/j.biopsycho.2020.108007

Viglione, A., Mazziotti, R., & Pizzorusso, T. (2023). From pupil to the brain: New insights for studying cortical plasticity through pupillometry. *Frontiers in Neural Circuits*, *17*, 1–11. https://doi.org/10.3389/fncir.2023.1151847

Wörgötter, F., Aksoy, E. E., Krüger, N., Piater, J., Ude, A., & Tamosiunaite, M. (2013). A simple ontology of manipulation actions based on hand-object relations. *IEEE Transactions on Autonomous Mental Development*, *5*(2), 117–134. https://doi.org/10.1109/TAMD.2012.2232291

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Ziaeetabar, F., Pomp, J., Pfeiffer, S., El-Sourani, N., Schubotz, R. I., Tamosiunaite, M., & Wörgötter, F. (2021). Using enriched semantic event chains to model human action prediction based on (minimal) spatial information. *PLoS ONE*, *15*, 1–25. https://doi.org/10.1371/journal.pone.0243829

3.3. Study 3: Me or Us? Mine or Yours? Here or There? Neural and behavioral responses of prediction in Joint Action

Selvan, R. N., Zahedi, A., Cheng, M., Dohr, K., Tamosiunaite, M.,
Wörgötter, F., Schubotz, R.I

To Be Submitted

Abstract

Everyday life is filled with instances of joint action, including cooking together, unloading groceries from the car, or playing games. Research suggests that our brains employ a dual hierarchical predictive model to anticipate our own and our partner's actions simultaneously. This hierarchical model has been suggested to have three layers: a sensorimotor layer, a sequence/subgoal layer, and a schema/common goal layer. This study aimed to investigate the interplay between shared intention, sub-goal coordination, and movement coordination. We hypothesized that these levels could be modeled as a flexibly coupled dual predictive hierarchy with varying degrees of coupling strength. To test this hypothesis, we employed a two-by-two-by-two factorial design, manipulating the presence/absence of shared common goal, sub-goal coordination, and movement coordination. N=80 participants engage in teams of two players in a novel card game simulating naturalistic joint action scenario. Hand and eye movements were recorded for two players playing the game, while EEG was collected from one player. Computer vision was employed to map game actions to players' eye movements and objectively define the action boundaries as Touching (T) and Untouching (U) events. Neural oscillations to different levels of coordination – goal, sub-goal, and movement – at the action boundaries were analyzed. By investigating these measures, our findings shed light on the mechanisms underlying joint action coordination at the action boundaries and the role of hierarchical predictive processing in facilitating efficient and adaptable interpersonal interactions.

Keywords: *Joint Action, Prediction, Shared Intention, Three Layer Coordination, Action Boundaries*

Introduction

Humans frequently engage in joint actions that require precise coordination of behavior with others, such as cooking together, moving furniture, or playing games. Successful joint action depends not only on executing one's own movements but also on anticipating the actions and intentions of interaction partners in real time (Sebanz et al., 2006; Vesper et al., 2010). The ability for interpersonal coordination is widely thought to rely on motor control (Wolpert et al., 2003) and predictive mechanisms (Friston et al., 2011; Friston & Frith, 2015a; Friston & Frith, 2015b) that enable agents to generate expectations about future states of both their own and others' actions, thereby minimizing uncertainty and facilitating smooth interaction. Within the predictive processing framework, the brain is conceptualized as a hierarchical generative model that continuously predicts incoming sensory input and updates its beliefs by minimizing prediction errors across multiple levels of abstraction (Clark, 2013; K. Friston, 2005). When applied to social interaction, predictive processing suggests that individuals use internal models not only to predict the sensory consequences of their own actions but also to infer and anticipate the actions, goals, and intentions of others (Kilner et al., 2007).

A prominent framework for understanding predictive mechanisms in joint action is the Predictive Joint Action Model (PJAM), which formalizes joint action as the dynamic coupling of hierarchical predictive models across interacting agents (Pesquita et al., 2018). PJAM posits that each agent maintains predictions at multiple levels, from concrete sensorimotor outcomes to action sequences and high-level goals and joint action emerges when these hierarchies become sufficiently aligned, with coupling strength flexibly modulated by task constraints and social context. Complementing this, Kahl & Kopp (2018) proposed a three-layer predictive processing model in which hierarchical generative models operate across sensorimotor, action-sequence, and goal/schema levels. The sensorimotor layer supports fine-grained movement prediction and synchronization, the action-sequence layer represents structured sub-goals, and the goal/schema layer encodes overarching intentions and shared plans, with active inference mechanisms enabling flexible adaptation and self-other distinction. Together, these frameworks suggest that joint action arises from the mutual entrainment of hierarchical predictions, in which each agent dynamically aligns sensorimotor, sub-goal, and goal-level expectations with those of their partners, enabling smooth, flexible, and context-sensitive coordination even when individual movements differ.

EEG studies provide direct evidence for these processes: anticipatory motor potentials such as the contingent negative variation (CNV) are modulated when participants prepare actions in coordination with a partner, reflecting expectation of others' movements (Fanghella et al., 2025; Kourtis et al., 2019b). Furthermore, event related potential (ERP) component P3 (Kourtis, Sebanz, & Knoblich, 2013) and inter-brain oscillatory coupling in theta and mu/alpha suppression have been observed in cooperative human–human interactions, reflecting shared timing and prediction (Markiewicz et al., 2024; Novembre et al., 2016; Shiraishi & Shimada, 2021). A human-robot interaction EEG study has also shown the difference in neural response where participants plan the action well in advance or anticipate it to be effortful (S. Cooper et al., 2020). These findings collectively indicate that the representation and predictive mechanisms that underlie joint action across both human and human–robot contexts, with EEG providing a sensitive tool for capturing anticipatory, and synchronized neural dynamics that support real-time coordination.

Despite these advances, empirical research particularly based on flexible coupling across hierarchical levels and differential neural signatures of coordination at each level of PJAM, has yet to be systematically examined. The present study addresses this gap by experimentally manipulating coordination at all three levels proposed within a single, naturalistic joint action task. Using a 2x2x2 factorial design, we independently manipulated (1) the presence versus absence of a shared common goal, (2) sub-goal coordination, and (3) movement coordination. Participants were tested in dyads while playing a novel card game designed to capture the nature of real-world joint action while maintaining experimental control. To investigate predictive processes across hierarchical levels at the action boundaries, we integrated behavioral, computer vision, and neural measures. EEG was recorded from one participant in each dyad to examine the neural correlates of prediction at varying levels of coordination. Eye tracking was recorded for both the players to investigate the gaze behavior. Concurrently, computer vision techniques were employed to identify and align action-boundary events, specifically the Touching (T) and Untouching (U) of the object (Aksoy et al., 2011; Wörgötter et al., 2013a).

We hypothesize that the three hierarchical layers of joint action Goal, Subgoal, and Movement modulate neural oscillations in a frequency-specific manner at Touching (T) versus Untouching (U) events. Theta power is expected to decrease at T events, influenced by subgoal and goal coordination. Alpha power will decrease at T events but increase with movement coordination, reflecting attentional and sensorimotor control. Beta power will primarily track movement coordination and complex interactions among Goal, Subgoal, and Movement, increasing during T events. Together, these patterns are predicted to reveal distinct cognitive, attentional, and motor contributions of each layer in hierarchical joint action.

Methods

Participants

Eighty healthy adults (67 female, 13 male) participated in the study. Participants were between 19 and 33 years of age (*M* = 24.78 years, *SD* = 3.05). Participants were tested in pairs, resulting in a total of 40 experimental sessions. Each pair consisted of player 1 who underwent EEG and Eye tracking recording and player 2, underwent with Eye tracking only. Data from five participants was not included in the final analysis due to technical errors during the acquisition and excessive movement artifacts. All participants reported normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Participants were assessed for handedness using Edinburgh Handedness Inventory (Oldfield, 1971) with scores varying between 50 to 100 (*M* = 92.77, SD = 11.95). All participants gave written informed consent to participate in this study and were compensated with money or course credits. Player 1 was compensated with three course credits or 45 Euros and Player 2 with two credits or 30 Euros. The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee of the University of Münster.

Stimuli

Game Overview and Physical Setup

The joint-action task was implemented as a custom-designed tabletop game played by two participants seated opposite each other. A square game board (6 × 6 grid; 36 cells) was fixed to the table and aligned with the overhead multi-camera computer vision system using floor markers. The board consisted of alternating rows of black and white squares (three rows of each), with the row closest to each participant defined their assigned color.

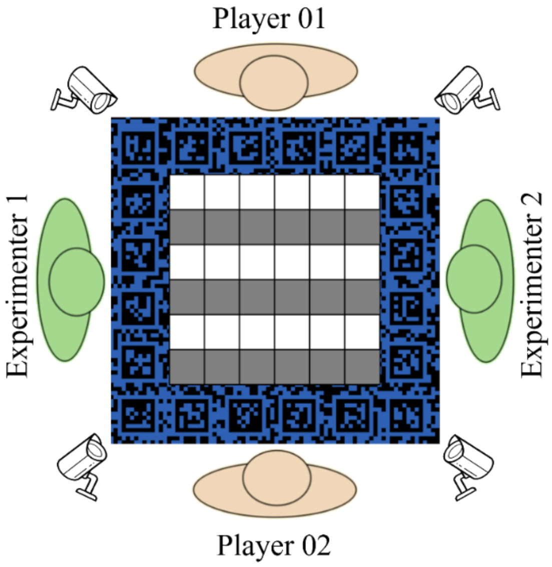


Fig.1 The experimental setup showing the two players and experimenters sitting across each other five computer vision cameras (including one at the top) to detect the action boundaries.

Each participant was provided with 18 playing cards, arranged in two stacks of nine directly in front of them. Cards measured 7.5 × 7.5 cm and were designed to be reliably detectable by the computer vision system. All cards were square, with either black or white frame and symbol (marking which player they should be assigned to) and colored shapes (colors: red, blue, green). For each player, the deck consisted of six cards of each color and six distinct symbols (star, crescent moon, circle, cross, heart, lightning), with each color – symbol combination occurring exactly once. This resulted in 36 unique cards for the game.

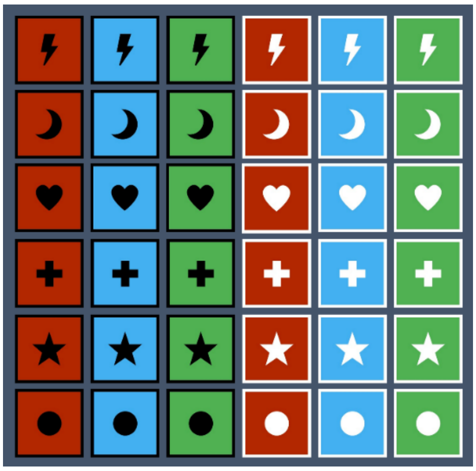


Fig 2. The two sets of cards (Black and White) that consist of six shapes in three colors used as stimuli for the experiment

General Game Procedure

Players were instructed to place cards onto the board as quickly and as accurately as possible within a one-minute trial period. They were required to use right hand only, draw only a single card at a time, and pick exclusively from their assigned card stacks unless otherwise specified by the movement manipulation. The first card of both the players should start in the middle four cells of the game board. Following that, the cards could be placed in any direction but with the new card touching at least one side of the existing card. The overarching goal across all conditions was to create "islands," i.e., orthogonally contiguous groups of cards displaying the same symbol. Diagonal adjacency did not count as part of an island.

Game Sets

The task comprised eight sets produced by crossing three two-level factors: (1) goal coordination, (2) sub-goal coordination, and (3) movement coordination.

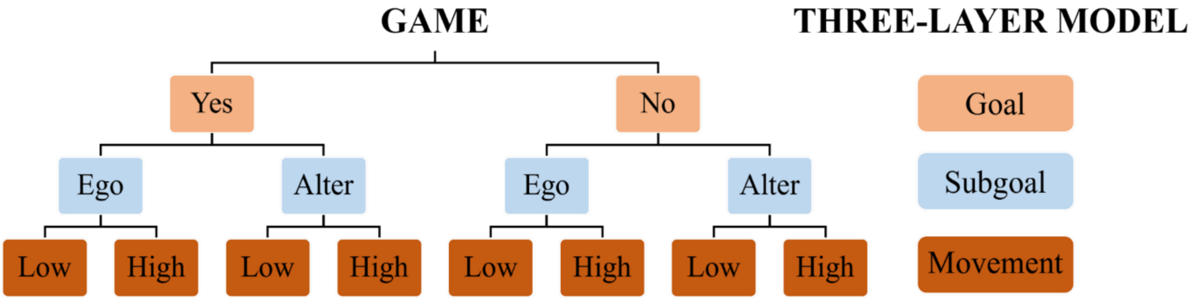


Fig 3. Outline of the game with three layers of Coordination in a 2x2x2 matrix resulting in eight game sets.

Goal Coordination: No Common Goal Vs Common Goal

For no common goal, players worked independently toward maximizing their own islands. "Independent" here means self-focused rather than competitive: participants were instructed not to obstruct each other but simply to optimize placement of their own cards. The maximal island size per symbol in this condition is three (the number of each symbol in a single deck). For common goal, players cooperatively attempted to build the largest possible shared island for each symbol, combining cards across players. The maximum island size in this condition is six cards (three from each player).

Sub-goal Coordination: Ego vs. Alter

This manipulation-imposed color-based adjacency restrictions. In the Ego condition (no sub-goal coordination) players were not allowed to place a card next to another card of the same color in their own rows. They did not need to monitor or avoid the other player's colors. In the Alter condition, where sub-goal coordination was required, players had to avoid placing a card next to an identical colored card belonging to the other player. This required monitoring both players' placements and introduced a need for explicit sub-goal coordination.

Movement Coordination: Low vs. High

In low movement coordination condition, each player's two card stacks were placed in front of them, allowing unobstructed access. During high movement coordination condition, the entire action space was supposed to be shared as the card stacks were swapped, placing each player's decks in front of the other participant. This forced both players to reach over the board to retrieve their own cards, thereby increasing movement interference and necessitating movement coordination.

Scoring and Error Handling

Only the largest island formed for each symbol was scored. Islands required a minimum of two cards (20 points), with 10 points awarded per card. In the common goal condition, the largest possible island size per symbol was six cards (60 points), totaling 360 points for all six symbols, which was then divided evenly between players (maximum 180 points per player). In the No Common Goal condition, the maximum island size per symbol was three cards (30 points), resulting in a maximum individual score of 180 points. If multiple islands of equal size for the same symbol were formed, only one was scored. Errors included violations of color adjacency (violating the Ego or Alter rule) or placing a card on a row of the wrong color (black card on white row, etc.). Erroneous cards were either removed immediately to prevent interference or removed after the trial to exclude them from scoring.

Procedure

The experimental setup involved two players seated opposite each other playing a card game (Fig.1). Prior to the experimental session, participants were contacted via email and asked to confirm that they were not acquainted with the other participant scheduled for the same session. Player 1 arrived approximately one hour earlier than Player 2 to allow sufficient time for EEG preparation. In addition to EEG, eye-tracking data were recorded using a Pupil

Labs Core system; however, these data are not the focus of the present manuscript and are not reported further. There were two experimenters sitting opposite each other in the table with each assigned to one player to look out for rule violations. Each session consisted of four parts and began with the experimenter reading the game instructions aloud, followed by a practice game. The experimental goal structures were clarified after the practice round. Participants were allowed to ask questions between trials but not during task execution. Following the practice round, the experimental phase began.

Each team played two rounds of eight game sets. The order of the game sets was randomized with the constraint that sets with low and high movement coordination were not played consecutively. Prior to each game set, the experimenter read the specific instructions, and participants were informed that instructions would not be repeated to ensure consistent exposure to the rules. Following the instructions, the game commenced with a timer. Participants played for one minute and were signaled to stop by an auditory tone. EEG triggers were recorded at the onset of the instruction and conclusion of each game set. The cards were scored at the end of each game set by the experimenter. After completing the first round of eight sets, participants were given a two-minute break before completing the second round under identical conditions. Upon completion of both rounds, participants were asked to complete a post-session questionnaire. A full experimental session lasted approximately 180 minutes.

EEG Data Acquisition and Preprocessing

EEG activity was recorded from the scalp using 62 Ag/AgCl electrodes integrated into an actiCAP snap electrode cap and collected with BrainVision Recorder software (Brain Products, Gilching, Germany). Electrodes were arranged in accordance with the international 10-20 layout. Electrode impedances were kept below 10 kΩ for the duration of the recording session. The sampling rate was set to 1000 Hz and subjected to an online bandpass filter ranging from 0.1 to 1000 Hz. EEG data were pre-processed offline using EEGLAB toolbox v2023.1 in MATLAB R2023a. The raw data was downsampled to 512 Hz and the EEG channels were re-referenced to the average. A bandpass filter of 0.5 Hz to 40 Hz was applied. Independent Component Analysis (ICA) was applied to the EEG data to identify and remove non-neural artifacts in accordance with the COBIDAS (Committee on Best Practices in Data

Analysis and Sharing) recommendations. And additional visual inspection for further artifact removal was performed.

EEG Time Frequency Analysis

EEG data was analyzed using EEGLAB toolbox on MATLAB 2023a. Continuous data was preprocessed and cleaned using independent component analysis, and only artifact-free channels were retained. EEG signals were epoched time-locked to touching (T) and untouching (U) events belonging to eight base conditions defined by a 2 × 2 × 2 factorial structure of Goal, Subgoal, and Movement, yielding 16 experimental conditions. Time–frequency decomposition was performed separately for each participant and channel using wavelet-based convolution over 1-40 Hz with variable cycle lengths (1-8 cycles). Spectral power was computed as the squared magnitude of complex coefficients. To account for variability in action timing, power was estimated only within the actual duration of each event, excluding time points beyond event offset. Power values were then averaged across trials within condition. Band-limited power was extracted for theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) bands. Condition-wise band power values were saved for each participant and used for group-level statistical analyses.

$$\mathit{Power} \sim \mathit{Goal} * \mathit{Subgoal} * \mathit{Movement} + \mathit{Action\ Boundary} * (\mathit{Goal} + \mathit{Subgoal} + \mathit{Movement}) + (1 \mid \mathit{ID}) + (1 \mid \mathit{Channel}) \quad \text{(Eq. 1)}$$

Behavioural Data Analysis

Game performance, indexed by the player’s score, was analyzed using a linear mixed-effects model. Goal, Subgoal, and Movement were entered as fixed effects in a full factorial design, allowing estimation of all main effects and their interactions on performance. To account for repeated measurements within participants, subject identity (ID) was included as a random effect. The random-effects structure allowed participant-specific variation in the pairwise interactions between Goal and Subgoal, Goal and Movement, and Subgoal and Movement, capturing individual differences in how these factors jointly influenced performance. This modeling approach accounts for the hierarchical structure of the data and controls for non-independence of observations, enabling robust inference on both population-level effects and inter-individual variability in game outcomes.

$$\mathit{Outcome} \sim \mathit{Goal} * \mathit{Subgoal} * \mathit{Movement} + (1 \mid \mathit{ID}) \quad \text{(Eq. 2)}$$

Results

Behavioral Results

The game scores were analyzed across different levels of coupling coordination. There was a significant main effect of Subgoal ($\chi^2(1) = 31.273$, $p < 0.001$, *Cramer's V* = 0.236), indicating that game difficulty increased and scores decreased (*Mean Score* = 105 ± 1.74) when players failed to coordinate their card colors, creating a potential for rule violations compared to absence of Subgoal (*Mean Score* = 87.6 ± 1.65). Similarly, Movement showed a significant main effect ($\chi^2(1) = 7.017$, $p = 0.008$, *Cramer's V* = 0.112), with scores declining when players had to pick cards from the opposite side of the board and coordinate their hand movements with their partner (*Mean Score* = 91.9 ± 1.68) compared to low movement condition when the cards were on their own side (*Mean Score* = 101 ± 1.82)

In contrast, there was neither significant main effect of Goal ($\chi^2(1) = 2.99$, $p = 0.084$, *Cramer's V* = 0.073), nor were there significant two-way interactions between Goal and Subgoal ($\chi^2(1) = 4.357$, $p = 0.428$, *Cramer's V* = 0.034), Goal and Movement ($\chi^2(1) = -1.286$, $p = 0.815$, *Cramer's V* = 0.01), or Subgoal and Movement ($\chi^2(1) = 0.429$, $p = 0.938$, *Cramer's V* = 0.003). Finally, the three-way interaction among Goal, Subgoal, and Movement was not significant ($\chi^2(1) = 6.143$, $p = 0.429$, *Cramer's V* = 0.033).

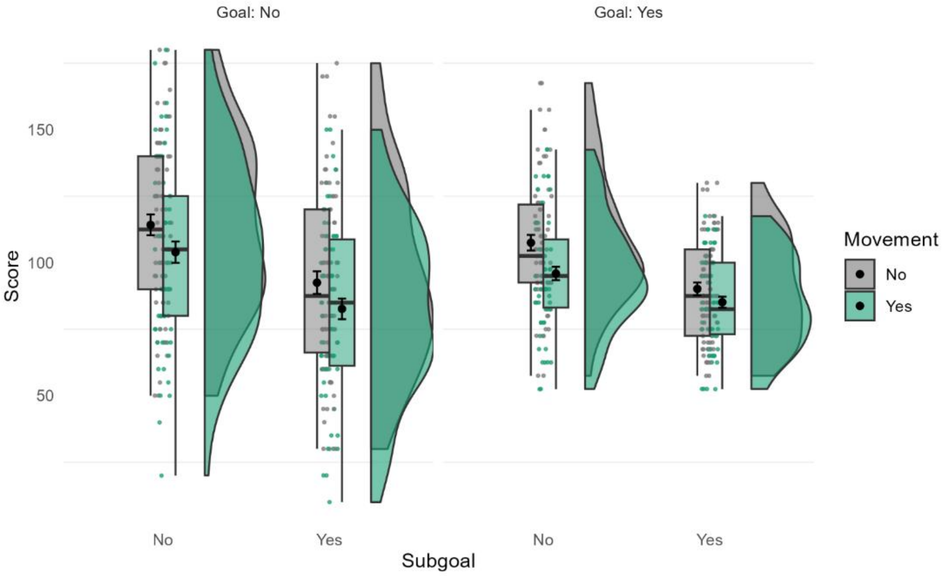


Fig 4. Game scores at different levels of coordination of Goal, Subgoal and Movement Layers

EEG Results

Theta Power

The theta power was analyzed across different coupling conditions similar to the behavioral data analysis. The Subgoal layer had a significant main effect ($\chi^2(1) = 12.161$, $p < 0.001$, *Cramer's V* = 0.019) with theta power decrease when the players coordinated at the subgoal level. Additionally, the action boundary also showed significant main effect ($\chi^2(1) = 98.334$, $p < 0.001$, *Cramer's V* = 0.054) showing theta decrease at the Touching events compared to the Untouching events. In case of interaction effects, Goal and Action Boundary had a significant effect ($\chi^2(1) = 5.894$, $p = 0.015$, *Cramer's V* = 0.013) indicating that the presence of goal coordination further decrease power at the T events. The interaction of Subgoal and Action Boundary also had a significant effect ($\chi^2(1) = 7.931$, $p = 0.005$, *Cramer's V* = 0.015) indicating that the presence of Subgoal coordination increased the power of Theta at the T events.

However, Goal ($\chi^2(1) = 0.631$, $p = 0.427$, *Cramer's V* = 0.004) and Movement ($\chi^2(1) = 1.851$, $p = 0.174$, *Cramer's V* = 0.007) did not show any main effect. Also, the interaction between Goal and Subgoal ($\chi^2(1) = 0.23$, $p = 0.61$, *Cramer's V* = 0.003), Goal and Movement ($\chi^2(1) = 1.761$, $p = 0.184$, *Cramer's V* = 0.007). Movement layer did not have significant interaction with Subgoal ($\chi^2(1) = 3.55$, $p = 0.06$, *Cramer's V* = 0.01) and Type ($\chi^2(1) = 2.008$, $p = 0.156$, *Cramer's V* = 0.008). Furthermore, there was no three way interaction among Goal, Subgoal and Movement ($\chi^2(1) = 0.752$, $p = 0.386$, *Cramer's V* = 0.005).

Alpha Power

The alpha power was analyzed across different coupling conditions, mirroring the approach used for the theta power analysis. The Subgoal layer showed a significant main effect ($\chi^2(1) = 7.323$, $p = 0.007$, *Cramer's V* = 0.015), with alpha power decreasing when players coordinated at the subgoal level. The Movement layer also yielded a significant main effect ($\chi^2(1) = 64.848$, $p < 0.001$, *Cramer's V* = 0.044), though unlike the other factors, it was associated with an increase in alpha power. Additionally, the Action Boundary showed a highly significant main effect ($\chi^2(1) = 172.377$, $p < 0.001$, *Cramer's V* = 0.071), showing a clear alpha power decrease at Touching (T) events compared to Untouching (U) events. Regarding interaction effects, only the interaction between Subgoal and Action Boundary was significant ($\chi^2(1) = 8.907$, $p = 0.003$, *Cramer's V* = 0.016), indicating that the presence of subgoal coordination increased alpha power at the T events (attenuating the general decrease seen at boundaries).

In contrast, several factors did not reach statistical significance. The Goal layer did not show a main effect ($\chi^2(1) = 0.150$, $p = 0.699$, *Cramer's* $V = 0.002$), and there were no significant interactions between Goal and Subgoal ($\chi^2(1) = 0.071$, $p = 0.789$, *Cramer's* $V = 0.001$) or Goal and Movement ($\chi^2(1) = 0.325$, $p = 0.569$, *Cramer's* $V = 0.003$). Similarly, the Movement layer did not significantly interact with Subgoal coordination ($\chi^2(1) = 0.068$, $p = 0.794$, *Cramer's* $V = 0.001$) or the Action Boundary ($\chi^2(1) = 0.746$, $p = 0.388$, *Cramer's* $V = 0.005$). Unlike the theta power results, the interaction between Goal and Action Boundary was not significant for alpha ($\chi^2(1) = 1.519$, $p = 0.218$, *Cramer's* $V = 0.007$). Finally, there was no significant three-way interaction among Goal, Subgoal, and Movement ($\chi^2(1) = 0.059$, $p = 0.807$, *Cramer's* $V = 0.001$).

Beta Power

The beta power was analyzed across different coupling conditions, revealing a more complex interaction structure than the lower frequency bands. The Movement layer showed a highly significant main effect ($\chi^2(1) = 320.34$, $p < 0.001$, *Cramer's* $V = 0.097$), as did the Action Boundary ($\chi^2(1) = 128.92$, $p < 0.001$, *Cramer's* $V = 0.061$). In terms of two-way interactions, beta power was significantly influenced by the interaction between Goal and Movement ($\chi^2(1)= 11.75$, $p < 0.001$, *Cramer's* $V = 0.019$), Subgoal and Movement ($\chi^2(1)= 7.51$, $p = 0.006$, *Cramer's* $V = 0.015$), and Subgoal and Action Boundary ($\chi^2(1) = 8.90$, $p = 0.003$, *Cramer's* $V = 0.016$). Additionally, the interaction between Movement and the Action Boundary was highly significant ($\chi^2(1) = 36.76$, $p < 0.001$, *Cramer's* $V = 0.033$). Most notably, a significant three-way interaction was observed between the Goal, Subgoal, and Movement layers ($\chi^2(1) = 4.45$, $p = 0.035$, *Cramer's* $V = 0.011$), indicating that the joint coordination of top-down goals and subgoals is modulated by physical movement coordination.

Conversely, several factors remained non-significant. The Goal layer ($\chi^2(1) = 0.436$, $p = 0.509$, *Cramer's* $V = 0.004$) and the Subgoal layer ($\chi^2(1) = 1.052$, $p = 0.305$, *Cramer's* $V = 0.006$) did not yield significant main effects. Furthermore, the two-way interaction between Goal and Subgoal ($\chi^2(1) = 1.251$, $p = 0.263$, *Cramer's* $V = 0.006$) was not significant. Similarly, the interaction between the Goal layer and Action Boundary failed to reach significance ($\chi^2(1) = 1.023$, $p = 0.312$, *Cramer's* $V = 0.005$).

The effects of coordination across layers varied by frequency band. Goal alone rarely influenced power but interacted with Action Boundary in theta and with Movement in beta, contributing to a three-way Goal × Subgoal × Movement effect (Fig. 5). Subgoal consistently modulated theta and alpha, decreasing power and interacting with Action Boundary, while also shaping beta through interactions with Movement and Action Boundary (Fig. 6). Movement had minimal impact on lower frequencies but strongly influenced beta power, both as a main effect and through interactions with Goal, Subgoal, and Action Boundary, highlighting its key role in complex, action-based coordination (Fig. 7).

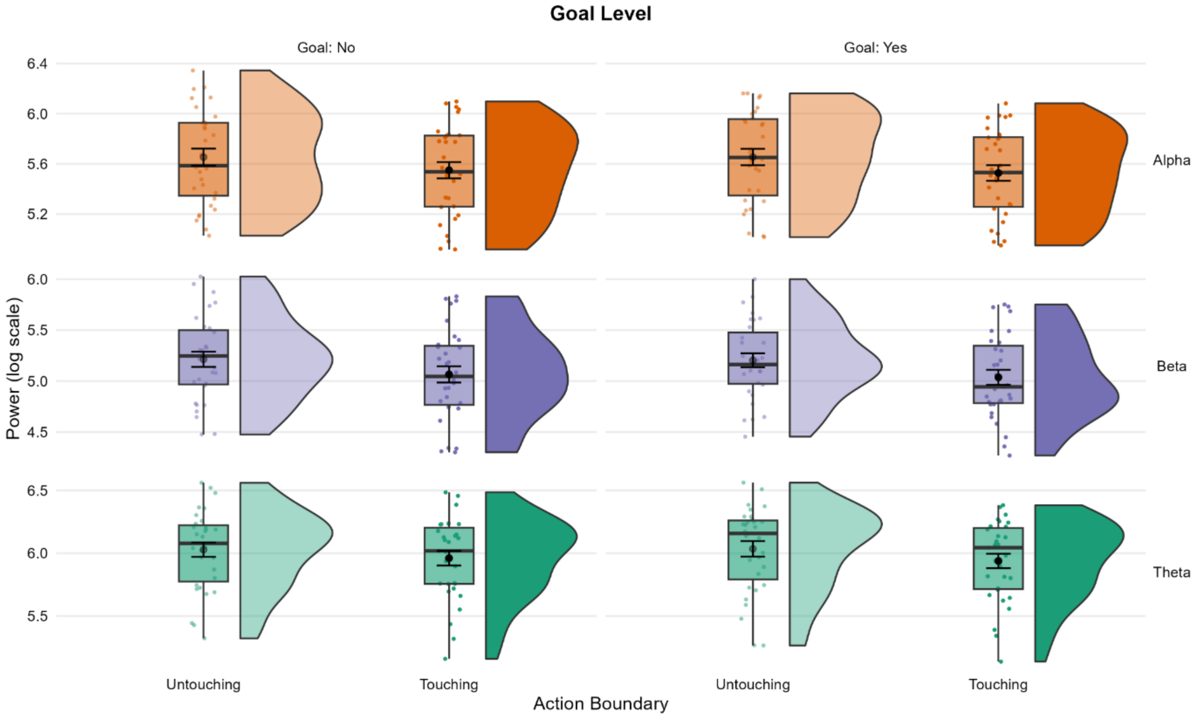


Fig 5. EEG Band Power differences at Touching and Untouching Events in the Goal Layer

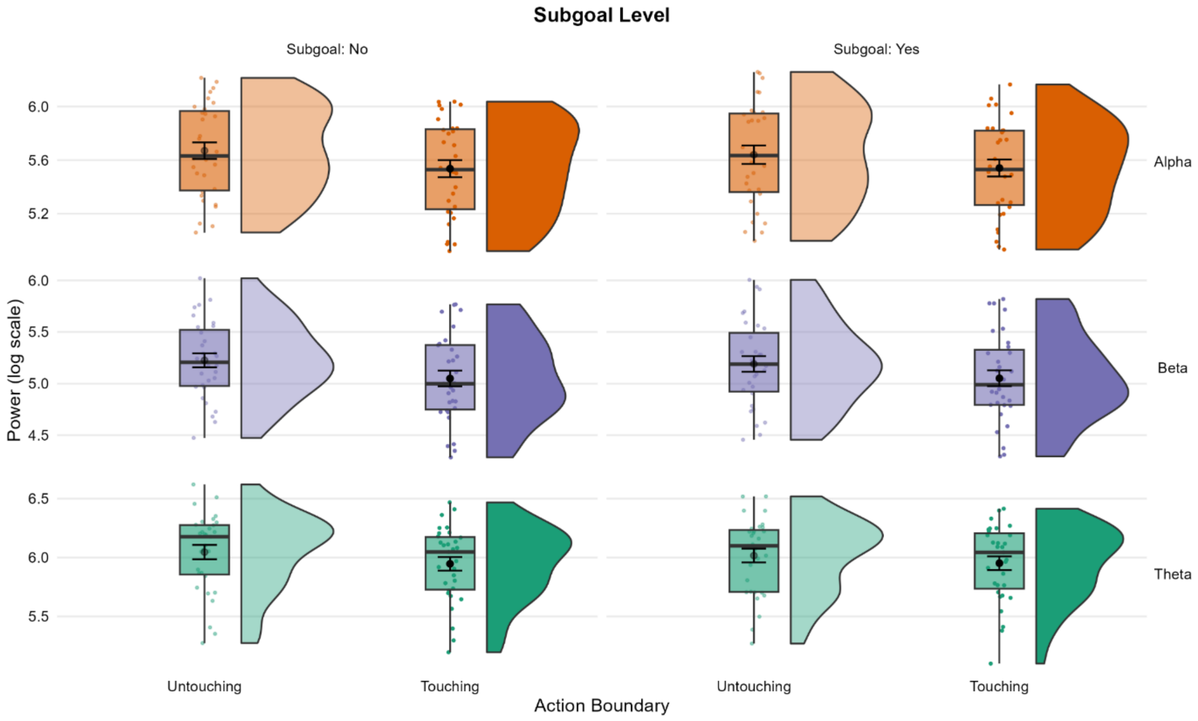


Fig 6. EEG Band Power differences at Touching and Untouching Events in the Subgoal Layer

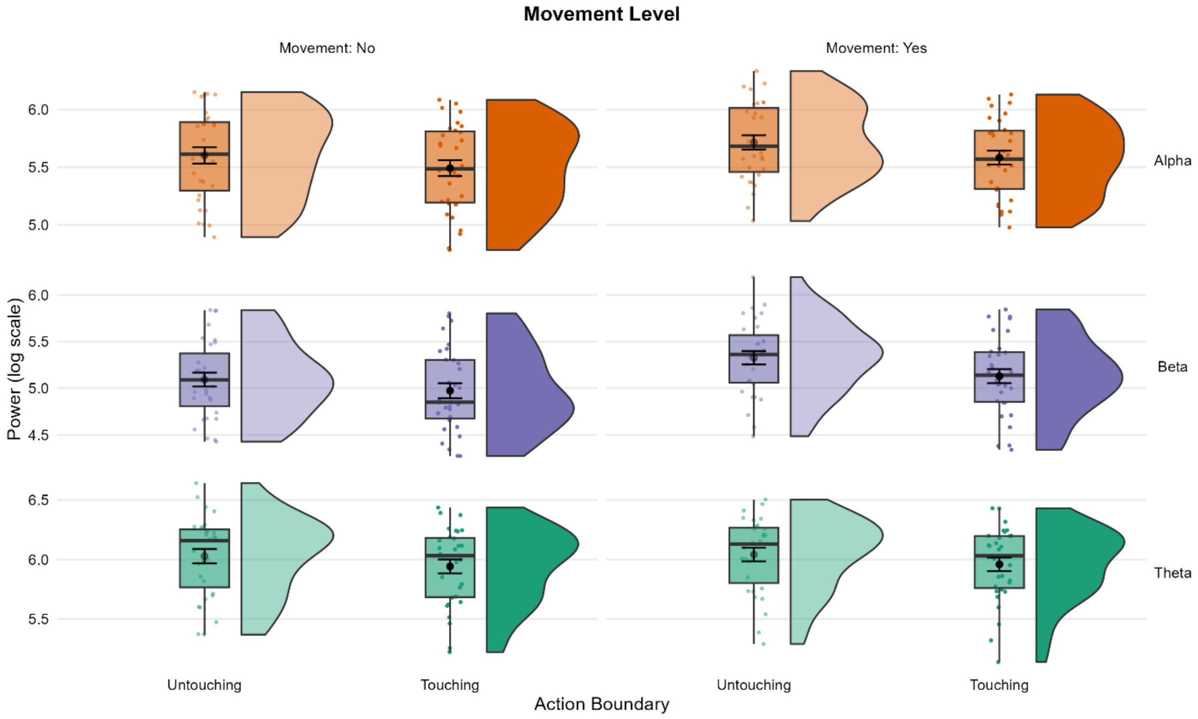


Fig 7. EEG Band Power differences at Touching and Untouching Events in the Movement Layer

Discussion

The current study investigated how hierarchical coordination demands in joint action shape neural dynamics at action boundaries, using a naturalistic, interactive game experiment. By orthogonally manipulating goal, subgoal, and movement coordination, the task dissociated multiple layers of shared intentional structure while preserving continuous, real-world action. EEG analyses revealed frequency-specific signatures of coordination at the action boundaries. Theta power decreased at action boundaries, with this suppression being further amplified when participants shared a common goal, but attenuated when subgoal coordination was required, suggesting differential roles of high-level and intermediate predictive constraints. Alpha power showed a similar boundary-related decrease, modulated selectively by subgoal coordination and movement demands. In contrast, beta power exhibited strong sensitivity to movement coordination and higher-order interactions among goal, subgoal, and movement layers. Together, these findings indicate that joint action is hierarchically structured and differentially expressed across neural frequency bands, reflecting distinct functional contributions to action monitoring and coordination.

Theta-band oscillations are widely implicated in various cognitive functions such as sensory processing, attentional processing, spatial navigation and memory (Karakas, 2020). The reduced theta power when the players had to coordinate at the subgoal likely reflects diminished demands on cognitive control and working memory once shared task representations are established, making joint action planning more efficient (Cavanagh & Frank, 2014; Sauseng et al., 2010). Similarly, the strong reduction in theta power at Touching events compared to Untouching events suggests that theta oscillations are sensitive to action boundaries, with Touching events marking moments of planning completion, hand touching a card to be transported, where uncertainty and predictive demands are minimized. Given the role of theta in prediction error processing and behavioral adaptation (Billeke et al., 2013; Huang et al., 2015), this decrease may reflect a transient release from control and predictive processing once outcomes are resolved. The interaction between Goal coordination and Action Boundary further supports this view, as shared high-level goals appear to enhance predictability at critical action boundaries, thereby reducing the need for control-related theta engagement.

In contrast, an increase in theta power at Touching events when the players are to be coordinated at the subgoal coordination may reflect enhanced attentional engagement or updating processes at meaningful coordination checkpoints (Williams et al., 2019). While subgoal alignment reduces overall control demands, touching events (T) under subgoal coordination may represent salient moments for confirming joint progress and updating shared task states. These could be the time points where the players recall the rule of the specific game set during which the working memory comes into play (Pomper & Ansorge, 2021), and after selecting the card, there is a high attentional engagement to place the card at the right place in the game board without violation of rules resulting in high cognitive load. Taken together, these results indicate that theta-band activity does not uniformly increase or decrease during joint action but is flexibly modulated by the hierarchical level of coordination and by action boundaries.

Alpha-band oscillations are generally understood as a key mechanism for functional inhibition and attentional regulation in the brain. Rather than reflecting cortical idling, contemporary accounts propose that alpha activity actively suppresses task-irrelevant information and gates sensory, cognitive, and motor processing in a context-dependent manner (Jensen & Mazaheri, 2010; Klimesch, 2012). The decrease in alpha power is typically associated with increased attentional engagement, information processing, and readiness for action, whereas increases in alpha power are thought to reflect inhibitory control, reduced processing demands, or the shielding of ongoing processes from interference (Foxe & Snyder, 2011).

The present findings largely align with this functional framework while also highlighting the flexibility of alpha modulation in joint action. The decrease in alpha power during Subgoal coordination is consistent with the notion that aligning at an intermediate hierarchical level requires enhanced attentional engagement and integration of shared task representations, thereby reducing inhibitory alpha activity. Similarly, the pronounced alpha suppression at Touching events supports the interpretation of action boundaries as moments of heightened salience, requiring increased monitoring, outcome evaluation, and confirmation of joint progress processes commonly associated with alpha desynchronization.

In contrast, the increase in alpha power associated with the Movement layer diverges from a purely attention-based interpretation and instead supports accounts linking alpha enhancement to the inhibition of higher-order cognitive processing during continuous action execution. When coordination is driven at the movement level, increased alpha power may reflect the suppression of unnecessary cognitive or sensory input, allowing smooth motor coordination with minimal interference. Finally, the interaction between Subgoal coordination and Action Boundary, showing elevated alpha power at Touching events when subgoal coordination is present, suggests that shared subgoal representations reduce the attentional demands typically associated with action boundaries. In this case, alpha activity may index efficient predictive processing and reduced need for attentional reallocation, attenuating the general boundary-related alpha decrease.

Beta-band oscillations are classically associated with motor control, sensorimotor integration, and the maintenance of the current cognitive–motor state. Rather than signaling movement per se, beta activity is often interpreted as reflecting the stabilization of ongoing motor and cognitive sets, with beta suppression accompanying movement initiation or change, and beta enhancement indexing the maintenance or prediction of the current state (Engel & Fries, 2010; Kilavik et al., 2013). In social and joint action contexts, beta oscillations have also been linked to action prediction, coordination, and the integration of top-down goals with bottom-up motor execution (Czeszumski et al., 2020).

The present results strongly align with this functional characterization of beta activity. The main robust effect of the Movement layer and the Action Boundary suggests that beta power is particularly sensitive to sensorimotor demands and transitions between action states. Modulation of beta at action boundaries likely reflects updating or reconfiguration of the current motor set, consistent with evidence that beta oscillations decrease when actions are adjusted or new motor plans are required. The absence of main effects for the Goal and Subgoal layers further supports the idea that beta activity is not primarily driven by abstract task structure in isolation, but rather by how these representations are implemented through movement.

Crucially, the pattern of two-way and three-way interactions indicates that beta oscillations serve as a neural interface between hierarchical task representations and physical coordination. Interactions between Movement and both Goal and Subgoal layers suggest that top-down action plans modulate beta activity only when they are expressed through coordinated movement. The significant interaction between Subgoal and Action Boundary further implies that beta dynamics are sensitive to how intermediate action plans are updated at salient event boundaries. Most notably, the three-way interaction between Goal, Subgoal, and Movement layers highlights that beta power reflects the integrated coordination of abstract goals and subgoals with ongoing motor execution, rather than any single level of the hierarchy alone.

The lack of significant effects involving Goal or Subgoal coordination in the absence of movement-related factors underscores a key distinction between beta and lower-frequency bands. Unlike theta and alpha oscillations, which appear more closely tied to cognitive control and attentional demands, beta activity in the present data predominantly indexes the maintenance, updating, and integration of coordinated motor states in joint action. Overall, these findings support the view that beta-band oscillations play a central role in binding hierarchical task structure to sensorimotor coordination, enabling stable yet flexible joint action.

Conclusion

The study demonstrates that neural oscillations across theta, alpha, and beta bands are differentially modulated by the hierarchical structure of joint action. Theta and alpha activity primarily reflected cognitive control, attention, prediction, and event segmentation, showing sensitivity to goal- and subgoal-level coordination and action boundaries. In contrast, beta oscillations were dominated by movement-related modulation and by interactions between hierarchical task representations and sensorimotor coordination. Together, these findings suggest that successful joint action relies on a frequency-specific division of labor, in which lower-frequency oscillations support shared representations and adaptive control, while beta-band dynamics integrate top-down goals with ongoing movement, enabling flexible yet stable interpersonal coordination.

CRediT authorship contribution statement

Rosari Naveena Selvan: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Anoushiravan Zahedi:** Writing – review & editing, Formal analysis, Validation, Methodology. **Minghao Cheng:** Writing – original draft, Software, Methodology, Data curation. **Katharina Dohr:** Methodology, Data curation, Writing – original draft. **Minija Tamosiunaite** – Writing – review & editing, Resources, Methodology. **Florentin Wörgötter:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Ricarda I. Schubotz:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Funding

This work was supported by the German Research Foundation (DFG) [WO 388/13–1]. The funder had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report.

Acknowledgments

We extend our sincere gratitude to Monika Mertens, Lena Innig, Thea Bröring, Anna Asdonk, Maike Denguth, Pia Niedrée and Yannah Schröder for helping to recruit participants and/or assisting during data collection for the study.

References

Aksoy, E. E., Abramov, A., Dörr, J., Ning, K., Dellen, B., & Wörgötter, F. (2011). Learning the semantics of object-action relations by observation. *International Journal of Robotics Research*, *30*(10), 1229–1249. https://doi.org/10.1177/0278364911410459

Billeke, P., Zamorano, F., Cosmelli, D., & Aboitiz, F. (2013). Oscillatory brain activity correlates with risk perception and predicts social decisions. *Cerebral Cortex*, *23*, 2872–2883. https://doi.org/10.1093/cercor/bhs269

Cavanagh, & Frank, M. J. (2014). Frontal Theta as a Mechanism for Affective and Effective Control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012.Frontal

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cooper, S., Fensome, S. F. P., Kourtis, D., Gow, S., & Dragone, M. (2020). An EEG investigation on planning human-robot handover tasks. *Proceedings of the 2020 IEEE International Conference on Human-Machine Systems, ICHMS 2020*. https://doi.org/10.1109/ICHMS49158.2020.9209543

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., & König, P. (2020). Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. *Frontiers in Human Neuroscience*, *14*, 1–17. https://doi.org/10.3389/fnhum.2020.00039

Engel, A. K., & Fries, P. (2010). Beta-band oscillations-signalling the status quo? *Current Opinion in Neurobiology*, *20*(2), 156–165. https://doi.org/10.1016/j.conb.2010.02.015

Fanghella, M., Mussini, E., Zazio, A., Genovese, F., Satta, E., Barchiesi, G., Battaglia-Mayer, A., Bortoletto, M., & Sinigaglia, C. (2025). Distinct neural dynamics in joint versus side-by-side actions: insights from dual EEG. *Social Cognitive and Affective Neuroscience*, *20*(1), 1–9. https://doi.org/10.1093/scan/nsaf075

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, *2*, 1–13. https://doi.org/10.3389/fpsyg.2011.00154

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K., & Frith, C. (2015). A Duet for one. *Consciousness and Cognition*, *36*, 390–405. https://doi.org/10.1016/j.concog.2014.12.003

Friston, K. J., & Frith, C. D. (2015). Active inference, Communication and hermeneutics. *Cortex*, *68*, 129–143. https://doi.org/10.1016/j.cortex.2015.03.025

Friston, K., Square, Q., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, *104*, 137–160. https://doi.org/10.1007/s00422-011-0424-z.Action

Huang, Y., Chen, L., & Luo, H. (2015). Behavioral oscillation in priming: Competing perceptual predictions conveyed in alternating theta-band rhythms. *Journal of Neuroscience*, *35*(6), 2830–2837. https://doi.org/10.1523/JNEUROSCI.4294-14.2015

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*, 1–8. https://doi.org/10.3389/fnhum.2010.00186

Kahl, S., & Kopp, S. (2018). A predictive processing model of perception and action for self-other distinction. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02421

Karakas, S. (2020). A review of theta oscillation and its functional correlates. *International Journal of Psychophysiology*, *157*, 82–99. https://doi.org/https://doi.org/10.1016/j.ijpsycho.2020.04.008

Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A., & Riehle, A. (2013). The ups and downs of beta oscillations in sensorimotor cortex. *Experimental Neurology*, *245*, 15–26. https://doi.org/10.1016/j.expneurol.2012.09.014

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Kourtis, D., Sebanz, N., & Knoblich, G. (2013). Predictive representation of other people's actions in joint action planning: An EEG study. *Social Neuroscience*, *8*(1), 31–42. https://doi.org/10.1080/17470919.2012.694823

Kourtis, D., Woźniak, M., Sebanz, N., & Knoblich, G. (2019). Evidence for we-representations during joint action planning. *Neuropsychologia*, *131*, 73–83. https://doi.org/10.1016/j.neuropsychologia.2019.05.029

Markiewicz, R., Segaert, K., & Mazaheri, A. (2024). Brain-to-brain coupling forecasts future joint action outcomes. *IScience*, *27*(9), 110802. https://doi.org/10.1016/j.isci.2024.110802

Novembre, G., Sammler, D., & Keller, P. E. (2016). Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action. *Neuropsychologia*, *89*, 414–425. https://doi.org/10.1016/j.neuropsychologia.2016.07.027

Pesquita, A., Whitwell, R. L., & Enns, J. T. (2018). Predictive joint-action model: A hierarchical predictive approach to human cooperation. *Psychonomic Bulletin and Review*, *25*(5), 1751–1769. https://doi.org/10.3758/s13423-017-1393-6

Pomper, U., & Ansorge, U. (2021). Theta-Rhythmic Oscillation of Working Memory Performance. *Psychological Science*, *32*(11), 1801–1810. https://doi.org/10.1177/09567976211013045

Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1015–1022. https://doi.org/10.1016/j.neubiorev.2009.12.006

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Shiraishi, M., & Shimada, S. (2021). Inter-brain synchronization during a cooperative task reflects the sense of joint agency. *Neuropsychologia*, *154*, 107770. https://doi.org/10.1016/j.neuropsychologia.2021.107770

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*, 998–1003. https://doi.org/10.1016/j.neunet.2010.06.002

Williams, C. C., Kappen, M., Hassall, C. D., Wright, B., & Krigolson, O. E. (2019). Thinking theta and alpha: Mechanisms of intuitive and analytical reasoning. *NeuroImage*, *189*, 574–580. https://doi.org/10.1016/j.neuroimage.2019.01.048

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 593–602. https://doi.org/10.1098/rstb.2002.1238

Wörgötter, F., Aksoy, E. E., Krüger, N., Piater, J., Ude, A., & Tamosiunaite, M. (2013). A simple ontology of manipulation actions based on hand-object relations. *IEEE Transactions on Autonomous Mental Development*, *5*(2), 117–134. https://doi.org/10.1109/TAMD.2012.2232291

4. General Discussion

4.1. Summary of the Presented Studies

Study I Selvan et al., (2024) investigated the neural mechanisms underlying action prediction and updating during the observation of structured action sequences, with the aim of clarifying how the brain anticipates upcoming actions based on learned regularities. Grounded in predictive processing theories, the study addressed the question of whether action prediction unfolds dynamically across distinct neural stages rather than as a single anticipatory response. Participants observed sequences of hand–object interactions composed of meaningful action segments, such as touching and releasing objects, which were arranged according to probabilistic transition rules. Over time, observers implicitly learned these transition probabilities, allowing them to form expectations about which action segment was likely to occur next. fMRI was used to measure neural responses at key moments in the action sequences, particularly at the boundaries between segments where predictions could be generated and updated.

The findings distinguished between two core processes involved in action prediction: the retrieval of possible upcoming actions and the selection of a specific prediction. When an observed action segment ended, multiple next actions were often plausible given the learned sequence structure. Neural activity in frontal regions, including lateral prefrontal cortex, was associated with this retrieval phase, suggesting that these areas support the representation of abstract action possibilities based on statistical knowledge. In contrast, activity in midline cortical regions and the hippocampus was linked to the selection of a specific upcoming action, indicating a role for these structures in binding predictions to sequential contexts. This dissociation provides evidence that action prediction is hierarchically organized, with different brain regions contributing to distinct inferential stages.

In addition to prediction generation, the study examined neural responses to unexpected action transitions. When observed actions violated learned probabilities, increased activation was found in regions associated with the action observation network, reflecting prediction error processing and model updating. These findings align with predictive coding accounts in which mismatches between expected and actual input drive learning and adaptation. Importantly, the observed neural effects could not be explained solely by low-level visual differences, underscoring that the brain's responses were sensitive to learned action structure rather than to sensory novelty alone.

Overall, **Study 1** demonstrates that action prediction during observation is an active and dynamic process supported by hierarchical neural mechanisms. The study advances existing models of action understanding by showing that the brain does not merely predict a single future action, but instead temporarily represents multiple possible outcomes before converging on a specific expectation. By integrating learning, memory, and predictive processing, this work provides a refined account of how humans anticipate others' actions and adapt their internal models in response to changing action sequences.

Study 2 (Selvan et al. in review) investigated how eye movements reflect prediction errors during the observation of structured, multi-step action sequences. While previous research has shown that gaze behavior anticipates upcoming actions and reacts to surprising events, little was known about the fine-grained temporal dynamics of oculomotor responses when a prediction error occurs within an ongoing action sequence. To address this gap, we employed a controlled yet ecologically valid action-observation paradigm in which participants first learned probabilistic action sequences and later observed both familiar sequences and "illegal" variants. In these illegal sequences, a single action step was substituted at a specific moment to violate learned expectations and induce a prediction error.

Eye tracking and pupillometry were analyzed around precisely defined action boundaries, which were identified using computer vision algorithms. The study focused on multiple oculomotor measures, including fixation count, fixation duration, saccade velocity, and pupil dilation, to capture how prediction errors influence visual sampling behavior. The results revealed a consistent and coordinated pattern of gaze responses to the points of violation. Prediction errors were associated with increased pupil dilation, suggesting heightened arousal and engagement of neuromodulatory systems such as the noradrenergic system. At the same time, saccade velocity increased, indicating intensified exploratory eye movements under conditions of uncertainty.

Fixation behavior also changed markedly during prediction errors. Participants exhibited shorter fixation durations and a higher number of incorrect fixations, reflecting a shift away from stable, anticipatory gaze toward rapid visual exploration. This pattern suggests that when expectations are violated, the visual system reallocates attentional resources to sample the environment more broadly to resolve uncertainty and update internal models. Importantly, these effects were time-locked to action boundaries rather than being driven by low-level visual novelty, supporting the interpretation that they reflect prediction error processing rather than simple stimulus-driven responses.

Overall, the study demonstrates that prediction errors during action observation are expressed through distinct oculomotor signatures that integrate attentional reorientation, arousal, and exploratory behavior. By showing how gaze dynamics change moment-by-moment when expected action trajectories break down, the study provides compelling evidence that eye movements serve as a sensitive behavioral marker of predictive processing. The findings support hierarchical predictive coding accounts of action perception, in which violations of expectation trigger increased uncertainty, drive model updating, and alter perceptual sampling strategies in real time.

Study 3 (Selvan et al. to be submitted) examined the neural dynamics of predictive processing during interactive joint action using electroencephalography (EEG). Whereas Studies 1 and 2 focused on prediction during passive action observation, Study 3 addressed how predictive mechanisms operate when individuals must coordinate their actions with another agent in real time. Specifically, the study investigated how violations at different hierarchical levels of joint action goal, sub-goal, and movement coordination are reflected in oscillatory brain activity.

Building directly on the hierarchical predictive framework established in Studies 1 and 2, Study 3 extends the investigation from observing structured action sequences to actively participating in them. By retaining the focus on prediction violations at well-defined action boundaries while embedding these violations within real-time interpersonal coordination, the study tests whether predictive mechanisms identified during action observation generalize to interactive, goal-directed behavior. In doing so, Study 3 bridges perceptual and interactive domains of predictive processing, offering a critical extension of earlier findings into social action contexts.

Participants performed a cooperative joint-action task in which they were required to coordinate movements with a partner to achieve shared goals under varying constraints. The task systematically manipulated whether partners shared the same goal, sub-goal, or movement trajectory, thereby inducing prediction violations at distinct hierarchical levels of action representation. This design allowed the authors to dissociate prediction-related neural responses associated with abstract goal representations from those linked to lower-level motor coordination.

EEG activity was analyzed around precisely defined action boundaries, identified using computer vision algorithms to ensure accurate temporal alignment between neural signals and behavior. Time–frequency analyses focused on oscillatory power changes across theta, alpha, and beta frequency bands, which have been implicated in cognitive control, prediction, and sensorimotor processing. The results revealed that prediction violations at different hierarchical levels elicited dissociable oscillatory signatures. Coordination of higher-level goal and sub-goal layers were associated with increased theta-band activity, consistent with enhanced cognitive control and the need to update internal models. In contrast, violations at the movement coordination level primarily modulated beta-band activity, reflecting disruption and recalibration of sensorimotor predictions.

In addition, alpha-band modulation was observed during conditions requiring increased attentional monitoring and suppression of irrelevant action tendencies, suggesting a role for attentional gating during joint prediction. These frequency-specific effects were time-locked to action boundaries, supporting the interpretation that they reflect prediction updating processes rather than ongoing motor execution alone.

Overall, Study 3 demonstrates that predictive processing during joint action is hierarchically organized and dynamically instantiated in distinct neural oscillatory regimes. By showing that violations at different levels of shared action structure elicit separable EEG signatures, the study provides strong evidence that prediction in social interaction operates across multiple representational levels. The findings extend predictive coding accounts from action perception to real-time interpersonal coordination, highlighting how neural prediction mechanisms support flexible and adaptive joint behavior.

4.2. Predictive Processing at Action Boundaries

Predictive processing has reshaped contemporary views of brain function by proposing that the nervous system continuously anticipates sensory inputs and minimizes the mismatch between top-down predictions and bottom-up evidence. Formally grounded in Bayesian inference and the free energy principle, this framework posits that neural computations across hierarchies aim to minimize prediction error, thereby reducing uncertainty about incoming signals (Sprevak & Smith, 2023). Although originally formulated to explain sensory perception in visual and auditory domains, predictive processing has been increasingly extended to encompass sensorimotor integration, motor control, and action cognition. The predictive brain not only interprets sensory data but actively anticipates the consequences of motor acts, structuring both perception and action as inferential, model-based processes rather than reactive responses (E. C. Brown & Brüne, 2012).

Within the domain of action perception and execution, the concept of hierarchical prediction is particularly salient. Natural behavior consists of structured sequences of motor acts, each with its own goals, dynamics, and temporal boundaries. Observers and actors alike extract these structures to anticipate what will occur next, thereby facilitating fluent interaction with the environment. Because predictive models guide expectations about which actions are likely to follow, they must be continuously updated as sequences unfold. Crucially, action boundaries, temporal junctures where one motor segment ends and the next begins representing moments of increased uncertainty that demand active model updating. At such boundaries, ongoing predictions about a completed action must be terminated and replaced with new predictions about forthcoming actions.

Study 1 investigated the neural responses at these action boundaries by objectively defining the end of an action as Untouching (U) hand releasing the object and beginning of an action as Touching (T) hand touching the object using computer vision. The *Legal Action Sequences,* the ones that participants were trained for were included for the analysis of predictive processing at the action boundaries. At the U events, one of the prominent activations seen is in the pre-Supplementary Motor Area (pre-SMA) and SMA, that are majorly responsible for motor planning and execution. The functional role of SMA is attributed to performing action sequences and pre-SMA to the changing of the sequences (Nachev et al., 2008). Additionally, pre-SMA activation was also reported when the attention was given to the intention of others (Lohse, 2006). Another distinct difference seen is the higher activation of pre-SMA for internally generated actions (Cunnington et al., 2002). On the other hand, the SMA is activated during the observation of graspable objects (Grèzes & Decety, 2002) on the underlying phenomenon called the *object affordance,* that activates the brain region to generate a motor action with respect to the object. A study further separated the role of SMA and pre-SMA in temporal processing, with activation of pre-SMA related to sensory, non-sequential, and supra-second temporal processing and SMA proper to sensorimotor, sequential, and sub-second temporal processing (Schwartze et al., 2012). The activation of the pre-SMA/SMA observed at the U events in **Study 1** could reflect the prediction of action coming to an end, matching the intention of the actor along with the releasing of the object that they learnt and formed the internal model during imitation training.

Another strong activation observed was at the Putamen and Insula that constitutes the Striatum which along with SMA are known to play a key role in arranging the sequences in the order in which they are to be executed (Bednark et al., 2015; Shima & Tanji, 2000; Tanji, 2001). The striatum serves to integrate individual actions into coherent, organized behavioral units, a process known as chunking (Favila et al., 2023; Graybiel, 1998). Additionally, the activation was seen at the Lingual gyrus, which is known to be responsible for spatial memory with respect to the scene perception (Kamps et al., 2016; Ruotolo et al., 2019; Sulpizio et al., 2013b). In this study, the scene being the action space where the action sequences were performed, at the U events, it was important to perceive the location of the objects and recall the next action step to be performed.

Though, at the probabilistic branching structure, the brain activation did not survive the multiple correction, it is important to discuss the activation of the Anterior Cingulate Cortex (ACC). The anterior cingulate cortex (ACC) plays a central role in error monitoring and prediction updating, particularly when strong prior expectations bias the internal model toward a specific outcome. Empirical studies show that ACC activity increases when outcomes violate learned expectations or when prediction uncertainty is high, reflecting its role in evaluating the reliability of internal models (Holroyd & Coles, 2002). Notably, ACC responses are amplified when the system is strongly committed to a particular prediction, as prediction errors under such conditions carry greater informational value (J. W. Brown & Braver, 2005). This function supports adaptive behavioral adjustments by triggering model revision and increased cognitive control following unexpected outcomes (Shenhav et al., 2013).

At the second action boundary T, there was activation of cortical midline structures medial prefrontal cortex (MPFC) and the posterior cingulate cortex (PCC) are increasingly understood as key nodes for high-level prediction within hierarchical predictive processing frameworks. The MPFC is responsible for encoding abstract, context-dependent predictions derived from prior experience, integrating information from memory systems to bias expectations about future events and action outcomes (Bar, 2007; Clark, 2013). This role is especially prominent in situations requiring the use of learned regularities or schemas to anticipate what will happen next (Schacter et al., 2017). The PCC, in contrast, has been proposed to support the monitoring and updating of internal models by evaluating the congruence between ongoing predictions and incoming evidence, thereby signaling when predictions should be revised (Pearson et al., 2012). Together, the MPFC and PCC form core components of the DMN, which is increasingly conceptualized not as task-negative but as centrally involved in internally driven simulation, anticipation, and model updating (Buckner et al., 2008; K. Friston, 2010).

Furthermore, at T events, hippocampus was activated, studies have shown the area to be sensitive to action boundaries (Ben-Yakov & Dudai, 2011; R. A. Cooper & Ritchey, 2020). The hippocampus has been increasingly characterized as a match-mismatch detector that supports predictive processing by comparing incoming sensory information with internally generated expectations (Duncan et al., 2012; Kumaran & Maguire, 2007; Sinclair et al., 2021). This function is especially critical for action prediction, where anticipating the next action segment depends on stored knowledge of temporal and causal regularities. Empirical studies show that hippocampal activity increases when expected sequences are violated, consistent with a role in detecting prediction errors at higher, relational levels rather than low-level sensory discrepancies (Chen et al., 2011). Moreover, hippocampal representations support prospective simulation of future actions, enabling observers to anticipate upcoming movements based on prior experience (Schapiro et al., 2016; Stachenfeld et al., 2017). In the context of action boundaries, hippocampal engagement reflects the evaluation of whether current observations match predicted action trajectories and the updating of internal models when mismatches occur.

Overall, the pattern of neural responses observed across the two action boundaries suggests that predictive processing during action observation unfolds as a two-step process distributed across the end and beginning of action segments. At the U events, corresponding to action termination, the engagement of pre-SMA, SMA, striatum, insula, and lingual gyrus indicates a first predictive step centered on evaluation and closure of the ongoing action. This stage likely involves monitoring whether the unfolding action matches the learned internal model, integrating sensory evidence with motor sequence representations, and signaling the completion of a predicted action outcome. In contrast, the T events appear to reflect a second predictive step focused on prospective model instantiation, supported by the recruitment of cortical midline structures and the hippocampus. Here, higher-level contextual and mnemonic information is used to generate predictions about the upcoming action segment, selecting among possible continuations based on learned probabilistic structure. Together, these findings support a temporally ordered predictive mechanism in which action termination triggers confirmation or revision of current predictions, while action initiation engages anticipatory processes that prepare the system for future action states, enabling fluent perception of complex action sequences.

4.3. Neural and Oculomotor Signatures of Prediction Error

Prediction error is a core computational signal in predictive processing frameworks, defined as the discrepancy between internally generated predictions and actual sensory or motor input, and it plays a central role in driving learning and model updating across the brain. When predictions fail, prediction error signals propagate through hierarchical neural systems to revise internal models, thereby optimizing future perception and action. In the domain of action observation, prediction errors emerge when observed movements deviate from expected kinematic, temporal, or goal-related structures, leading to increased activity in sensorimotor, control, and attentional networks (Kilner et al., 2007; Press et al., 2011). Importantly, predictive mechanisms are tightly coupled with oculomotor behavior, as eye movements are proactively guided by expectations about upcoming action outcomes. Gaze is often directed ahead of observed actions, reflecting anticipatory predictions, whereas violations of expectation elicit delayed saccades, increased fixations, or gaze corrections that mirror neural prediction error responses (Elsner et al., 2013; Flanagan & Johansson, 2003). Together, neural and oculomotor measures provide complementary windows into how prediction errors shape dynamic action understanding.

Neuroimaging studies have provided converging evidence that the action observation network (AON) supports predictive processing by encoding expected kinematic, temporal, and goal-related features of observed actions. fMRI research has shown that the posterior superior temporal sulcus (pSTS) is selectively sensitive to biological motion and signals violations of predicted movement trajectories, making it a key site for visual prediction error during action perception (Jastorff et al., 2011; Rens et al., 2020; Stehr et al., 2021). The inferior parietal lobule (IPL) integrates visual input with stored motor knowledge and is preferentially activated when observed actions deviate from expected goals or object-related outcomes, reflecting higher-level, goal-based prediction errors (Fogassi et al., 2005; A. F. D. C. Hamilton & Grafton, 2006). A recent fMRI study (Pomp et al., 2025) showed that the aIPS activated is associated with the Touching and Untouching structures that are crucial time points for predictive processing. Additionally, premotor regions, particularly the ventral premotor cortex (PMv), show increased activation when predicted action outcomes are violated, consistent with their role in mapping observed actions onto internal motor representations (Cross et al., 2013; Press et al., 2011). Effective connectivity analyses further demonstrate that prediction error enhances feedback from premotor and parietal regions to visual cortices, supporting rapid updating of action models (Urgen & Saygin, 2020). Together, these findings indicate that the AON is functionally organized to detect and resolve prediction errors across multiple representational levels during action observation. Thus, the findings from the *Illegal Action Sequences* of **Study 1** served as sanity check to the neural responses to prediction error.

Following that, **Study 2** investigated the oculomotor change after encountering the prediction error in a probabilistically structured action sequence. Oculomotor behavior provides a sensitive behavioral marker of predictive processing and prediction error during action perception. Eye movements are not purely reactive but are proactively guided by internal models that anticipate future sensory states, allowing gaze to be directed toward expected locations of upcoming events. During action observation, observers typically generate predictive saccades to future action targets, such as the goal of a reaching or grasping movement, reflecting successful action prediction (Gredebäck & Falck-Ytter, 2015b; Vannuscorps & Caramazza, 2017). When observed actions violate expectations, prediction errors are expressed behaviorally as delayed saccades, increased fixation durations, or corrective eye movements, indicating the need to update internal models (Rotman et al., 2006;

Wills et al., 2007). Further studies suggest that prediction error signals in fronto-parietal and sensorimotor networks influence oculomotor control by modulating activity in regions such as the frontal eye fields and intraparietal sulcus (Corbetta et al., 2008; Kowler, 2011). Thus, oculomotor behavior and neural prediction error signals are tightly coupled, providing complementary measures of anticipatory and error-driven processes during action understanding.

Furthermore, pupillometry offers a sensitive, noninvasive measure of arousal-related responses to surprise during action observation. Pupil diameter is closely linked to activity in the locus coeruleus–noradrenergic (LC–NE) system, which modulates cortical gain and responds to unexpected or salient events (Mathôt, 2018; Preuschoff et al., 2011). Within predictive processing frameworks, violations of expectation generate prediction errors that increase arousal, reflected as transient pupil dilation. During action observation, larger pupil responses have been observed when actions deviate from expected goals, timing, or kinematics, indicating heightened uncertainty and the need for model updating (Nassar et al., 2013; Preuschoff et al., 2011). These pupillary responses occur even in the absence of overt behavioral change, suggesting that they index internal belief updating rather than motor output alone. Neuroimaging studies further link pupil-linked arousal to increased activity in fronto-parietal and cingulate networks involved in error monitoring and cognitive control (Urai et al., 2017). Thus, pupillometry provides a complementary physiological marker of prediction error during action observation, capturing arousal-driven adjustments in predictive processing. In **Study 2,** we also observed the pupil dilation at the critical action boundary exactly where the violation of expectation occurred, indicating the arousal.

Saccadic eye movements, and particularly saccade velocity, are sensitive to prediction error and uncertainty during action observation, reflecting the interaction between perceptual expectations and oculomotor control. During action observation, gaze dynamics reveal predictive processes that anticipate forthcoming movements and outcomes, with saccadic behavior tightly linked to action prediction and internal forward models. Observers generate anticipatory saccades toward predicted action outcomes even before they occur, reflecting sensorimotor predictions about future states of ongoing actions (Gouret & Pfeuffer, 2023). These anticipatory eye movements vary systematically with the timing and relevance of action effects, suggesting that prediction is both time-sensitive and context-dependent. Eye movement patterns support the notion that action perception functions as active hypothesis testing: saccades are directed to the most informative visual locations to disambiguate competing action hypotheses under uncertainty (Donnarumma et al., 2017). Predictive gaze also occurs during complex goal-directed tasks, such as tracking throwing outcomes, where predictive saccades precede visual feedback and are modulated by outcome uncertainty and prediction value.

Contemporary models propose that saccade dynamics are shaped by probabilistic beliefs about future sensory states, such that higher confidence in predictions leads to faster and more vigorous saccades (Schütz et al., 2012). When observed actions violate expected kinematic or goal-related trajectories, prediction error increases uncertainty, resulting in reduced saccade velocity and altered movement profiles (Brunyé & Gardony, 2017a; Harris & Wolpert, 2006). These effects are thought to arise from changes in evidence accumulation and motor gain within fronto-parietal oculomotor circuits, including the frontal eye fields and superior colliculus (Ding & Gold, 2013; Munoz & Everling, 2004). During action observation, unexpected transitions or ambiguous action outcomes therefore modulate saccade velocity, indexing the need to update internal predictive models and resolve uncertainty (Diaz et al., 2013). However, the findings of **Study 2,** after the critical action boundary, the saccade velocity increased after the point of violation, and this could be due to internal model formed during the imitation training about the four quadrants in which the action could take place. And in the current study, the change in velocity might be the reflection of arousal that is in line with pupil dilation as a response to the prediction error.

Another important measure of eye tracking that could complement the prediction error signal in gaze behavior is Fixation. Specifically, fixation counts, and fixation duration provides a sensitive window into prediction error and uncertainty during action observation. Within predictive processing accounts, fixations reflect active information sampling aimed at reducing uncertainty about upcoming action states. When observed actions unfold as expected, gaze behavior is characterized by fewer fixations and shorter fixation durations, indicating efficient prediction and low prediction error. In contrast, unexpected action transitions or ambiguous goals increase fixation frequency and dwell time, reflecting heightened uncertainty and the need to update internal models (Henderson, 2017; Võ et al., 2019). Eye-tracking studies of action observation show that observers fixate longer on task-relevant body parts or objects when action outcomes are uncertain, suggesting that fixations are strategically allocated to resolve prediction errors (Ambrosini et al., 2011; Gredebäck & Falck-Ytter, 2015b).

During action observation, fixation patterns dynamically reflect the interplay between attentional allocation, prediction error, and uncertainty. Observers tend to increase fixation counts and prolong fixation durations when action outcomes are ambiguous or violate expectations, suggesting an active process of monitoring and updating internal models (Hooge & Erkelens, 1996; Unema et al., 2005). Eye-tracking research shows that such fixation adjustments are sensitive to probabilistic cues and prior experience, with uncertainty leading to more distributed gaze patterns and higher cognitive load (Torrents-Rodas et al., 2021). In **Study 2**, there was change in both fixation count and fixation duration. After the critical action boundary, the incorrect fixations were significantly higher for the illegal sequences and as a result, the fixation duration decreased. This is due to the increased sampling to minimize the uncertainty after the expectation being violated. The action space opened for the next action step yet to occur resulting in attentional exploration (Beesley et al., 2015).

Together, the findings across neural and oculomotor measures converge on a unified account of prediction error processing during action observation. At the neural level, violations of expected action structure engage distributed components of the action observation network, including temporal, parietal, premotor, and control regions, reflecting hierarchical detection and resolution of mismatches between predicted and observed actions. These neural prediction error signals are mirrored in oculomotor behavior, where changes in saccades, fixations, and pupil dilation provide temporally precise, behaviorally grounded markers of belief updating. The deviations from expected action sequences elicited arousal-related pupil dilation, altered saccade dynamics, and increased exploratory fixation patterns, indexing uncertainty and model revision. Importantly, the alignment between neural responses and eye-tracking measures supports the view that prediction error is not a purely perceptual signal but a system-wide mechanism linking perception, action, attention, and arousal. By jointly examining brain activity and oculomotor behavior, the present studies **(Study 1** and **2)** demonstrate how prediction errors dynamically reorganize both internal representations and outward sampling strategies to support adaptive action understanding.

4.4. Brain Oscillations of Coordination in Joint Action

Study 3 examined how hierarchical coordination demands in joint action are reflected in neural oscillatory dynamics at action boundaries. Building on the predictive processing framework outlined in Sections 4.2 and 4.3, study extends the investigation from individual action perception and prediction error to interactive, shared action. By orthogonally manipulating coordination at the level of goals, subgoals, and movements within a naturalistic joint task, the study provides a fine-grained account of how shared intentional structure shapes neural dynamics across frequency bands. Crucially, the findings demonstrate that joint action is supported by hierarchically organized neural processes, with distinct oscillatory signatures indexing different levels of coordination and prediction during action transitions (K. Friston, 2010; Kilner et al., 2007; Sebanz et al., 2006).

A key contribution of this study lies in its focus on action boundaries as critical moments for coordination and predictive updating. As discussed in Section 1.2, action boundaries are points of increased uncertainty where ongoing predictions must be evaluated, terminated, or revised, and new predictions instantiated for upcoming actions (Baldassano et al., 2017; Zacks et al., 2007). While prior work has largely examined such boundaries in passive action observation, the present results demonstrate that similar principles apply to joint action, where predictions must be aligned not only with the environment but also with a co-actor (Knoblich et al., 2011; Vesper et al., 2010). The modulation of theta, alpha, and beta oscillations at Touching and Untouching events suggests that action boundaries serve as neural coordination checkpoints at which shared predictive models are confirmed or adjusted.

Theta-band dynamics provide compelling evidence for hierarchical predictive processing in joint action. Theta power showed a robust decrease at action boundaries, particularly at Touching events, consistent with the interpretation that these moments correspond to the resolution of uncertainty and completion of a predictive cycle. Theta oscillations have been widely implicated in cognitive control, conflict monitoring, working memory, and prediction error processing (Cavanagh & Frank, 2014; Karakas, 2020; Sauseng et al., 2010). Within predictive processing accounts, reduced theta activity may reflect a transient release from control demands once prediction error is minimized and an action outcome is confirmed (Billeke et al., 2013; Huang et al., 2015).

Importantly, theta suppression was further amplified when participants shared a common goal, suggesting that high-level goal alignment enhances predictability at action boundaries. Shared goals constrain the space of possible future actions, thereby reducing uncertainty and the need for control-related theta engagement. This finding aligns with the evidence that shared intentionality facilitates anticipatory coordination by aligning internal models across agents (Tomasello et al., 2005; Vesper et al., 2017).

However, theta dynamics were not uniform across coordination levels. When subgoal coordination was required, theta suppression at Touching events was attenuated. This pattern suggests that intermediate coordination levels place specific demands on attentional updating and working memory. Subgoals structure action sequences more concretely than overarching goals but still require active maintenance and confirmation at salient checkpoints (M. M. Botvinick, 2008; Koechlin & Summerfield, 2007). In this context, Touching events may function as moments for confirming joint progress, recalling task rules, and updating shared representations, processes known to engage theta oscillations (Pomper & Ansorge, 2021; Williams et al., 2019). Thus, theta activity flexibly reflects hierarchical coordination demands rather than task difficulty per se.

Alpha-band oscillations further elucidate attention and inhibitory mechanisms supporting joint action. Contemporary accounts conceptualize alpha activity as an active mechanism for functional inhibition and information gating rather than cortical idling (Jensen & Mazaheri, 2010; Klimesch, 2012). Decreases in alpha power are typically associated with enhanced attentional engagement and information processing, whereas increases reflect inhibitory control and the suppression of task-irrelevant processes (Foxe & Snyder, 2011). The present findings largely conform to this framework.

Alpha power decreased during Subgoal coordination and at Touching events, indicating heightened attentional demands and increased monitoring at salient action boundaries. Coordinating at the subgoal level requires the integration of shared intermediate plans, which likely increases attentional engagement and reduces inhibitory alpha activity

(Womelsdorf et al., 2014). Similarly, Touching events mark the initiation of new action segments and therefore demand increased outcome evaluation and readiness for subsequent actions, processes consistently associated with alpha desynchronization (Klimesch et al., 2007; Pfurtscheller & Lopes Da Silva, 1999).

In contrast, alpha power increased during Movement-level coordination, suggesting a functional shift toward efficient motor execution. When coordination is driven primarily at the movement level, higher-order cognitive processing may become redundant or disruptive. Increased alpha activity in this context may therefore reflect the suppression of unnecessary sensory and cognitive input, allowing smooth and automatic motor coordination (Haegens et al., 2011). The interaction between Subgoal coordination and Action Boundary, showing elevated alpha power at Touching events under subgoal coordination, further suggests that shared intermediate representations reduce the attentional demands typically associated with action transitions. Within predictive processing frameworks, this pattern is consistent with increased predictive confidence and reduced need for attentional reallocation (Clark, 2013; K. Friston & Frith, 2015).

Beta-band oscillations emerged as the most sensitive marker of sensorimotor coordination and hierarchical integration. Beta activity is classically associated with motor control, sensorimotor integration, and the maintenance of the current cognitive–motor state (Engel & Fries, 2010; Kilavik et al., 2013). Rather than signaling movement execution alone, beta oscillations are thought to reflect the stabilization of ongoing motor sets and the prediction of forthcoming sensorimotor states (C. Palmer et al., 2016; C. E. Palmer et al., 2019) .

Consistent with this view, beta power showed strong sensitivity to movement coordination and to action boundaries, indicating its role in updating and maintaining coordinated motor states (Limanowski et al., 2024). The absence of main effects of Goal and Subgoal coordination supports the idea that beta activity is not driven by abstract task structure in isolation but by how such structure is instantiated through movement (Czeszumski et al., 2020). Crucially, the observed two-way and three-way interactions among Goal, Subgoal, and Movement layers indicate that beta oscillations serve as a neural interface between hierarchical task representations and physical coordination. Top-down goals and subgoals modulated beta activity only when expressed through coordinated movement, highlighting beta's role in binding abstract intentions to sensorimotor execution.

The significant interaction between Subgoal coordination and Action Boundary further suggests that beta dynamics are sensitive to how intermediate plans are updated at salient transition points. Most notably, the three-way interaction among Goal, Subgoal, and Movement layers indicates that beta power reflects the integrated coordination of multiple hierarchical levels rather than any single level alone. This integrative function distinguishes beta oscillations from lower-frequency bands, which are more closely tied to cognitive control and attentional regulation.

Taken together, the oscillatory findings of Study 3 complement and extend the neural and oculomotor results from Studies 1 and 2. While earlier studies demonstrated how prediction and prediction error are processed during action observation, the present study shows how similar predictive principles govern real-time joint action. Action boundaries remain central moments for prediction updating, but in joint contexts these updates must incorporate shared goals, intermediate plans, and coordinated motor execution.

In conclusion, Study 3 demonstrates that brain oscillations provide a powerful window into the hierarchical organization of joint action. Theta-band activity reflects flexible control and updating at higher and intermediate levels of coordination, alpha-band activity indexes attentional engagement and predictive confidence, and beta-band activity binds abstract task structure to sensorimotor execution. Situated within a predictive processing framework, these findings advance our understanding of how humans achieve fluent coordination with others. Joint action emerges as a deeply predictive, hierarchically organized process in which shared intentions are continuously evaluated, updated, and enacted at action boundaries

4.5. Critical Evaluations and Methodological Considerations

Together, the three studies form a coherent investigation of predictive processing in action understanding and joint action, progressing from controlled action observation (Studies 1 and 2) to interactive, real-time coordination (Study 3). Importantly, Studies 1 (fMRI) and 2 (eye tracking) constitute two methodological perspectives on the same paradigm, allowing complementary neural and oculomotor signatures of action prediction to be examined under identical task constraints. Study 3 then extends this framework into a naturalistic joint-action setting, addressing limitations of the earlier studies while introducing new methodological challenges.

Studies 1 and 2 contribute to the literature on action prediction by combining probabilistic action learning with fine-grained neural and oculomotor measurements. A major strength is the use of structured, multi-step action sequences, moving beyond simplified single-step paradigms. This design enables prediction to be examined as a temporally unfolding process, aligned with real-world action perception. Careful definition of action boundaries using computer-vision algorithms strengthens rigor by enabling precise time-locked analyses of prediction generation and violation.

However, the ecological validity of complex sequences introduces interpretational challenges. Actions remain highly constrained, repetitive, and explicitly trained, promoting strong motor representations and explicit knowledge of action structure. Consequently, it is difficult to disentangle whether observed predictive processes reflect general action understanding or task-specific strategies. In everyday observation, predictions often rely on partial experience, contextual cues, and social inference rather than explicit sequence learning, engaging potentially distinct neural and oculomotor mechanisms.

Reliance on probabilistic transition structures assumes uniform internalization of transition probabilities, though individual differences in learning rate, strategy, or motor proficiency may influence prediction strength. Branch number captures only objective uncertainty, not subjective confidence. Future work could model individual learning trajectories or incorporate trial-by-trial expectation measures to better link behavioral, oculomotor, and neural indices of prediction.

In Study 1, the dissociation between neural processes involved in retrieving possible actions and selecting a specific prediction represents a compelling theoretical advance. However, the spatial and temporal limitations of fMRI constrain the conclusiveness of this interpretation. Observed activations may reflect overlapping processes such as working memory, attentional control, or decision-making rather than prediction-specific computations. In particular, frontal activations are attributed to retrieving possible actions that could alternatively increase cognitive control demands at action boundaries. Without converging evidence from methods with higher temporal resolution, such as MEG, it remains difficult to definitively attribute these activations to distinct predictive stages.

Similarly, in Study 2, the interpretation of oculomotor signatures as markers of prediction error is theoretically well motivated but inherently ambiguous. Measures such as fixation duration, saccade velocity, and pupil dilation are influenced by multiple interacting factors, including arousal, attentional load, task difficulty, and motor urgency. Although the authors carefully argue for a prediction-error interpretation and control for low-level confounds, the observed gaze patterns likely reflect a combination of surprise, uncertainty, and attentional reorientation rather than a unitary prediction-error signal. Notably, the increase in saccade velocity following prediction violations contrasts with findings from other prediction-error paradigms, underscoring how task context and stimulus dynamics shape oculomotor responses.

An additional limitation of Study 2 concerns the relatively small final sample size due to substantial data exclusion. While eye tracking, particularly in scanner-adjacent or constrained environments is technically challenging, the resulting reduction in statistical power raises concerns about the robustness and generalizability of some effects, especially marginal interactions. Moreover, participants were aware that they would complete a recognition task following each sequence, which may have encouraged strategic monitoring rather than spontaneous action observation. This task demand could amplify prediction-related responses, making it unclear how strongly the findings generalize to passive or natural viewing conditions.

Despite these limitations, the combined approach of Studies 1 and 2 is a strength. Convergence between neural and oculomotor markers provides cross-modal support for hierarchical predictive processing. Identical or closely related action materials enhance comparability and theoretical coherence. Future research could extend this to less explicitly trained actions, socially meaningful interactions, and goal violations, incorporating computational modeling of prediction strength and uncertainty.

Study 3 extends this framework into interactive joint action, requiring real-time coordination with shared goals, sub-goal constraints, and movement interference. The factorial manipulation of Goal, Subgoal, and Movement allows hierarchical levels of joint action to be dissociated. Integration of behavioral measures, computer-vision action boundaries, and EEG captures cognitive and sensorimotor aspects of joint prediction, improving ecological validity and addressing limitations of explicit training and recognition demands.

The factorial manipulation of goal, sub-goal, and movement coordination constitutes a notable methodological strength, allowing hierarchical levels of joint action to be dissociated within a single task. Moreover, the integration of behavioral performance measures, computer-vision defined action boundaries, and EEG provides a rich multimodal dataset capable of capturing both cognitive and sensorimotor aspects of joint prediction. Relative to Studies 1 and 2, Study 3 thus offers improved ecological validity and directly addresses concerns that earlier findings may be driven by explicit training and recognition demands.

Nevertheless, Study 3 introduces limitations. Tasks remain rule-based, with explicit instructions, time pressure, and scoring, which may shape strategies. Complexity increases cognitive load and complicates isolation of prediction-specific processes. EEG provides high temporal but limited spatial resolution, and recording from one participant per dyad restricts conclusions about shared neural dynamics. Extensive instruction and practice may encourage strategic planning, limiting spontaneity characteristic of real-world joint action.

Across all studies, a trade-off exists between experimental control and ecological validity. Studies 1 and 2 provide fine-grained insight under controlled conditions, while Study 3 moves toward naturalistic joint action but sacrifices some interpretability and neural specificity. The progression from Studies 1 and 2 to Study 3 reflects a deliberate methodological trajectory, with each study compensating for others' limitations and offering complementary perspectives on predictive processing.

5. Conclusion

The work presents a comprehensive investigation into predictive processing in action understanding and joint action, employing a multimodal approach combining neuroimaging, eye tracking, and electrophysiology to examine hierarchical prediction across perception and interaction. Studies 1 and 2 form a tightly linked experimental pair, using probabilistic action repertoires and precisely defined action boundaries identified through computer vision algorithms to probe prediction from complementary perspectives. Study 1 (fMRI) revealed neural correlates of action prediction, dissociating retrieval of possible outcomes from selection of specific predictions, highlighting fronto-parietal engagement time-locked to action boundaries. Study 2 (eye tracking and pupillometry) captured fine-grained oculomotor responses to prediction errors. Anchored to the same action boundaries, violations of learned sequences triggered coordinated changes in fixation patterns, saccade velocity, and pupil dilation, reflecting heightened uncertainty and model updating. Together, these studies provide converging evidence that predictive mechanisms are hierarchically organized, temporally precise, and expressed across multiple neural and behavioral domains.

Building on these insights, Study 3 (EEG) investigated predictive processing in real-time joint action, where individuals coordinate movements with a partner under hierarchical constraints of goals, sub-goals, and movement coordination. Oscillatory brain activity was analyzed with respect to computer-vision-defined action boundaries, linking neural dynamics directly to levels of coordination in joint action. Results showed that predictive computations dynamically shape interpersonal coordination, influencing anticipation of partner actions and joint actions at multiple hierarchical levels. This extension addresses limitations of earlier studies, including reliance on explicitly trained sequences and passive observation, and highlights the complexity of prediction in socially embedded, real-time environments.

Across all three studies, combining fMRI, eye tracking, and EEG enabled a nuanced understanding of predictive processing from distinct but complementary perspectives, with analyses anchored to precise action boundaries. fMRI provided spatially precise insights into cortical networks supporting hierarchical prediction, eye tracking captured behavioral manifestations of prediction errors, and EEG elucidated temporal dynamics of prediction during coordination. This methodological triangulation strengthens theoretical claims, as converging evidence from neural, oculomotor, and oscillatory measures supports hierarchical predictive coding accounts of action understanding and joint action.

At the same time, the thesis highlights trade-offs between experimental control and ecological validity. Studies 1 and 2 offered tightly controlled paradigms with fine-grained analyses but relied on structured, trained sequences that may limit generalizability. Study 3 increased ecological validity through interactive coordination but introduced interpretational challenges, including overlapping cognitive and motor processes and constraints in neural localization. Nonetheless, the progression from controlled observation to real-time interaction reflects a coherent trajectory, with each study addressing limitations of the others while advancing understanding of predictive mechanisms in human action.

Overall, this thesis demonstrates that predictive processing is a core feature of both action perception and joint action, expressed across neural, oculomotor, and behavioral domains. By integrating multiple methods, hierarchical analyses, and computer-vision-defined action boundaries, it provides a detailed account of how the brain anticipates, monitors, and updates predictions in complex, structured, and interactive contexts. Future work can explore less explicitly trained, socially meaningful actions, computational modeling of prediction strength, and hyperscanning to capture genuinely shared neural dynamics during real-world interaction.

6. References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Aksoy, E. E., Abramov, A., Dörr, J., Ning, K., Dellen, B., & Wörgötter, F. (2011). Learning the semantics of object-action relations by observation. *International Journal of Robotics Research*, *30*(10), 1229–1249. https://doi.org/10.1177/0278364911410459

Alexander, W. H., & Brown, J. W. (2015). Hierarchical Error Representation: A Computational Model of Anterior Cingulate and Dorsolateral Prefrontal Cortex. *Neural Computation*, *27*, 2354–2410. https://doi.org/doi:10.1162/NECO_a_00779

Alexander, W. H., & Brown, J. W. (2018). Frontal cortex function as derived from hierarchical predictive coding. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-21407-9

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, *106*(3), 1437–1442. https://doi.org/10.1152/jn.00118.2011

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*(3), 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Bednark, J. G., Campbell, M. E. J., & Cunnington, R. (2015). Basal ganglia and cortical networks for sequential ordering and rhythm of complex movements. *Frontiers in Human Neuroscience*, *9*, 1–13. https://doi.org/10.3389/fnhum.2015.00421

Beesley, T., Nguyen, K. P., Pearson, D., & Le Pelley, M. E. (2015). Uncertainty and predictiveness determine attention to cues during human associative learning. *Quarterly Journal of Experimental Psychology*, *68*(11), 2175–2199. https://doi.org/10.1080/17470218.2015.1009919

Ben-Yakov, A., & Dudai, Y. (2011). Constructing realistic engrams: Poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *Journal of Neuroscience*, *31*(24), 9032–9042. https://doi.org/10.1523/JNEUROSCI.0702-11.2011

Billeke, P., Zamorano, F., Cosmelli, D., & Aboitiz, F. (2013). Oscillatory brain activity correlates with risk perception and predicts social decisions. *Cerebral Cortex*, *23*, 2872–2883. https://doi.org/10.1093/cercor/bhs269

Bolt, N. K., & Loehr, J. D. (2024). Motor-related cortical oscillations distinguish one's own from a partner's contributions to a joint action. *Biological Psychology*, *190*. https://doi.org/10.1016/j.biopsycho.2024.108804

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208. https://doi.org/10.1016/j.tics.2008.02.009

Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience*, *6*, 1–19. https://doi.org/10.3389/fnhum.2012.00147

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*(5712), 1118–1121. https://doi.org/10.1126/science.1105783

Brunyé, T. T., & Gardony, A. L. (2017). Eye tracking measures of uncertainty during perceptual decision making. *International Journal of Psychophysiology*, *120*, 60–68. https://doi.org/10.1016/j.ijpsycho.2017.07.008

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38. https://doi.org/10.1196/annals.1440.011

Burunat, I., Levitin, D. J., & Toiviainen, P. (2024). Breaking (musical) boundaries by investigating brain dynamics of event segmentation during real-life music-listening. *Proceedings of the National Academy of Sciences*, *121*(36), 1–11. https://doi.org/10.1073/pnas

Cavanagh, & Frank, M. J. (2014). Frontal Theta as a Mechanism for Affective and Effective Control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012.Frontal

Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. *Learning and Memory*, *18*(8), 523–528. https://doi.org/10.1101/lm.2135211

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cooper, R. A., & Ritchey, M. (2020). Progression from feature-specific brain activity to hippocampal binding during episodic encoding. *Journal of Neuroscience*, *40*(8), 1701–1709. https://doi.org/10.1523/JNEUROSCI.1971-19.2019

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Cross, E. S., Stadler, W., Parkinson, J., Schütz-Bosbach, S., & Prinz, W. (2013). The influence of visual training on predicting complex action sequences. *Human Brain Mapping*, *34*(2), 467–486. https://doi.org/10.1002/hbm.21450

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2002). The preparation and execution of self-initiated and externally-triggered movement: A study of event-related fMRI. *NeuroImage*, *15*(2), 373–385. https://doi.org/10.1006/nimg.2001.0976

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., & König, P. (2020). Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. *Frontiers in Human Neuroscience*, *14*, 1–17. https://doi.org/10.3389/fnhum.2020.00039

Diaz, G. J., Cooper, J., Rothkopf, C., & Hayhoe, M. (2013). Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of Vision*, *13*(1), 1–14. https://doi.org/10.1167/13.1.21

Ding, L., & Gold, J. I. (2013). The basal ganglia's contributions to perceptual decision making. *Neuron*, *79*(4), 640–649. https://doi.org/10.1016/j.neuron.2013.07.042

Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex*, *89*, 45–60. https://doi.org/10.1016/j.cortex.2017.01.016

Duncan, K., Ketz, N., Inati, S. J., & Davachi, L. (2012). Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus*, *22*(3), 389–398. https://doi.org/10.1002/hipo.20933

Eisenberg, M. L., Zacks, J. M., & Flores, S. (2018). Dynamic prediction during perception of everyday events. *Cognitive Research: Principles and Implications*, *3*(53), 1–12. https://doi.org/10.1186/s41235-018-0146-z

Elsner, C., D'Ausilio, A., Gredebäck, G., Falck-Ytter, T., & Fadiga, L. (2013). The motor cortex is causally related to predictive eye movements during action observation. *Neuropsychologia*, *51*(3), 488–492. https://doi.org/10.1016/j.neuropsychologia.2012.12.007

Engel, A. K., & Fries, P. (2010). Beta-band oscillations-signalling the status quo? *Current Opinion in Neurobiology*, *20*(2), 156–165. https://doi.org/10.1016/j.conb.2010.02.015

Fanghella, M., Mussini, E., Zazio, A., Genovese, F., Satta, E., Barchiesi, G., Battaglia-Mayer, A., Bortoletto, M., & Sinigaglia, C. (2025). Distinct neural dynamics in joint versus side-by-side actions: insights from dual EEG. *Social Cognitive and Affective Neuroscience*, *20*(1), 1–9. https://doi.org/10.1093/scan/nsaf075

Favila, N., Gurney, K., & Overton, P. G. (2023). Role of the basal ganglia in innate and learned behavioural sequences. *Reviews in the Neurosciences*, *35*(1), 35–55. https://doi.org/10.1515/revneuro-2023-0038

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolotti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, *308*(5722), 662–667. https://doi.org/10.1126/science.1106138

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, *2*, 1–13. https://doi.org/10.3389/fpsyg.2011.00154

Fragkiadaki, K., Levine, S., Felsen, P., & Malik, J. (2015). Recurrent network models for human dynamics. *Proceedings of the IEEE International Conference on Computer Vision*, *2015 Inter*, 4346–4354. https://doi.org/10.1109/ICCV.2015.494

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., & Frith, C. (2015). A Duet for one. *Consciousness and Cognition*, *36*, 390–405. https://doi.org/10.1016/j.concog.2014.12.003

Friston, K., & Kiebel, S. (2009). Cortical circuits for perceptual inference. *Neural Networks*, *22*(8), 1093–1104. https://doi.org/10.1016/j.neunet.2009.07.023

Frith, C. D., & Frith, U. (1999). Interacting minds - A biological basis. *Science*, *286*(5445), 1692–1695. https://doi.org/10.1126/science.286.5445.1692

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Girdhar, R., Carreira, J., Doersch, C., & Zisserman, A. (2019). Video Action Transformer Network Rohit. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. http://arxiv.org/abs/1812.02707

Go, H., Danckert, J., & Anderson, B. (2022). Saccadic eye movement metrics reflect surprise and mental model updating. *Attention, Perception, and Psychophysics*, *84*(5), 1553–1565. https://doi.org/10.3758/s13414-022-02512-4

Gouret, F., & Pfeuffer, C. U. (2023). Anticipatory Saccades Towards the Future Consequences of One’s Actions – an Online Eye Tracking Study. *Journal of Cognition*, *6*(1), 1–16. https://doi.org/10.5334/joc.261

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, *70*(1–2), 119–136. https://doi.org/10.1006/nlme.1998.3843

Gredebäck, G., & Falck-Ytter, T. (2015). Eye Movements During Action Observation. *Perspectives on Psychological Science*, *10*(5), 591–598. https://doi.org/10.1177/1745691615589103

Grèzes, J., & Decety, J. (2002). Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia*, *40*(2), 212–222. https://doi.org/10.1016/S0028-3932(01)00089-6

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(48), 19377–19382. https://doi.org/10.1073/pnas.1117190108

Hamilton, A. F. D. C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, *26*(4), 1133–1137. https://doi.org/10.1523/JNEUROSCI.4551-05.2006

Hamilton, A. F. de C., & Grafton, S. T. (2012). The motor hierarchy: from kinematics to goals and intentions. *Sensorimotor Foundations of Higher Cognition*, (July), 381–407. https://doi.org/10.1093/acprof:oso/9780199231447.003.0018

Harris, C. M., & Wolpert, D. M. (2006). The main sequence of saccades optimizes speed-accuracy trade-off. *Biological Cybernetics*, *95*(1), 21–29. https://doi.org/10.1007/s00422-006-0064-x

Hayhoe, M. M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research*, *217*(1), 125–136. https://doi.org/10.1007/s00221-011-2979-2

Henderson, J. M. (2017). Gaze Control as Prediction. *Trends in Cognitive Sciences*, *21*(1), 15–23. https://doi.org/10.1016/j.tics.2016.11.003

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Hooge, I. T. C., & Erkelens, C. J. (1996). Control of fixation duration in a simple search task. *Perception and Psychophysics*, *58*(7), 969–976. https://doi.org/10.3758/BF03206825

Huang, Y., Chen, L., & Luo, H. (2015). Behavioral oscillation in priming: Competing perceptual predictions conveyed in alternating theta-band rhythms. *Journal of Neuroscience*, *35*(6), 2830–2837. https://doi.org/10.1523/JNEUROSCI.4294-14.2015

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*(2), 318–329. https://doi.org/10.1093/cercor/bhq098

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*, 1–8. https://doi.org/10.3389/fnhum.2010.00186

Kahl, S., & Kopp, S. (2018). A predictive processing model of perception and action for self-other distinction. *Frontiers in Psychology*, *9, 2421*. https://doi.org/10.3389/fpsyg.2018.02421

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *NeuroImage*, *132*, 417–424. https://doi.org/10.1016/j.neuroimage.2016.02.062

Karakas, S. (2020). A review of theta oscillation and its functional correlates. *International Journal of Psychophysiology*, *157*, 82–99. https://doi.org/https://doi.org/10.1016/j.ijpsycho.2020.04.008

Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A., & Riehle, A. (2013). The ups and downs of beta oscillations in sensorimotor cortex. *Experimental Neurology*, *245*, 15–26. https://doi.org/10.1016/j.expneurol.2012.09.014

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88. https://doi.org/10.1016/j.brainresrev.2006.06.003

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological Research on Joint Action. Theory and Data. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 54). https://doi.org/10.1016/B978-0-12-385527-5.00003-6

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*(6), 229–235. https://doi.org/10.1016/j.tics.2007.04.005

Koster-Hale, J., & Saxe, R. (2013). Theory of Mind: A Neural Prediction Problem. *Neuron*, *79*(5), 836–848. https://doi.org/10.1016/j.neuron.2013.08.020

Kourtis, D., Sebanz, N., Knoblich, G., Kourtis, D., Sebanz, N., & Knoblich, G. (2013). Predictive representation of other people' s actions in joint action planning: An EEG study. *Social Neuroscience*, *8*(1), 31–42. https://doi.org/10.1080/17470919.2012.694823

Kourtis, D., Woźniak, M., Sebanz, N., & Knoblich, G. (2019). Evidence for we-representations during joint action planning. *Neuropsychologia*, *131*, 73–83. https://doi.org/10.1016/j.neuropsychologia.2019.05.029

Kowler, E. (2011). Eye movements: The past 25years. *Vision Research*, *51*(13), 1457–1483. https://doi.org/10.1016/j.visres.2010.12.014

Kumaran, D., & Maguire, E. A. (2007). Match-mismatch processes underlie human hippocampal responses to associative novelty. *Journal of Neuroscience*, *27*(32), 8517–8524. https://doi.org/10.1523/JNEUROSCI.1677-07.2007

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Li, S. C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the Couplings Among Intellectual Abilities and Constituent Cognitive Processes Across the Life Span. *Psychological Science*, *15*(3), 155–163. https://doi.org/10.1111/j.0956-7976.2004.01503003.x

Limanowski, J., Adams, R. A., Kilner, J., & Parr, T. (2024). The Many Roles of Precision in Action. *Entropy*, *26*(9), 1–21. https://doi.org/10.3390/e26090790

Lohse, B. (2006). Attention to Intention. *Journal of Nutrition Education and Behavior*, *38*(3), 207. https://doi.org/10.1016/j.jneb.2006.04.003

Markiewicz, R., Segaert, K., & Mazaheri, A. (2024). Brain-to-brain coupling forecasts future joint action outcomes. *IScience*, *27*(9), 110802. https://doi.org/10.1016/j.isci.2024.110802

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/joc.18

Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*(3), 218–228. https://doi.org/10.1038/nrn1345

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, *9*(11), 856–869. https://doi.org/10.1038/nrn2478

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2013). Rational regulation of learning dynamics by pupil–linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. https://doi.org/10.1038/nn.3130.Rational

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28–38. https://doi.org/10.1037/h0035584

Novembre, G., Sammler, D., & Keller, P. E. (2016). Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action. *Neuropsychologia*, *89*, 414–425. https://doi.org/10.1016/j.neuropsychologia.2016.07.027

Palmer, C. E., Auksztulewicz, R., Ondobaka, S., & Kilner, J. M. (2019). Sensorimotor beta power reflects the precision-weighting afforded to sensory prediction errors. *NeuroImage*, *200*, 59–71. https://doi.org/10.1016/j.neuroimage.2019.06.034

Palmer, C., Zapparoli, L., & Kilner, J. M. (2016). A New Framework to Explain Sensorimotor Beta Oscillations. *Trends in Cognitive Sciences*, *20*(5), 321–323. https://doi.org/10.1016/j.tics.2016.03.007

Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y., & Platt, M. L. (2012). Posterior Cingulate Cortex: Adapting Behavior to a Changing World. *Trends in Cognitive Sciences*, *15*(4), 143–151. https://doi.org/10.1016/j.tics.2011.02.002.Posterior

Pesquita, A., Whitwell, R. L., & Enns, J. T. (2018). Predictive joint-action model: A hierarchical predictive approach to human cooperation. *Psychonomic Bulletin and Review*, *25*(5), 1751–1769. https://doi.org/10.3758/s13423-017-1393-6

Pfurtscheller, G., & Lopes Da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*, *110*(11), 1842–1857. https://doi.org/10.1016/S1388-2457(99)00141-8

Pomp, J., Heins, N., Trempler, I., Kulvicius, T., Tamosiunaite, M., Mecklenbrauck, F., Wurm, M. F., Wörgötter, F., & Schubotz, R. I. (2021). Touching events predict human action segmentation in brain and behavior. *NeuroImage*, *243*, 118534. https://doi.org/10.1016/j.neuroimage.2021.118534

Pomp, J., Wurm, M. F., Selvan, R. N., Wörgötter, F., & Schubotz, R. I. (2025). Touching-untouching patterns organize action representation in the inferior parietal cortex. *NeuroImage*, *310*. https://doi.org/10.1016/j.neuroimage.2025.121113

Pomper, U., & Ansorge, U. (2021). Theta-Rhythmic Oscillation of Working Memory Performance. *Psychological Science*, *32*(11), 1801–1810. https://doi.org/10.1177/09567976211013045

Press, C., Cook, J., Blakemore, S. J., & Kilner, J. (2011). Dynamic modulation of human motor activity when observing actions. *Journal of Neuroscience*, *31*(8), 2792–2800. https://doi.org/10.1523/JNEUROSCI.1595-10.2011

Preuschoff, K., ’t Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline’s role in decision making. *Frontiers in Neuroscience*, *5*, 1–12. https://doi.org/10.3389/fnins.2011.00115

Rens, G., van Polanen, V., Botta, A., Gann, M. A., de Xivry, J. J. O., & Davare, M. (2020). Sensorimotor expectations bias motor resonance during observation of object lifting: The causal role of pSTS. *Journal of Neuroscience*, *40*(20), 3995–4009. https://doi.org/10.1523/JNEUROSCI.2672-19.2020

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Roseberry, S., Richie, R., Hirsh-Pasek, K., Golinkoff, R. M., & Shipley, T. F. (2011). Babies catch a break: 7- to 9-month-olds track statistical probabilities in continuous dynamic events. *Psychological Science*, *22*(11), 1422–1424. https://doi.org/10.1177/0956797611422074

Rotman, G., Troje, N. F., Johansson, R. S., & Flanagan, J. R. (2006). Eye movements when observing predictable and unpredictable actions. *Journal of Neurophysiology*, *96*(3), 1358–1369. https://doi.org/10.1152/jn.00227.2006

Ruotolo, F., Ruggiero, G., Raemaekers, M., Iachini, T., van der Ham, I. J. M., Fracasso, A., & Postma, A. (2019). Neural correlates of egocentric and allocentric frames of reference combined with metric and non-metric spatial relations. *Neuroscience*, *409*, 235–252. https://doi.org/10.1016/j.neuroscience.2019.04.021

Sasaki, A. T., Okamoto, Y., Kochiyama, T., Kitada, R., & Sadato, N. (2018). Distinct sensitivities of the lateral prefrontal cortex and extrastriate body area to contingency between executed and observed actions. *Cortex*, *108*, 234–251. https://doi.org/10.1016/j.cortex.2018.08.003

Sauseng, P., Griesmayr, B., Freunberger, R., & Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1015–1022. https://doi.org/10.1016/j.neubiorev.2009.12.006

Schacter, D. L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic future thinking: mechanisms and functions. *Current Opinion in Behavioral Sciences*, *17*, 41–50. https://doi.org/10.1016/j.cobeha.2017.06.002

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*(1), 3–8. https://doi.org/10.1002/hipo.22523.Statistical

Schwartze, M., Rothermich, K., & Kotz, S. A. (2012). Functional dissociation of pre-SMA and SMA-proper in temporal processing. *NeuroImage*, *60*(1), 290–298. https://doi.org/10.1016/j.neuroimage.2011.11.089

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Selvan, R. N., Cheng, M., Siestrup, S., Mecklenbrauck, F., Jainta, B., Pomp, J., Zahedi, A., Tamosiunaite, M., Wörgötter, F., & Schubotz, R. I. (2024). Updating predictions in a complex repertoire of actions and its neural representation. *NeuroImage*, *296, 120687*. https://doi.org/10.1016/j.neuroimage.2024.120687

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *NeuronNeuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007.The

Shima, K., & Tanji, J. (2000). Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *Journal of Neurophysiology*, *84*(4), 2148–2160. https://doi.org/10.1152/jn.2000.84.4.2148

Simonyan, K., & Zisserman, A. (2014). Two-Stream Convolutional Networks for Action Recognition in Videos. *Advances in Neural Information Processing Systems*, 1–9. https://doi.org/10.1016/0006-2952(83)90587-7

Sinclair, A. H., Manalili, G. M., Brunec, I. K., Alison Adcock, R., & Barense, M. D. (2021). Prediction errors disrupt hippocampal representations and update episodic memories. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(51), 1–12. https://doi.org/10.1073/pnas.2117625118

Sprevak, M., & Smith, R. (2023). An Introduction to Predictive Processing Models of Perception and Decision-Making. *Topics in Cognitive Science*, *00*, 1–28. https://doi.org/10.1111/tops.12704

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643–1653. https://doi.org/10.1038/nn.4650

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development*, *85*(5), 1821–1826. https://doi.org/10.1111/cdev.12247

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*(7), 3522–3535. https://doi.org/10.1093/cercor/bhab029

Sulpizio, V., Committeri, G., Lambrey, S., Berthoz, A., & Galati, G. (2013). Selective role of lingual/parahippocampal gyrus and retrosplenial complex in spatial memory across viewpoint changes relative to the environmental reference frame. *Behavioural Brain Research*, *242*(1), 62–75. https://doi.org/10.1016/j.bbr.2012.12.031

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event Boundaries in Perception Affect Memory Encoding and Updating. *Journal of Experimental Psychology: General*, *138*(2), 236–237. https://doi.org/10.1037/a0015631

Tanji, J. (2001). Sequential Organization of Multiple Movements: Involvement of Cortical. *Annual Review of Neuroscience*, *24*, 631–651.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Torrents-Rodas, D., Koenig, S., Uengoer, M., & Lachnit, H. (2021). A rise in prediction error increases attention to irrelevant cues. *Biological Psychology*, *159*, 108007. https://doi.org/10.1016/j.biopsycho.2020.108007

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494. https://doi.org/10.1080/13506280444000409

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8,* 14637. https://doi.org/10.1038/ncomms14637

Urgen, B. A., Pehlivan, S., & Saygin, A. P. (2019). Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. *Neuropsychologia*, *127*, 35–47. https://doi.org/10.1016/j.neuropsychologia.2019.02.006

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. https://doi.org/10.1016/j.cortex.2020.03.014

Vannuscorps, G., & Caramazza, A. (2017). Typical predictive eye movements during action observation without effector-specific motor simulation. *Psychonomic Bulletin and Review*, *24*(4), 1152–1157. https://doi.org/10.3758/s13423-016-1219-y

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., Hristova, D., Karlinsky, A., McEllin, L., Nijssen, S. R. R., Schmitz, L., & Wahn, B. (2017). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology*, 1–7. https://doi.org/10.3389/fpsyg.2016.02039

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*, 998–1003. https://doi.org/10.1016/j.neunet.2010.06.002

Viglione, A., Mazziotti, R., & Pizzorusso, T. (2023). From pupil to the brain: New insights for studying cortical plasticity through pupillometry. *Frontiers in Neural Circuits*, *17*, 1–11. https://doi.org/10.3389/fncir.2023.1151847

Võ, M. L. H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, *29*, 205–210. https://doi.org/10.1016/j.copsyc.2019.03.009

Wang, C., Li, H., Jia, L., Li, F., & Wang, J. (2020). Theta band behavioral fluctuations synchronized interpersonally during cooperation. *Psychonomic Bulletin and Review*, *27*(3), 563–570. https://doi.org/10.3758/s13423-020-01711-0

Williams, C. C., Kappen, M., Hassall, C. D., Wright, B., & Krigolson, O. E. (2019). Thinking theta and alpha: Mechanisms of intuitive and analytical reasoning. *NeuroImage*, *189*, 574–580. https://doi.org/10.1016/j.neuroimage.2019.01.048

Wills, A. J., Lavric, A., Croft, G. S., & Hodgson, T. L. (2007). Predictive learning, prediction errors, and attention: Evidence from event-related potentials and eye tracking. *Journal of Cognitive Neuroscience*, *19*(5), 843–854. https://doi.org/10.1162/jocn.2007.19.5.843

Wörgötter, F., Aksoy, E. E., Krüger, N., Piater, J., Ude, A., & Tamosiunaite, M. (2013). A simple ontology of manipulation actions based on hand-object relations. *IEEE Transactions on Autonomous Mental Development*, *5*(2), 117–134. https://doi.org/10.1109/TAMD.2012.2232291

Wurm, M. F., von Cramon, D. Y., & Schubotz, R. I. (2011). Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. *Human Brain Mapping*, *32*(12), 2141–2150. https://doi.org/10.1002/hbm.21176

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked. *Nature Neuroscience*, *4*, 651–655. https://doi.org/10.1038/88486

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*(12), 4057–4066. https://doi.org/10.1162/jocn_a_00078

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event Perception: A Mind/Brain Perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273.Event

Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*, 1–15. https://doi.org/10.3389/fnhum.2010.00168

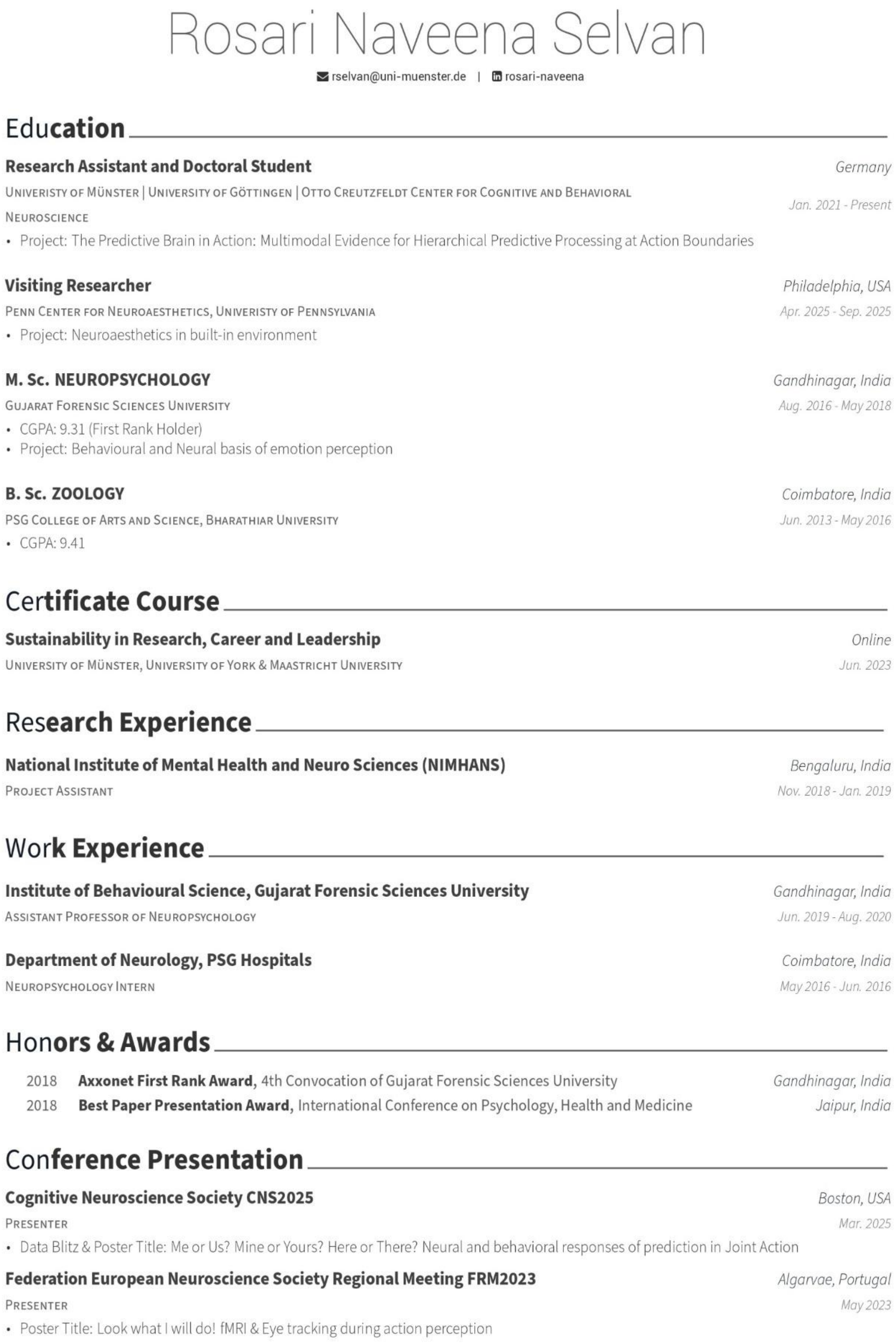
Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Ziaeetabar, F., Pomp, J., Pfeiffer, S., El-Sourani, N., Schubotz, R. I., Tamosiunaite, M., & Wörgötter, F. (2021). Using enriched semantic event chains to model human action prediction based on (minimal) spatial information. *PLoS ONE*, *15*, 1–25. https://doi.org/10.1371/journal.pone.0243829

7. Abbreviations

ACC	Anterior Cingulate Cortex
aIPS	anterior Inferior Parietal Sulcus
AON	Action Observation Network
CNV	Contingent Negative Variation
DMN	Default Mode Network
EEG	Electroencephalogram
ERP	Event Related Potential
EST	Event Segmentation Theory
fMRI	functional Magnetic Resonance Imaging
MNS	Mirror Neuron System
MPFC	Medial Prefrontal Cortex
PCC	Posterior Cingulate Cortex
PE	Prediction Error
PJAM	Predictive Joint Action Model
SEC	Semantic Event Chain
SMA	Supplementary Motor Area
T	Touching
U	Untouching

Curriculum Vitae



Rosari Naveena Selvan

✉ rselvan@uni-muenster.de | in rosari-naveena

Education

Research Assistant and Doctoral Student *Germany*

Univeristy of Münster | University of Göttingen | Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience *Jan. 2021 - Present*

- Project: The Predictive Brain in Action: Multimodal Evidence for Hierarchical Predictive Processing at Action Boundaries

Visiting Researcher *Philadelphia, USA*

Penn Center for Neuroaesthetics, Univeristy of Pennsylvania *Apr. 2025 - Sep. 2025*

- Project: Neuroaesthetics in built-in environment

M. Sc. NEUROPSYCHOLOGY *Gandhinagar, India*

Gujarat Forensic Sciences University *Aug. 2016 - May 2018*

- CGPA: 9.31 (First Rank Holder)
- Project: Behavioural and Neural basis of emotion perception

B. Sc. ZOOLOGY *Coimbatore, India*

PSG College of Arts and Science, Bharathiar University *Jun. 2013 - May 2016*

- CGPA: 9.41

Certificate Course

Sustainability in Research, Career and Leadership *Online*

University of Münster, University of York & Maastricht University *Jun. 2023*

Research Experience

National Institute of Mental Health and Neuro Sciences (NIMHANS) *Bengaluru, India*

Project Assistant *Nov. 2018 - Jan. 2019*

Work Experience

Institute of Behavioural Science, Gujarat Forensic Sciences University *Gandhinagar, India*

Assistant Professor of Neuropsychology *Jun. 2019 - Aug. 2020*

Department of Neurology, PSG Hospitals *Coimbatore, India*

Neuropsychology Intern *May 2016 - Jun. 2016*

Honors & Awards

2018 **Axxonet First Rank Award**, 4th Convocation of Gujarat Forensic Sciences University *Gandhinagar, India*
2018 **Best Paper Presentation Award**, International Conference on Psychology, Health and Medicine *Jaipur, India*

Conference Presentation

Cognitive Neuroscience Society CNS2025 *Boston, USA*

Presenter *Mar. 2025*

- Data Blitz & Poster Title: Me or Us? Mine or Yours? Here or There? Neural and behavioral responses of prediction in Joint Action

Federation European Neuroscience Society Regional Meeting FRM2023 *Algarvae, Portugal*

Presenter *May 2023*

- Poster Title: Look what I will do! fMRI & Eye tracking during action perception

International Conference on Psychology, Health and Medicine *Jaipur, India*
Presenter *Feb. 2018*

- Paper Title: Temporal Continuity in Emotional Facial Expression Recognition Task: A behavioural Analysis
- Paper Title: Electrophysiological correlates of cognitive workload during Music perception: Revisiting Mozart Effect

Technical Skills

Techniques	EEG, fMRI, Eye tracking
Programming	Python, MATLAB, R
Experimental Psychology Software	Presentation, Qualtrics, E-Prime, SuperLab

Workshop and Training

Lab Rotation on Neurobiology of Drosophila *Sep. 2024*
Institute of Neuro- and Behavioral Biology, University of Münster *Münster, Germany*

IMPRS NeuroCom Summer School in Cognitive Neuroscience *Jun. 2023*
Max Planck Institute for Human Cognitive and Brain Sciences *Leipzig, Germany*

Cognitive Neuroscience Skill Training COGNESTIC 2022 *Sep. 2022*
MRC Cognition and Brain Sciences Unit, University of Cambridge *Cambridge, England*

European Summer School on Eye Movements ESSEM 2022 *Sep. 2022*
Department of Psychology, University of Bonn *Bonn, Germany*

SPM Course for fMRI/VBM *May 2022*
The Wellcome Centre for Human Neuroimaging, University College London *Online*

Lab Rotation on Eye Tracking and Virtual Reality *Sep. 2021*
Institute of Psychology, University of Münster *Münster, Germany*

Summer School on NeuroImaging *Jul. 2017*
International Institute of Information Technology *Hyderabad, India*

Basics of Polysomnography and Sleep Physiology *Mar. 2017*
National Institute of Mental Health and NeuroSciences (NIMHANS) *Bangalore, India*

Skills

Soft Skills	Work Ethics, Communication, Adaptability, Problem Solving, Time Management, Team Work
Language Skills	English, German (B1), Tamil (Mother Tongue)

Publications

- Pomp, J., Wurm, M. F., **Selvan, R. N.**, Wörgötter, F., & Schubotz, R. I. (2025). Touching-untouching patterns organize action representation in the inferior parietal cortex. *Neuroimage*, 310, 121113. https://doi.org/10.1016/j.neuroimage.2025.121113
- Papatzikis, E., Agapaki, M., **Selvan, R. N.**, Hanson-Abromeit, D., Gold, C., Epstein, S., Lok, U. W. V., Barda, E., & Pandey, V. (2024). Music Therapy and Music Medicine in Neonatal Care: A scoping review of passive music listening research applications and findings on infant development and medical practice. *BMC Paediatrics*, 24, 829. https://doi.org/10.1186/s12887-024-05275-z.
- **Selvan, R. N.**, Cheng, M., Siestrup, S., Mecklenbrauck, F., Jainta, B., Pomp, J., Zahedi, A., Tamosiunaite, M., Wörgötter, F., & Schubotz, R. I. (2024). Updating predictions in a complex repertoire of actions and its neural representation. *Neuroimage*, 296, 120687. https://doi.org/10.1016/j.neuroimage.2024.120687
- Papatzikis, E., Agapaki, M., **Selvan, R. N.**, Pandey, V., & Zeba, F. (2023). Quality standards and recommendations for research in music and neuroplasticity. *Annals of the New York Academy of Sciences*, 1520(1), 20–33. https://doi.org/10.1111/nyas.14944
- Papatzikis, E., Elhalik, M., Inocencio, S. A. M., Agapaki, M., **Selvan, R. N.**, Muhammed, F. S., Haroon, N. A., Dash, S. K., Sofologi, M., & Bezoni, A. (2021). Key Challenges and Future Directions When Running Auditory Brainstem Response (ABR) Research Protocols with Newborns: A Music and Language EEG Feasibility Study. *Brain sciences*, 11(12), 1562. https://doi.org/10.3390/brainsci11121562

Declarations

Declarations by the candidate documenting **open-science activities** and **on the consideration of ethical aspects** as part of the doctoral process and disclosure of **personal contribution** for manuscripts completed by two or more authors (cumulative dissertations)

Doctoral candidate : Rosari Naveena Selvan

Title of dissertation : The Predictive Brain in Action: Multimodal Evidence for Hierarchical Predictive Processing at Action Boundaries

1. Documentation of **open-science activities**

Manuscript 1

	yes	no	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials	X		
Open access publication	X		https://doi.org/10.1016/j.neuroimage.2024.120687

Manuscript 2

	yes	no	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials	X		
Open access publication	X		

Manuscript 3

	yes	no	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials	X		
Open access publication	X		

2. Declaration on **the consideration of ethical aspects**

Study number	**Source (manuscript / chapter of dissertation):** e.g. study 1 in paper 2, study 1 described in chapter 4	**Was the study reviewed by an ethics commission?**	
		yes	**no**
1	Manuscript 1	X	
1	Manuscript 2	X	
2	Manuscript 3	X	

3. Declaration of one's **personal contribution** to the submitted academic manuscripts by two or more authors

Manuscript 1

Title	Updating predictions in a complex repertoire of actions and its neural representation		
Author(s)	Rosari Naveena Selvan, Minghao Cheng, Falko Mecklenbrauck, Benjamin Jainta, Sophie Siestrup, Jennifer Pomp, Anoushiravan Zahedi, Minija Tamosiunaite, Florentin Wörgötter, Ricarda I. Schubotz		
Publication status:	not yet submitted		(please mark with X)
	submitted		
	in review		
	in revision		
	accepted		
	published	X	
Journal	Neuroimage		
Year of publication	2024		

Description of your own contribution in the case of joint authorship:
- partly responsible for the study's conception and design
- responsible for data collection
- mainly processing, analyzing and interpreting data
- mainly responsible for writing and revising the manuscript
- marked as corresponding author

Manuscript 2

Title	Seeing the Unexpected: Oculomotor Signatures of Prediction Error at Action Boundaries		
Author(s)	Rosari Naveena Selvan, Anoushiravan Zahedi, Minghao Cheng, Minija Tamosiunaite, Florentin Wörgötter, Ricarda I. Schubotz		
Publication status:	not yet submitted		(please mark with X)
	submitted		
	in review	X	
	in revision		
	accepted		
	published		
Journal	Journal of Vision		
Year of publication			

Description of your own contribution in the case of joint authorship:

- partly responsible for the study’s conception and design
- responsible for data collection
- mainly processing, analyzing and interpreting data
- mainly responsible for writing and revising the manuscript
- marked as corresponding author

Manuscript 3

Title	Me or Us? Mine or Yours? Here or There? Neural and behavioral responses of prediction in Joint Action		
Author(s)	Rosari Naveena Selvan, Anoushiravan Zahedi, Minghao Cheng, Katharina Dohr, Minija Tamosiunaite, Florentin Wörgötter, Ricarda I. Schubotz		
Publication status:	not yet submitted	X	(please mark with X)
	submitted		
	in review		
	in revision		
	accepted		
	published		
Journal			
Year of publication			
Description of your own contribution in the case of joint authorship: - partly responsible for the study's conception and design - responsible for data collection - mainly processing, analyzing and interpreting data - mainly responsible for writing and revising the manuscript - marked as corresponding author			

Münster, 06.02.2026

Place, Date

Signature of the doctoral candidate