THE INTERPLAY OF PRIOR INFORMATION AND MOTION CUES IN RESOLVING VISUAL AMBIGUITY IN AGENT PERCEPTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
NEUROSCIENCE

By
Sena Er Elmas
August 2023

The Interplay of Prior Information and Motion Cues in Resolving Visual Ambiguity in Agent Perception
By Sena Er Elmas
August 2023

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Burcu Ayşen Ürgen(Advisor)

Murat Perit Çakır

Yasemin Afacan

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

THE INTERPLAY OF PRIOR INFORMATION AND MOTION CUES IN RESOLVING VISUAL AMBIGUITY IN AGENT PERCEPTION

Sena Er Elmas
M.S. in Neuroscience
Advisor: Burcu Ayşen Ürgen
August 2023

Agent perception, a complex cognitive task that involves interpreting an agent's actions to infer their internal states and adjust our behavior accordingly, is a fundamental aspect of human cognition. Despite its importance, a significant gap persists in our understanding of how low-level factors, such as motion and form information, interact with top-down factors, such as prior information about the agents.

To address this gap, we conducted two electroencephalogram (EEG) experiments investigating the interplay of prior knowledge and motion information on the temporal dynamics of agent perception in the human brain. We used human, android and robot agents engaged in various actions. The chosen agents were designed to form a set wherein form information alone could not easily resolve the ambiguity in agent identities.

In the first experiment, participants were informed about agent identities before the experiment (Prior Experiment), while in the second experiment, participants remained uninformed (Naive Experiment). We controlled the availability of motion information by presenting stimuli in either video and image formats. We recorded scalp EEG and utilized event-related potential (ERP) analysis and model-based representational similarity analysis (RSA) to uncover the temporal course of agent representation in the human brain.

Our results revealed that the processing of agents in EEG depends on the availability of motion information and prior information. Specifically, in the Naive Experiment, agent information was available longer during the still condition than

in the moving condition. In contrast, agent information was present for similar durations in still and moving conditions of the Prior Experiment.

These findings suggest that prior knowledge and motion information interactively modulate the duration of the processing of agent information. Our results underscore the critical role of prior knowledge and motion cues in shaping the processing of agent information, highlighting the complex interplay between top-down modulation and bottom-up cues in driving the perception of agents and their actions. This study contributes to our understanding of the temporal dynamics of agent perception and the role of top-down and bottom-up processes in this complex cognitive task.

Keywords: Agent perception, prior knowledge, visual ambiguity, action observation, ERP, RSA.

ÖZET

AKTÖR ALGILAMA SÜRECİNDE GÖRSEL BELİRSİZLİĞİ ÇÖZMEDE ÖN BİLGİ VE HAREKET İPUÇLARININ ETKİLEŞİMİ

Sena Er Elmas
Nörobilimler, Yuksek Lisans
Tez Danışmanı: Burcu Ayşen Ürgen
Ağustos 2023

Aktör algısı, bir aktörün içsel durumlarını anlamak için eylemlerini yorumlamayı ve davranışlarımızı buna göre ayarlamayı içeren karmaşık bir bilişsel görev, insan bilişinin temel bir yönüdür. Önemine rağmen, hareket ve form bilgisi gibi aşağıdan-yukarıya faktörlerin, ön bilgi gibi üstten-aşağıya faktörlerle nasıl etkileşime girdiğine dair anlayışımızda önemli bir boşluk var.

Bu boşluğu gidermek için, insan beynindeki aktör algısının zamansal dinamikleri üzerindeki ön bilgi ve hareket bilgisinin etkileşimini araştıran iki elektroensefalogram (EEG) deneyi gerçekleştirdik. Çeşitli eylemlerde bulunan insan, android ve robot aktörleri kullandık. Seçilen aktörler, form bilgisinin tek başına aktör kimliklerindeki belirsizliği kolayca çözemeyeceği bir küme oluşturması için tasarlandı.

İlk deneyde, katılımcılara deneyden önce aktör kimlikleri hakkında bilgi verildi (Ön Bilgi Deneyi), ikinci deneyde ise katılımcılar bilgilendirilmedi (Naif Deney). Hareket bilgisini kontrol etmek için uyaranları video ve resim formatlarında sunduk. Kafa derisi EEG'sini kaydettik ve olaya ilişkin potansiyel (ERP) analizi ve model tabanlı temsili benzerlik analizi (RSA) kullanarak insan beynindeki aktör temsilinin zaman içindeki seyrini ortaya çıkardık.

Sonuçlarımız, aktörlerin görsel işlenmesiyle ilişkili EEG sinyallerinin hareket bilgisi ve ön bilginin varlığına bağlı olduğunu ortaya koydu. Özellikle, Naif Deneyde, aktör bilgisi hareketsiz koşulda hareketli koşulunda olduğundan daha uzun süre boyunca devam etti. Buna karşılık, aktör bilgisi Ön Bilgi Deneyinde hareketsiz ve hareketli koşullarda benzer süreler için mevcuttu.

Bu bulgular, ön bilginin ve hareket bilgisinin aktör bilgisi işleme süresini etkileşimli bir şekilde modüle ettiğini öne sürmektedir. Sonuçlarımız, ön bilginin ve hareket ipuçlarının aktör bilgisinin işlenmesini şekillendirmedeki kritik rolünü vurgulamakta, aktörlerin ve eylemlerinin algılanmasını yönlendiren üstten aşağıya modülasyon ve alttan yukarıya ipuçları arasındaki karmaşık etkileşimi vurgulamaktadır. Bu çalışma, aktör algısının zamansal dinamikleri ve bu karmaşık bilişsel görevde üstten aşağıya ve alttan yukarıya süreçlerin rolü hakkındaki anlayışımıza katkıda bulunmaktadır.

Anahtar sözcükler: Aktör algılama, ön bilgi, görsel belirsizlik, eylem algısı, OİP, Temsili Benzerlik Analizi .

Acknowledgement

Embarking on this academic journey in the realm of neuroscience, I've been blessed to find myself surrounded by an incredible support system. I'd like to take a moment to extend my deepest gratitude to those who have been instrumental in shaping my journey.

Firstly, my heartfelt appreciation goes out to Dr. Burcu Ayşen Ürgen. Her journey, transitioning from a Bachelor's in Computer Engineering to a distinguished career in neuroscience, served as a beacon for me, showcasing the boundless possibilities of interdisciplinary research. As my thesis advisor, her expertise and guidance have been invaluable. While I was exploring my path in neuroscience, her example emboldened my decisions and brought clarity to my pursuits. Dr. Ürgen's unwavering support, insightful advice, and consistent encouragement have been pivotal in shaping me into the researcher I am today.

To my husband, Hüseyin, my partner in all aspects of life: we embarked on this journey hand-in-hand, facing each challenge and celebrating every success. You've been my confidant, my favorite research discussion partner, and my anchor. The hurdles seemed less daunting and the obstacles easier to overcome with you by my side. I am truly fortunate to have you in my life.

To my family, your support has been the wind beneath my wings. Mom, Pervin, and Dad, Mükremin, your boundless support and open-mindedness have always given me the courage to chase my dreams. My late grandfather, Mustafa, and my grandmother, Hacer, have been pillars of strength, instilling in me a love for hard work and learning. Their spirit and encouragement continue to resonate with me, and I miss them every day.

A heartfelt thank you to my friends who have walked this path with me, sharing in my challenges and joys. Your patience, understanding, and encouragement during my countless rants and moments of self-doubt were much needed and greatly appreciated.

Lastly, I extend my gratitude to TÜBİTAK for their financial support through the BIDEB scholarship program. Their contribution has been instrumental in my academic pursuits.

Contents

List of Figures

Chapter 1

Introduction

Consider a scenario in which an individual is operating a vehicle during a rainstorm. The visual field is filled with numerous moving stimuli: branches oscillating in the wind, raindrops descending on the windshield, traffic signs passing by. Despite this visual complexity, a pedestrian crossing the road immediately captures the driver's attention, illustrating the salience of agent motion in our cognitive processing.

In another scenario, an observer is situated in a crowded public square. Amidst the multitude of moving individuals, the observer identifies a familiar figure based on a distinctive gait - a close acquaintance. This scenario underscores the critical role of motion in agent identification.

Finally, envision a situation involving a stray dog in a park. If the dog exhibits aggressive movements, an observer might perceive it as a threat and maintain distance. Conversely, if the dog displays playful behavior, the observer might perceive it as friendly and approach it. These reactions highlight how an agent's movements can inform our perception of their internal state, such as their intentions or emotions, and influence our subsequent behavior.

These examples demonstrate the profound significance of agent perception in our daily interactions. Agent perception involves more than passive observation

or simple categorization of stimuli. It requires interpreting an agent's actions to infer their internal states, evaluating potential threats or rewards, and adjusting our behavior accordingly. This process carries immense evolutionary and social importance. For instance, correctly interpreting an agent's actions can help individuals avoid danger or seize opportunities. Misinterpretations, on the other hand, can lead to social misunderstandings or even life-threatening situations.

In conclusion, the perception of agents is a fundamental aspect of human cognition, with far-reaching implications for our interactions and decision-making processes. Whether it involves recognizing a familiar face in a crowd, interpreting the intentions of a stray dog, or navigating through a busy traffic scenario, our ability to perceive and interpret the actions of other agents profoundly shapes our responses and behaviors. This intricate process, which is deeply rooted in our evolutionary history, offers a rich field of study.

1.1 Perception of agents in Action Observation framework

Action observation, a part of higher-order vision, refers to the visual process of observing the actions of agents. It involves processing movements of effectors and assessing the goal of the action [2]. Perception of agents has traditionally been studied within the context of action observation, as agent (or "actor") being the main extrinsic factor in action observation. While an action cannot occur in the absence of an agent, the agent is still considered an extrinsic factor due to the fact that the identity of the action remains unchanged regardless of the actor involved. The concept of an agent encompasses various layers of information, including biological attributes such as age and gender, motor characteristics like posture, as well as more high-level cognitive properties such as intention or determination [3]. The neural subtrate underpinning action observation is a network of brain regions collectively known as the Action Observation Network.

Action observation network (AON) is a large-scale bilateral network that has three main nodes: premotor, parietal, and occipito-temporal nodes [1, 4]. Premotor node covers frontal areas BA 44/45 (Broca's area), lateral dorsal premotor cortex (dPMC, BA 6), supplementary motor area (SMA, BA 6); parietal node covers rostral inferior parietal lobule (IPL, area PFt), primary somatosensory cortex (SI, BA 1/2), superior parietal (SPL, area 7A), intraparietal cortex (IPS, area hIP3), occipito-temporal node covers posterior middle temporal gyrus (pMTG) at the transition to visual area V5, and fusiform face area/fusiform body area (FFA/FBA) [1] (Refer to the Figure 1.1 below).

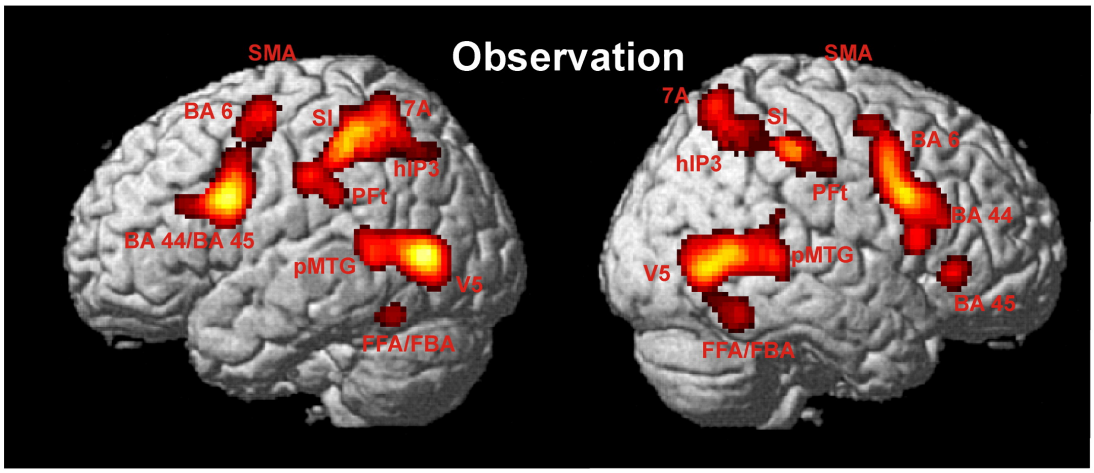

Figure 1.1: Action Observation Network. Adapted from [1]

While the specific brain regions responsible for representing the agent of an action are not definitively established, preliminary evidence suggests a role for the occipitotemporal node of the Action Observation Network (AON). Specifically, the posterior middle temporal gyrus/posterior superior temporal gyrus (pMTG/pSTG) region may be more involved in representing the observed agent or actor, rather than the observed actions [3]. Neuroimaging studies in monkeys also support the role of occipitotemporal nodes in agent processing; research in macaque monkeys has shown that the middle and rostral parts of the superior temporal sulcus (STS), corresponding to the pMTG/pSTG in the macaque brain, are activated when observing monkey agents [5].

In a study not directly focused on agent representation, Tarhan and Konkle proposed that the distinctions between sub-networks in action observation could be explained by the sociality and interaction scale. They identified a right-lateralized "agent-focused" sub-network, which is primarily engaged in actions directed at other people and the actor themselves. This sub-network includes regions traditionally specialized for body parts, whole bodies, and faces. Its responses align with previous evidence on animacy and object-size based organization in the occipitotemporal cortex (OTC) [6]. Further supporting the role of OTC in processing social aspects of an observed action, Wurm et al. (2017) reported that the dorsal lateral OTC is involved in processing the sociality features of an action, independent from action sub-categories [7]. Sociality-related organization of OTC is relevant to agent representation in the brain, as the processing of social aspects of an observed action shares commonalities with agent processing. In essence, understanding the social context of an action, which includes recognizing the agent, is a crucial aspect of interpreting and predicting behavior. Therefore, the research on sociality also implicate at the possible involvement of OTC in representing relevant agent features.

In conclusion, while the understanding of brain regions involved in representing the agent of an observed action is still limited, the current evidence supports the involvement of OTC.

1.2 Bottom-Up Processing of Agents

1.2.1 Perception of Agents Involves Both Dorsal and Ventral Visual Pathways

The perception of agents is a complex process that involves the integration of both form and motion information. This process is facilitated by two distinct visual pathways in the brain, as described by Ungerleider and Mishkin based on lesion studies in primates [8]. The ventral pathway, often referred to as the

"what" pathway, is primarily involved in object recognition and form representation. In contrast, the dorsal pathway, also known as the "where/how" pathway, is responsible for processing spatial location and motion. Given that form and motion cues each provide rich and unique information about agents, the perception of agents necessitates the involvement of both these visual pathways.

Perception of agents involves processing both the form (body form) and motion (body motion) of the agent. For instance, static body posture serves as a body form cue, while a gait pattern can serve as a body motion cue, both informing the observer about the agents and their actions. Two main theoretical models have been proposed to explain the neural mechanisms for processing body form and body motion cues [9]. The first theory posits that movements of agents are recognized solely based on the concatenation of static snapshots of body poses [10].

In contrast, the second theory, proposed by Giese and Poggio, suggests that the recognition of biological movements is achieved through parallel feedforward processing of body form and body motion cues in the ventral and dorsal visual streams, respectively [11]. This model comprises two parallel processing streams for form and motion information, each containing hierarchical layers of feature detectors. As one ascends the hierarchy, feature detectors at higher levels detect more complex features and have larger receptive fields [11].

Recent empirical evidence supports the view of separate parallel processing for form and motion cues. Vangeneugdan et al. reported that the extrastriate body area (EBA) responds to body form cues but not to body motion cues, while the posterior superior temporal sulcus (pSTS) responds to body motion cues but not to body form cues. Moreover, the discrimination of body form and body motion can be selectively impaired by manipulating different visual cues. Finally, applying transcranial magnetic stimulation (TMS) over the EBA selectively disrupted body form discrimination, while TMS over the pSTS selectively disrupted body motion discrimination. Taken together, these results suggest that body form and body motion cues are processed by functionally and neurally distinct mechanisms [9]. However, while this parallel processing model provides a robust framework

for understanding agent perception, it is not without its limitations.

1.2.2 Limitations of Feedforward Parallel Processing Model

The model proposed by Giese and Poggio offers a comprehensive framework for understanding the neural mechanisms underlying the recognition of biological motion. However, it primarily posits a feedforward processing mechanism, which may not fully account for the substantial evidence for top-down effects of attention and expectation on the recognition of movements of agents [12, 13, 14, 15, 16]. These top-down effects become especially impactful in cases where the stimuli are ambiguous and noisy, a common occurrence in natural scenarios [17]. Therefore, while the model's emphasis on the importance of both form and motion cues in agent perception is valuable, a comprehensive understanding of agent perception also requires considering the role of top-down processes.

Moreover, one key assumption of the Giese & Poggio's is that body motion cues are processed in the dorsal pathway. However, evidence from human lesion case studies challenges this assumption. Remarkably, 'motion-blind' patients, who have lesions in the dorsal pathway, still retain the ability to recognize biological motion [18, 19].

This counterintuitive finding suggests that the processing of an agent's body motion cues is not solely confined to the dorsal pathway. Instead, it implies the existence of separate neural mechanisms that can compensate for the impaired intermediate stages of the dorsal stream. These compensatory mechanisms could involve interconnections with the ventral stream, which might allow for the inference of motion from static frames [20, 21, 22]. Alternatively, they could involve top-down connections from hierarchically higher regions in the dorsal stream [23, 24], or regions outside the traditional ventral and dorsal streams might also play a role in processing body motion cues [25, 26].

In conclusion, while the feedforward parallel processing model provides a valuable starting point, a more comprehensive understanding of the neural mechanisms underlying the perception of agents in motion will require a broader perspective. This includes considering the potential role of other neural pathways and the interplay between top-down and bottom-up processes, especially in resolving visual ambiguities. The complexity and dynamism inherent in agent perception often result in ambiguous visual cues, necessitating the integration of top-down predictions with bottom-up sensory data to generate a coherent perception.

1.2.3 Perception of Agents beyond the Ventral and Dorsal Visual Pathways

Building on this broader perspective, recent research suggests that our understanding of visual processing, based on the two-visual pathway model, may require revision in light of new empirical findings. Specifically, a third visual pathway, specialized for social perception, has been proposed [26]. This proposed pathway, located on the lateral surface of the brain, begins in the primary visual cortex (V1) and projects into the posterior banks of the superior temporal sulcus (STS) via the V5/MT+ cluster. Unlike the traditional ventral and dorsal streams, this third pathway is proposed to process moving faces and moving bodies, rather than static faces or bodies. Face-selective areas in the posterior and anterior STS have been shown to exhibit a greater response to dynamic faces, whereas face-selective areas in the ventral stream, such as the fusiform face area (FFA) and occipital face area (OFA), do not exhibit any preference for dynamic faces [26]. This suggests that the STS, which has been identified as a hub for social perception and cognition, may play a crucial role in the processing of dynamic social cues [27, 28, 29].

The identification of a third pathway, specifically dedicated to social perception, underscores the importance of dynamic information in the perception of agents. This pathway's focus on moving faces and bodies aligns with the fundamental role of agent perception in understanding and predicting social behavior.

As such, the third pathway offers a crucial piece of the puzzle in our understanding of how the brain processes and interprets agent-related information.

Despite the fact that lateral and feedback connectivity is a well-documented feature of the brain, most computational models of visual perception, including models of biological motion perception, predominantly rely on feedforward computations [11, 30, 31, 32]. While feedforward processes continue to form the foundation of our understanding of vision, recent research has increasingly focused on the role of recurrent processing in resolving ambiguity [33, 34, 35, 36, 37]. Concurrently, there is a growing body of research examining the influence of top-down factors, such as prior expectations and attention, in shaping our perception of agents (attention: [12, 13, 14], prior expectation: [15, 38, 39, 40, 41, 42]).

1.3 Top-Down Processing of Agents

The role of top-down processing in visual perception, particularly in resolving ambiguity, has been the focus of numerous studies. It is understood that top-down and bottom-up processing work in tandem. This interplay is crucial in situations where sensory input is ambiguous or incomplete, as it is often the case in nature. The combination of bottom-up sensory input and top-down initial guesses or predictions, informed by prior knowledge or expectations, facilitates perception by helping to resolve such ambiguities.

Bar (2006) found that activity in the orbitofrontal cortex (OFC), modulated by the stimulus, preceded activity in object-recognition related areas in the temporal cortex. This suggests a role for the OFC in the top-down facilitation of object recognition [23]. While this study focused on object recognition, its findings may have implications for the perception of agents as well. Agents, being more complex and dynamic than inanimate objects, likely require even more top-down processing for accurate perception. However, the specific role of the OFC and other brain regions in the top-down processing of agent perception remains to be fully elucidated.

In terms of recurrent processing, Kar et al. (2019) found that recurrent convolutional neural networks (CNNs) predicted late IT responses in primates more accurately than feedforward deep CNNs [35]. This suggests that recurrent processing, which involves feedback from higher to lower levels of the visual system, plays a crucial role in the perception of agents. Moreover, they found that the performance of shallow-but-recurrent CNNs was similar to that of feedforward-very-deep CNNs, suggesting that recurrence has a function equivalent to additional nonlinear transformations. This finding implies that recurrent processing could be a potential neural mechanism for efficiently processing ambiguous and dynamic stimuli, such as those encountered in the perception of agents.

Kietzmann et al. (2019) used magnetoencephalography (MEG), time-resolved representational similarity analysis (RSA), Granger causality, and deep neural network modeling to suggest that recurrence is essential in the human visual system [43]. The study examined three levels of regions along the ventral-visual stream, including early-level (V1-V3), intermediate-level (V4t/lateral occipital cortex [LO]), and high-level (inferior temporal cortex [IT]/parahippocampal cortex [PHC]) visual areas. The emergence of low-level image features was observed early in the early visual areas and slightly later in the high-level visual areas. The influence of these low-level image features on high-level areas diminished early, coinciding with the emergence of effects related to categorical components such as face, animate, and real-world size. Interestingly, the face category emerged in similar temporal windows at all levels. The animacy category, on the other hand, emerged in a sequence opposite to that of low-level features: it first appeared in the high-level areas, then in lower levels, and finally reemerged strongly in the intermediate level and weakly in high-level areas. Additional Granger Causality analysis revealed an early feedforward influence from lower areas to higher areas and a later feedback influence from higher areas to lower areas. The categorical components that are used in the study (face, animate, and real-world size) are also pertinent to the perception of agents.

Their results suggests that the low-level visual features of agents may first emerge in low-level visual areas and then in higher-level visual areas. Meanwhile, categorical features of agents, including agent identity, might first emerge

in higher-level areas and then in lower-level areas. This sequence of emergence implies that the categorical features of agents could modulate bottom-up processing in a top-down manner, further highlighting the interplay between top-down and bottom-up processes in agent perception. Finally, recurrent DNN models closely matched RDM time-traces and outperformed feedforward models in capturing representational dynamics in the brain [43]. This suggests that computational models of agent perception need to incorporate both bottom-up and top-down processes to accurately capture the complexity of the task.

The neural mechanisms through which top-down prior expectations integrate with bottom-up sensory input have been the subject of extensive research, yet they remain a topic of ongoing debate. Predictive processing theories, such as predictive coding, offer a valuable theoretical framework for understanding the effect of prior knowledge on perception [44, 45, 46]. According to this framework, our brain continually generates and updates predictions about the sensory input it will receive, based on prior knowledge. When sensory input deviates from these predictions, prediction errors are generated, which then update our internal models of the world. Consistent with predictive coding, there is evidence that expectations can enhance early sensory representations. For instance, Kok et al. (2012) suggest that expectations improve stimulus representations in V1, demonstrating the influence of top-down priors on early visual processing [47]. However, some argue that expectations do not modulate early sensory responses, but instead interfere with later processes, such as response selection and execution. For instance, Rungratsameetaweemana et al. did not find expectation-related modulations in early ERPs related to early sensory processing and sensory evidence accumulation; however, they found that violation of expectations impacted posterior alpha and frontal theta oscillations [48].

Biological motion perception, a key aspect of agent perception, is also shaped by prior information and expectations, especially when the stimulus is vague and distorted, as is often the case in natural environments. For instance, Hunt et al. (2008) found that when the simple dots of a point-light display (PLD) walker are replaced with everyday objects, naive observers could not identify the walker in the stimulus, despite the presence of form and motion information. However,

informed observers could reliably detect the walker, suggesting an effect of prior information about the motion and the insufficiency of bottom-up processing alone in biological motion processing [15].

In conclusion, the role of prior knowledge in agent perception underscores the importance of top-down processing in our understanding of visual perception. This is particularly crucial in situations where visual cues are ambiguous and open to multiple interpretations. Prior knowledge helps resolve such ambiguities, guiding our perception towards the most likely interpretation based on our past experiences. This highlights the need for models of visual perception to incorporate both bottom-up and top-down processes, to fully capture the complexity of how we perceive and understand agents in our environment, especially in the face of visual ambiguity.

1.4 Ambiguous Stimuli in Visual Perception Literature

The study of visual perception often necessitates the use of ambiguous stimuli to effectively investigate the interplay between bottom-up and top-down processing [49]. Ambiguous stimuli, such as luminance or contrast modulations [50, 51], adding Gaussian noise [51], Mooney images [52], Necker cubes [37, 53], and various optical illusions [53, 54], have been widely used in the literature to create situations where visual cues are open to multiple interpretations. These stimuli force the perceptual system to rely more heavily on top-down processes to resolve the ambiguity.

However, creating ambiguity with dynamic stimuli presents a unique challenge. Some approaches include adding dynamic noise [55, 2] or presenting the stimulus for a brief duration [56, 57, 2], but these methods often involve a trade-off between controlling sensory evidence and maintaining ecological validity. In other words, while these methods effectively introduce ambiguity, they often do so at

the expense of the stimuli's naturalness or realism.

An alternative approach to introducing ambiguity in the study of agent perception is the use of non-biological agents, such as robots [58, 59]. Robots can be designed to resemble humans in form, creating a unique type of ambiguity. For instance, a robot and an android may be the same machine, but the android is modeled after a human. The mechanical motion cues suggest that the android is a mechanical agent, but the biological form cues imply a biological agent. This creates a unique opportunity to study the role of motion cues and prior information in resolving visual ambiguity.

Importantly, non-biological agents are not merely perceived as objects; they are perceived as agents, activating the Action Observation Network (AON) in a similar way to biological agents [58, 60]. This makes them particularly suitable for use in studies of action perception. In the following section, we delve into the perception of non-biological agents and discuss how they can be used to further our understanding of the complex interplay between bottom-up and top-down processes in visual perception.

1.5 Perception of Non-Biological Agents

Envisioning the near future, the use of robots for assistance is becoming increasingly commonplace as we design more capable robots. Consequently, the interdisciplinary field of human-robot interaction has garnered significant attention. Currently, a growing body of research is delving into how human and robot actions distinctly activate the AON [61, 62]. However, the results remain inconsistent across functional neuroimaging studies. Evidence varies, indicating that robot actions could trigger weaker, equivalent, or even stronger activation of the AON compared to human actions.

For instance, A PET study [63] and an fMRI study [64] found greater AON activation to human than robot actions. Conversely, Gazzola et al. reported

that while simple repetitive robotic actions did not activate AON, goal-oriented actions performed by both human and robot agents activated parietal, premotor, and inferior frontal areas of AON equivalently [60]. This finding was echoed in a TMS study that reported equivalent corticospinal excitability [65] and an EEG study that reported equivalent mu suppression during action observation of both human and robot agents[66].

Interestingly, Cross et al. reported more robust AON activation in response to a mechanically dancing robot as compared to a naturally dancing human [67]. This effect was observed even when the form was biological (human agent) but the kinematics were non-biological (rigid movements). Similarly, another fMRI study found stronger responses in the parietal and premotor nodes of AON to a robot grasping than to a human grasping [68].[68].

An fMRI repetition suppression study examined the neural response to three different agents: a human, a humanoid android, and a robot [58]. The findings revealed similar levels of AON activation for both the human and robot agents, except in the extrastriate body area (EBA). This suggests that the AON is not selective for biological appearance and movement. However, the android agent uniquely induced greater suppression, notably in the bilateral anterior intraparietal sulcus (aIPS). The authors attributed this stronger suppression to an increased prediction error, which arose from the incongruity between the android agent's form and motion. This interpretation aligns with the findings of [67, 68], who also observed stronger AON activation for robot actions and proposed greater prediction errors for robot actions as a potential explanation.

Overall, contradicting results can be reconciled by considering the observer's familiarity with the stimulus. It is hypothesized that AON response has a "U-shaped" profile: both extremely familiar/predictable and extremely unfamiliar/unpredictable actions induce strong responses in the AON, with weaker responses in the middle of the continuum [67]. This hypothesis is in line with "predictive coding framework" [44, 45, 46] has been proposed as the mechanism underlying the phenomenon of the "uncanny valley" [58]. The "uncanny valley" refers to the discomfort or eeriness experienced when humanoid objects closely

resemble humans but are not quite convincingly realistic [69, 70]. This suggests that the perception of non-biological agents, such as robots, may be influenced by complex interactions between bottom-up sensory input and top-down expectations.

1.6 Current Study

Despite recent advancements, our understanding of agent perception still contains significant gaps, particularly concerning the interplay between low-level factors, such as motion and form information, and top-down factors, such as prior information about the agents. The intricate interaction between these factors and their influence on the temporal dynamics of agent perception is an area that remains largely uncharted.

Our study aims to fill this gap by exploring the interplay between prior knowledge and motion information on the temporal dynamics of agent perception in the human brain. In our experiments, we presented participants with three agents engaged in various actions: **Human**, **Android**, and **Robot** (Figure 1.2). **Android** and **Robot** agents are the same machinery, but the **Android**'s appearance is modelled after the **Human** agent. In essence, **Human** agent exhibits both a biological appearance and biological motion; **Robot** agent displays a mechanical appearance and mechanical motion; and **Android** agent combines a biological appearance with mechanical motion, sharing appearance characteristics with **Human** and motion characteristics with **Robot**.

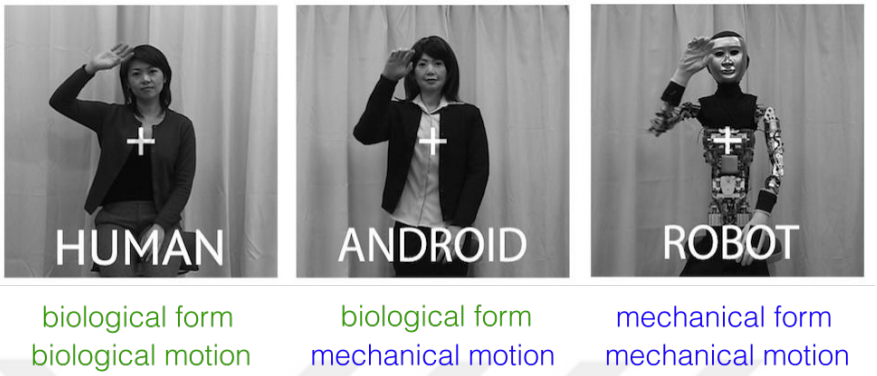

Figure 1.2: Agents in the study: **Human**, **Android** and **Robot**

Human has biological appearance and biological motion. **Robot** has mechanical appearance and mechanical motion. **Android** has biological appearance and mechanical motion; **Android** shares appearance with **Human** but shares motion characteristics with **Robot**. With this agent set, **Android** agent's identity remains ambiguous when only static images of the agents are presented.

We controlled the availability of motion information by presenting stimuli in video and image formats. Moreover, to regulate prior information, we conducted two experiments, with participants being informed about agent identities before the experiment in the first one (**Prior Experiment**) and participants remaining uninformed in the second (**Naive Experiment**). This control allowed us to test the involvement of top-down processing when form and motion cues ambiguously predict agent identities. We investigated the influence of prior knowledge and motion information on temporal dynamics of agent perception, focusing on temporal differences in the EEG signals while controlling for available cues. Additionally, we applied a combination of multiple regression and a model-based temporal variant of representational similarity analysis (RSA) [71], similar to the methodology used in [43] to pinpoint the time points of agent representation processing.

Our central hypothesis posits that the availability of motion information and prior knowledge interactively impact the temporal characteristics of agent processing. We further anticipate a more substantial influence of prior knowledge when no motion information is available, arising from the ambiguity introduced

by the absence of motion information.

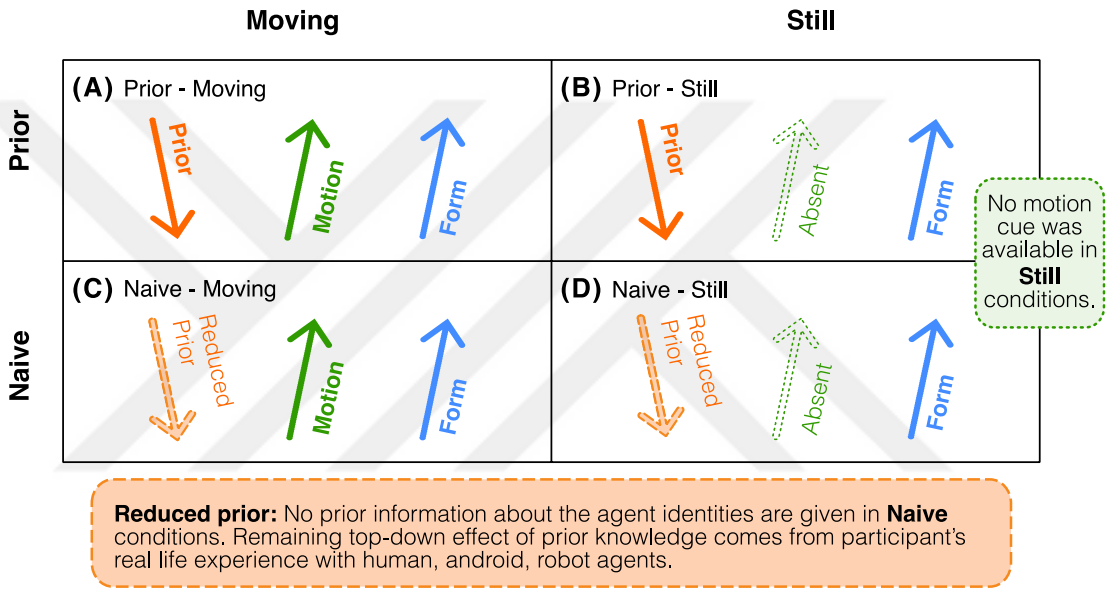

Figure 1.3: Comparison of Top-Down and Bottom-Up Effects in Experimental Conditions

The descending arrows represent top-down prior effect and the ascending arrows represent bottom-up form and motion cues. In **(A) Prior-Moving** and **(B) Prior-Still**, participants have been informed about agents identities, thus, the top-down effect of the prior is present. In **(C) Naive-Moving** and **(D) Naive-Still**, participants have not been informed about agents identities, thus, the top-down effect of the prior is reduced, remaining top-down effect of prior comes from participants' real life experience with the agents. In **(A) Prior-Moving** and **(C) Naive-Moving**, videos of an agent performing an action are presented to the participants, thus, both form and motion cues are available in the presented stimuli. In **(B) Prior-Still** and **(D) Naive-Still**, static images of an agent performing an action are presented to the participants, thus, form cues are available but motion cues are absent in the presented stimuli.

Chapter 2

Methods & Materials

2.1 Participants

A total of 35 subjects from the student community of UC San Diego participated in the study (20 participants in **Prior Experiment**; 15 participants in **Naive Experiment**). Participants had no history of neurological disorders and had a normal or corrected vision. They were given course credit upon their participation. The Human Subjects Ethics Committee of UC San Diego approved the study. The data from Experiment 1 was taken from Urgen, Kutas, and Saygin, 2018 [59].

2.2 Stimuli

The study utilized a combination of short videos (2 seconds in length, referred to as **Moving** stimuli) and images (**Still** stimuli) featuring three types of agents: **Human**, **Android**, and **Robot**, each performing a variety of actions (refer to Figure 2.1). **Android** and **Robot** were essentially the same machine, but presented in different forms. **Android** was designed to mimic the human agent in

appearance, yet it shared the same movement characteristics as **Robot**. Consequently, the form or motion alone could not definitively identify the agents in our set. To control the availability of motion information, we presented the stimuli in either short video or image formats. This approach, using visually ambiguous stimuli and controlling motion information, enabled us to examine the progression of agent processing when the visual features of the agents were either sufficient or limited. In the short videos, all agents performed eight distinct actions: drinking from a cup, examining an object with a hand, waving, turning the body, wiping a table, nudging, introducing themselves, and throwing a piece of paper. The images were derived from the first frames of each corresponding video. Further details about the stimuli can be found in [59].

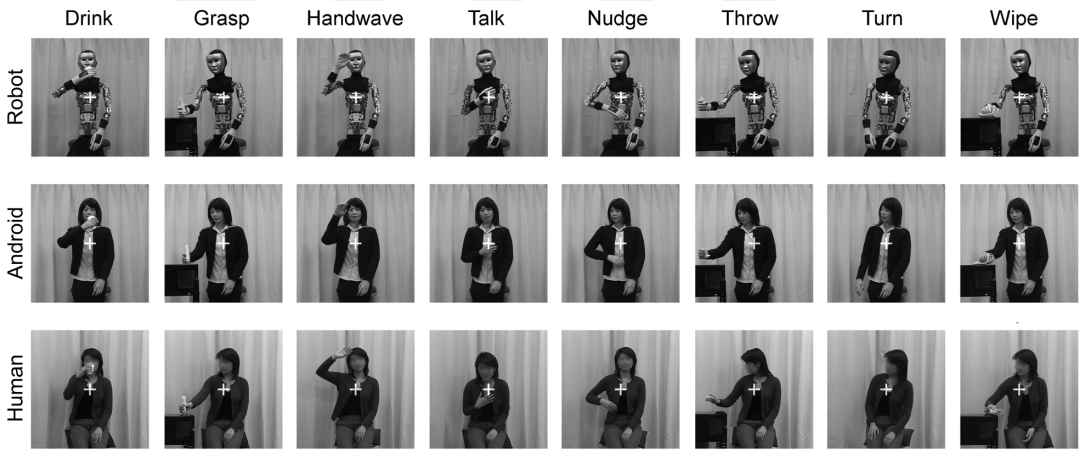

Figure 2.1: Stimuli used in the experiments.
There were three agents (robot, android, and human) and eight actions (drink, grasp, handwave, talk, nudge, throw, turn, wipe), a total of 24 combinations

2.3 Design and Procedure

Our study comprised two EEG experiments, with the primary distinction between them being the level of prior knowledge imparted to the participants. In the first experiment, referred to as the **Prior Experiment**, participants were introduced to the stimuli and briefed about the agents they would encounter

during the experiment. Conversely, in the second experiment, referred to as the **Naive Experiment**, participants were not given any specific information about the stimuli. They were simply informed that they would be viewing some images and videos and would be asked to make judgments about them. Despite the differences in the level of prior knowledge provided, both experiments manipulated two other factors: the identity of the agent (**Human**, **Android**, or **Robot**) and the availability of motion in the stimuli (either **Moving** or **Still**).

2.3.1 Prior Experiment

In the **Prior Experiment**, participants were shown videos in a pseudo-randomized sequence to ensure that the same video was not repeated in two consecutive trials. Each video was either displayed on its own or preceded by a static image (the first frame of the video) for a duration of 600–1000 ms, with a uniform probability jitter. Every video (alone) or image trial was introduced by a baseline period where a fixation cross was displayed at the center of the screen for 900–1200 ms, also with a uniform probability jitter. To maintain the participants' focus on the stimuli throughout the experiment, a comprehension question (e.g., "Was the action you just saw drinking?") was posed every 6-10 trials. Participants responded to these questions with a simple key press indicating "yes" or "no".

2.3.2 Naive Experiment

In the **Naive Experiment**, participants were shown images and videos in two separate blocks, with the image block always preceding the video block. This order was based on findings from pilot experiments that suggested participants could identify the three agents (Human, Android, Robot) once they had viewed the videos, but not always when they had only viewed the images. Therefore, we aimed to prevent the knowledge gained from the videos, primarily the motion information, from influencing the processing of the agents in the still conditions.

Within each block, images or videos were displayed in a pseudo-randomized sequence to ensure that the same image or video was not repeated in two consecutive trials. Each stimulus (image for 0.6 s or video for 2 s) was introduced by a baseline period where a fixation cross was displayed at the center of the screen for 900–1200 ms, with a uniform probability jitter. Similar to the **Prior Experiment**, a comprehension question (e.g., "Was the action you just saw drinking?") was posed every 6-10 trials to maintain the participants' focus on the stimuli throughout the experiment. Participants responded to these questions with a simple key press indicating "yes" or "no".

2.3.3 EEG Data Collection and Preprocessing

EEG was recorded with 64 ActiveTwo Ag/AgCl electrodes (Brain Vision, Inc.) at 512 Hz following the International 10/20 system. The electrode impedance level was kept below 25 k-Ohm. Two additional electrodes were placed above and below the right eye to monitor oculomotor activity. The data were preprocessed with the EEGLAB toolbox [72]. Each participant's data were high-pass filtered at 1 Hz and low-pass filtered at 50 Hz, and re-referenced to average mastoid electrodes behind the right and left ear. In **Naive Experiment**, after visual inspection, we observed a noisy electrode (Fp2) in two subjects. We interpolated this electrode only for these subjects using spherical interpolation. After the preprocessing, the data were epoched ranging from 200 ms preceding video or first frame onset to 600 ms after stimulus onset, and were time-locked to the onset of the video or the static picture to compare the moving and still forms of the agents (we refer to these as **Moving** and **Still** conditions). Atypical epochs of eye movement and electromyographic activity were removed from further analysis by semi-automated epoch rejection procedures (kurtosis and probability-based procedures with standard deviation = 6). After preprocessing, grand average event-related brain potentials (ERP) were computed and plotted for each condition using ERP Lab [73] and custom scripts using R [74] and tidyverse package [75].

2.4 Event-Related Potential Analysis

First, we calculated and examined grand average ERPs to investigate the effects of prior knowledge knowledge, agent type, and motion information availability. We adopted a data-driven approach and did not limit our investigation to a specific time period or a set of electrodes. We performed a 3-way mixed ANOVA on each electrode over the scalp and each time point with two within-subjects factors: **motion availability** (levels: **Still**, **Moving**), and **Agent** (levels: **Robot**, **Android**, and **Human**), and one between-subjects factor: **Prior Knowledge** (levels: **Naive** and **Prior**).

2.5 Time-Resolved Representational Similarity Analysis

We employed Representational Similarity Analysis (RSA) [71] in a time-resolved manner to explore the time points at which agent information becomes discernible in the EEG data, and to determine whether the time course of this information is influenced by prior knowledge and/or motion information. To this end, we created 24x24 representational dissimilarity matrices (RDM) (comprising combinations of 3 agents, eight actions) for each time point in the EEG data (refer to the EEG RDMs subsection below) and a categorical Agent model (refer to the Model RDMs subsection below) with which we correlated the EEG RDMs using Kendall-tau correlation. We also constructed two visual models (*Pixel-wise Intensity Model* and *Pixel-wise Motion Model*) and a categorical *Action model* to control and account for the low-level and other observable effects (refer to the Model RDMs subsection below).

2.5.1 EEG RDMs

We created EEG RDMs by calculating the pairwise correlation distance between EEG data (all channels combined) corresponding to each stimulus. Thus, we created a 24×24 matrix for each time point and subject. Since EEG was recorded at 512Hz, the temporal difference between each time point corresponded to 2ms. Considering that all channels might not be equally informative, we also analyzed subsets of electrodes based on their distribution over the scalp. We divided electrodes to **frontal** (Fz, F1, F2, F3, F4, AF3, AF4, AF7, AF8, AFz, Fp1, Fp2, FCz, FC1, FC2, FC5, FC6), **central** (Cz, C1, C2, C3, C4, C5, C6), **temporal** (T7, T8, FT9, FT10, FT7, FT8), **parietal** (Pz, P1, P2, P3, P4, P7, P8, POz, PO3, PO4, PO7, PO8) and **occipital** (Oz, O1, O2) subsets. We used EEG data exclusively from these electrodes to create region-specific EEG RDMs applying the same method. We used custom Python scripts and functions to create the EEG RDMs.

2.5.2 Model RDMs

To understand when agent information becomes available in the EEG data and the factors that may affect it, we utilized an attribute-based categorical model, namely, the **Agent model** (see below). Since a significant correlation between *Agent RDM* and *EEG RDMs* might also be explained by the observable visual differences between the stimuli (e.g., **Robot** stimuli might have higher pixel intensities than **Human** stimuli), we created three additional models: an attribute-based categorical *Action model*, a *Pixel-wise Motion model*, and a *Pixel-wise Intensity model*. We then performed multiple regression analysis to determine the time points where the **Agent model** independently explains the neural data significantly (Refer to the subsection Multiple Linear Regression Analysis below).

We generated Attribute-based categorical model RDMs using **Agent** (with attributes: *Robot*, *Human*, and *Android*) or **Action** (with attributes: *Drink*, *Grasp*,

Handwave, *Talk*, *Nudge*, *Throw*, *Turn*, *Wipe*) as categories to create one-hot encoded vector representations of the stimuli. A stimulus with a particular attribute was labeled with 1 in the model's respective attribute entry and labeled with 0 otherwise. We created 24×24 RDM for each categorical model by calculating the pairwise Hamming distance between each binary representation.

We generated the low-level model RDMs by converting each stimulus to a feature vector using the particular model's output (see *Pixel-wise Intensity RDM* and *Pixel-wise Motion RDM* below) for that stimulus and calculating the pairwise correlation distance between these feature vectors, resulting in a 24×24 RDM for each low-level visual model.

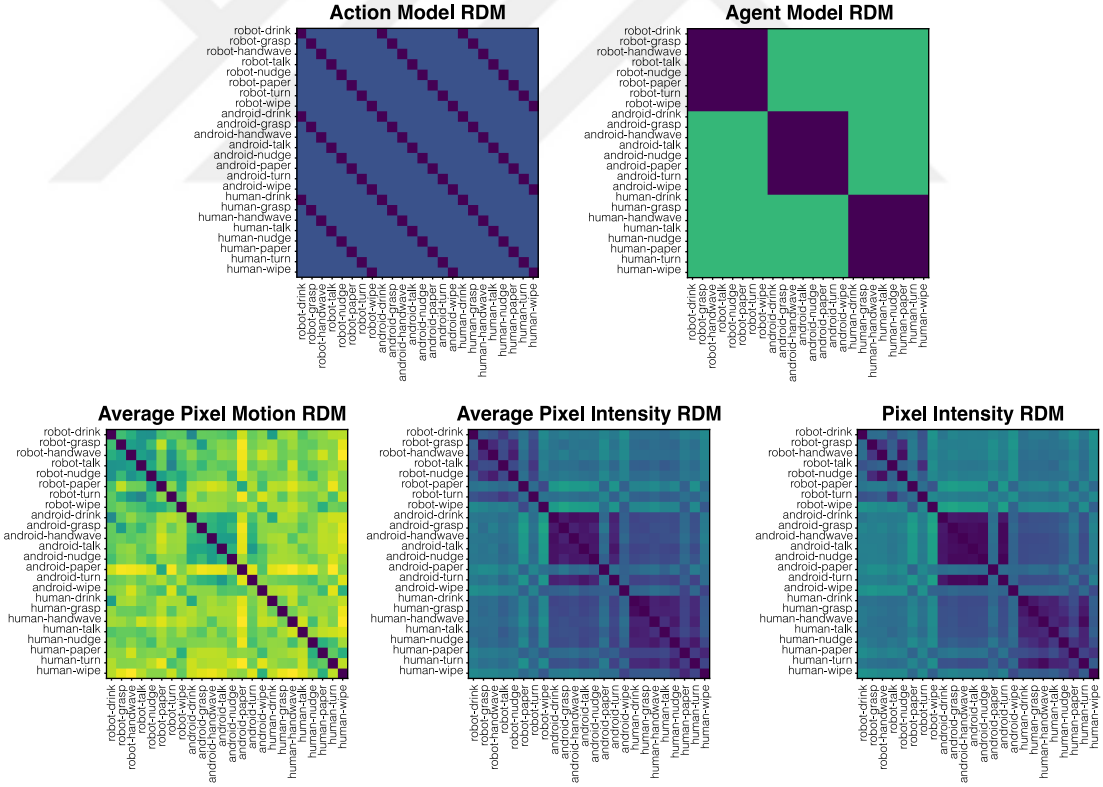

Figure 2.2: Model RDMs used in the multiple linear regression analysis

We generated the low-level model RDMs by converting each stimulus to a vector using the corresponding model's output for that stimulus. We calculated pixel-wise intensity models differently for moving and still conditions. For the

image stimulus, we used pixel intensities of the frames, and for each video stimulus, we averaged the intensities of each pixel over all frames of the video. To create a pixel-wise motion model, we estimated optic flow for all pixels in each frame and averaged it across frames in the video stimulus (A pixel-wise motion model was not created for still stimuli). These processes resulted in 2D pixel-wise intensity and pixel-wise motion matrices per stimulus; we converted these 2D matrices to feature vectors and calculated the correlation distance between each pair of vectors, resulting in a 24×24 RDM for each low-level model.

2.5.3 Multiple Linear Regression Analysis

We performed multiple linear regression analysis (MLR) in a time-resolved manner to investigate the time course of agent information in the EEG data while regressing out potential confounding variables such as motion, pixel intensity, and action type. We predicted EEG RDM using Model RDMs as explanatory variables of a linear regression model at every time point. This process yielded parameter estimates and corresponding two-tailed p values for each model RDM and each time point. We used Variance Inflation Factor (VIF) analysis to ensure that model RDMs used for multiple linear regression analysis do not have a multicollinearity issue.

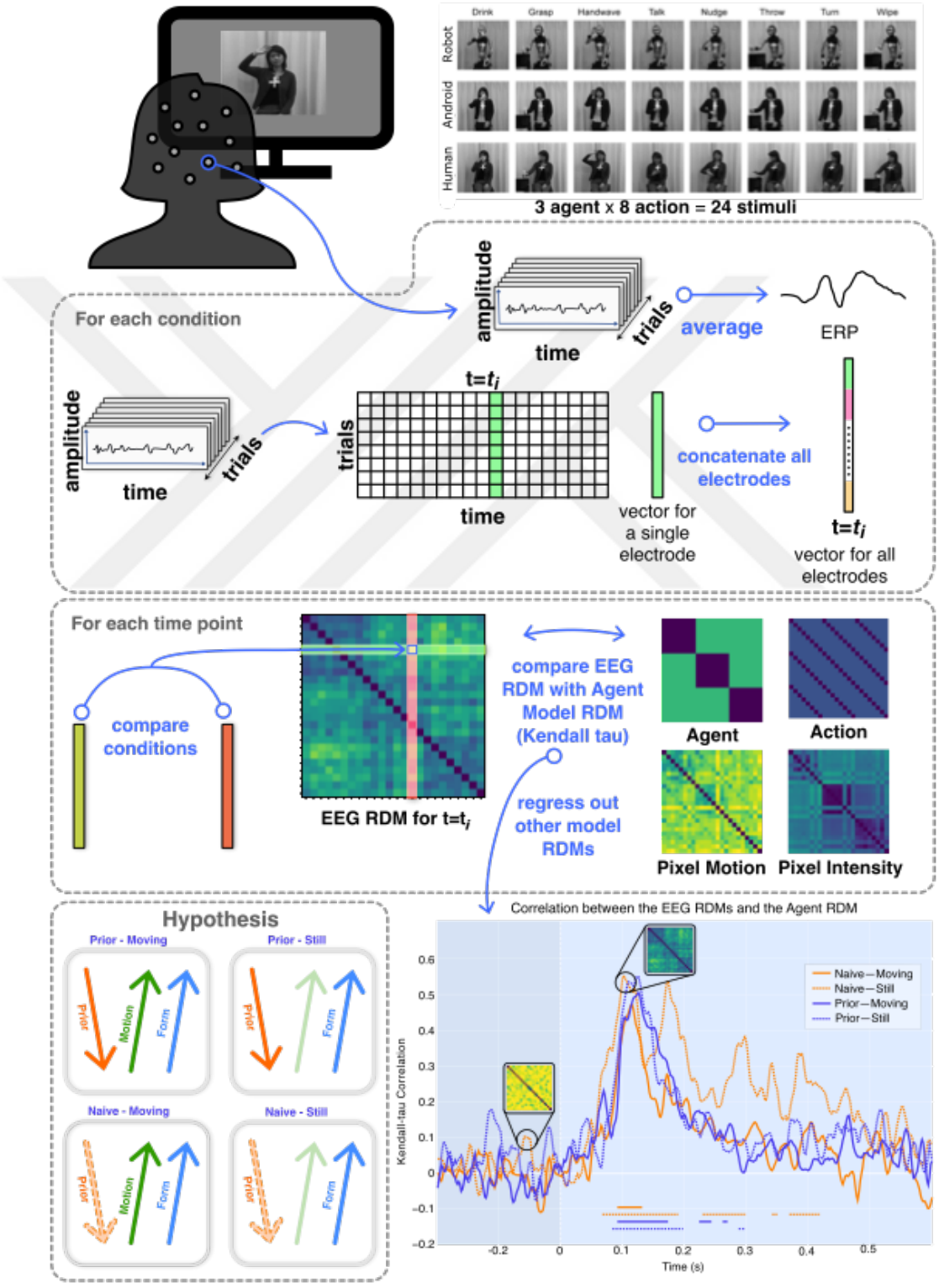

Figure 2.3: Study summary

Chapter 3

Results

3.1 ERP Analyses

We did not have a priori knowledge of the latency and location of the investigated effects of agent identity, availability of motion information, and prior knowledge of agent identities. Thus, we have not focused on a priori ERP component, location, or time window and instead opted for a mass univariate analysis approach. In Figure 3.1, grand-average ERPs for each agent identity and motion availability for both experiments are given. Figure 3.1 only shows a representative subset of electrodes for visual clarity. Both agent identity and motion availability have observable effects in both experiments.

Agent identities are most clearly distinguished in the Oz channel in **Prior Experiment**. Between 80–175 ms, human, android, and robot have clearly distinguishable ERP signatures in the Oz channel. Furthermore, motion availability does not significantly affect the ERP amplitudes during this time period. Again in the Oz channel, between 200–275 ms, robot stimuli have a divergent ERP signature from android and human stimuli. Interestingly, ERP signatures of moving and still robot stimuli are similar. This effect persists in all midline electrodes and time points in **Prior Experiment**; however, the motion availability invariance

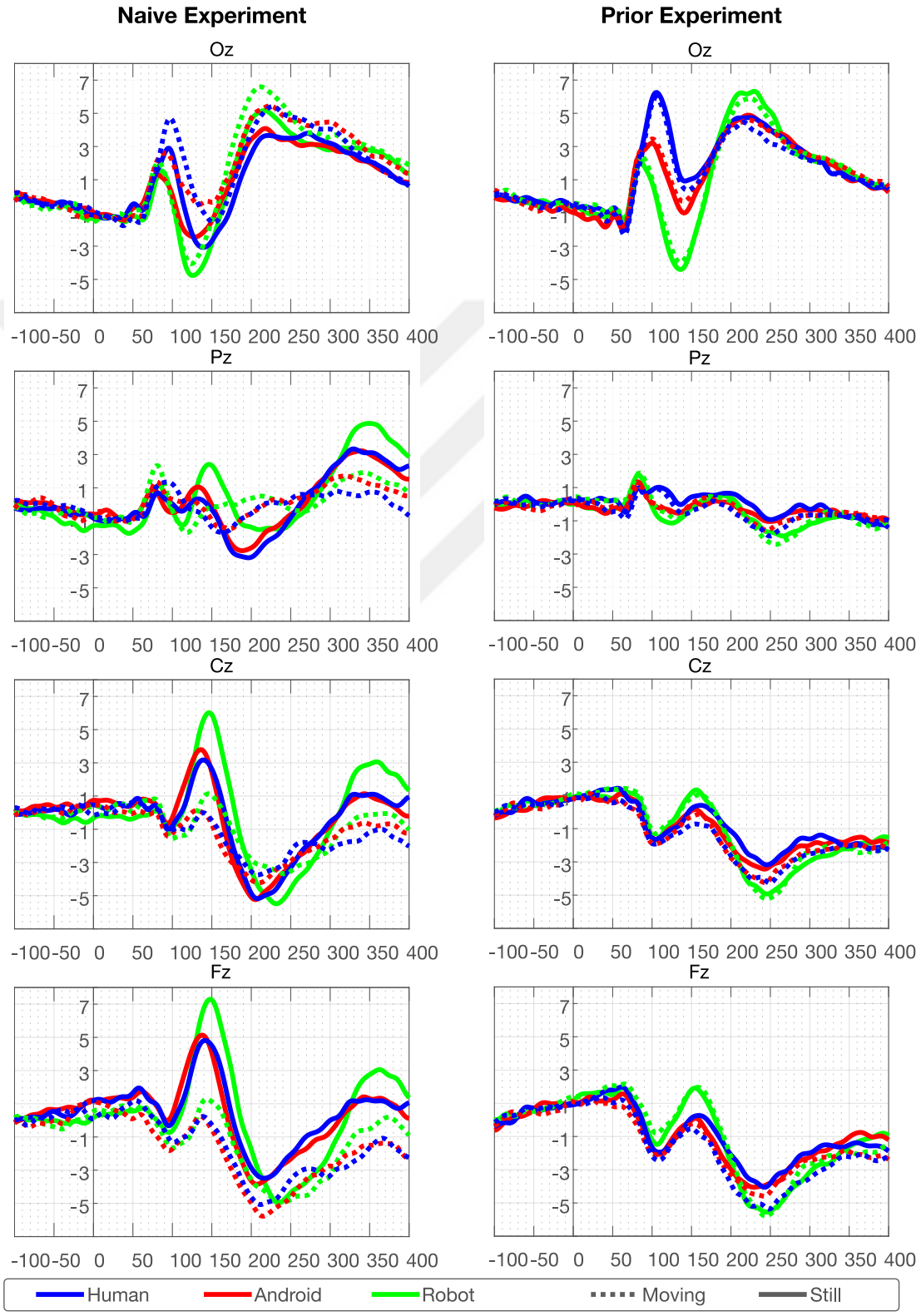

Figure 3.1: Grand-average ERPs from representative midline electrodes

is not just absent in the **Naive Experiment** but the motion availability is a dissociative factor in the ERPs. The availability of motion information selectively affects the ERPs when there is no prior information about the agent identities.

In **Naive Experiment**, ERPs from still stimuli condition have more positively skewed amplitudes in the Pz channel between 250–300 ms and 300–500 ms, in the Cz channel between 100–200 ms and 300–500 ms, in the Fz channel between 100–200 ms and 300–500 ms. The effect is reversed in the Fz channel after 500 ms; moving stimuli are more positive. In the Oz channel, the ERPs cannot be differentiated based on the motion availability at any time period; however, each agent ERP can be differentiated based on the motion availability. ERPs of each agent identity have a more positive amplitude when presented with motion in the Oz channel between 175 and 275 ms.

3.2 Mass-Univariate Analysis on ERPs

Visual inspection of ERPs offered that there may be agent and motion availability effects on the ERPs. The three-way mixed ANOVA supported the observed effects (Figures 3.2, 3.3, 3.4).

The main effect of agent identity (3.2) was observed in the frontal and central channels around 150–180 ms and in the occipital electrodes around 100–150 and 210–240 ms ($p<0.05$, Bonferroni corrected for time points and electrodes).

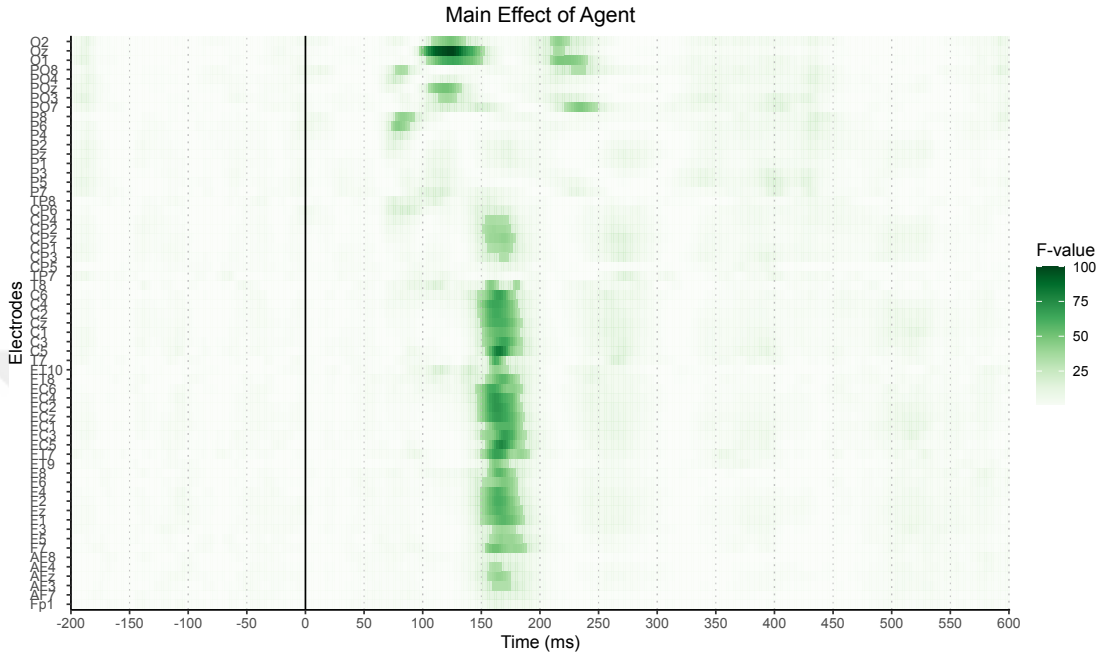

Figure 3.2: Mass-univariate analysis of ERPs: The Main Effect of **Agent**

The heatmap presents the results of a mass-univariate 3x2x2 ANOVA, showcasing the main effect of **Agent**. Time, marked in milliseconds along the x-axis, begins at 0, which denotes the initiation of stimulus presentation. The EEG electrodes, depicted on the y-axis, are organized from posterior to anterior based on their scalp location, adhering to the standard 10-20 system. The color variations in the heatmap, ranging from lighter to darker shades of green, correspond to the F-values procured from the ANOVA. These F-values represent the magnitude of the Agent variable's effect at each individual electrode and time point, with darker hues indicating a stronger effect. Notably, only the statistically significant effects (those with $p < 0.05$ after a Bonferroni correction for both time points and electrodes) are solidly colored. Non-significant effects, in contrast, are shown with transparency.

The main effect of motion availability (3.3) is observed in the frontal electrodes, mostly around 125–200 ms and later around 300 ms ($p<0.05$, Bonferroni corrected). The motion availability effect was earlier and more prominent in the left hemisphere.

The interaction effect of agent identity and motion availability (3.4) is observed in the parietal channels around 200 and 260 ms and in the left frontal

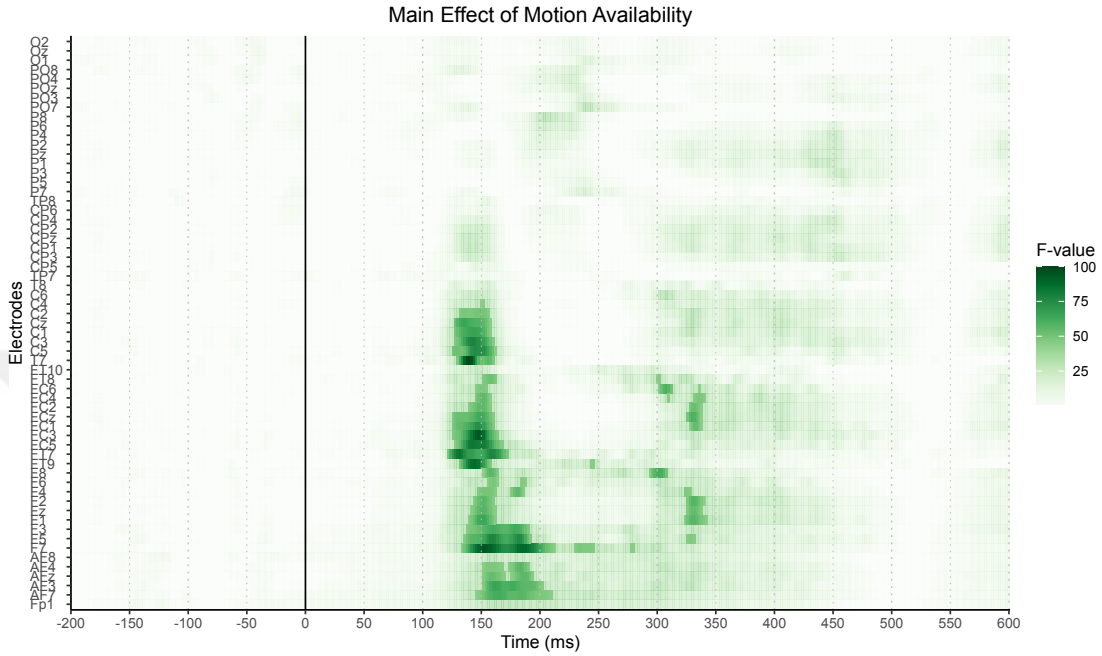

Figure 3.3: Mass-univariate analysis: The Main Effect of **Motion Availability**

The heatmap presents the results of a mass-univariate 3x2x2 ANOVA, showcasing the main effect of **Motion Availability**. Time, marked in milliseconds along the x-axis, begins at 0, which denotes the initiation of stimulus presentation. The EEG electrodes, depicted on the y-axis, are organized from posterior to anterior based on their scalp location, adhering to the standard 10-20 system. The color variations in the heatmap, ranging from lighter to darker shades of green, correspond to the F-values procured from the ANOVA. These F-values represent the magnitude of the Agent variable's effect at each individual electrode and time point, with darker hues indicating a stronger effect. Notably, only the statistically significant effects (those with $p < 0.05$ after a Bonferroni correction for both time points and electrodes) are solidly colored. Non-significant effects, in contrast, are shown with transparency.

channels around 120–200 ms. Prior information had no significant main effect on any electrode or time point. Additionally, the results of the ANOVA revealed no significant interaction effects for prior information and agent type, prior information and motion availability, or prior information, motion availability, and agent type at any electrode or time point ($p>0.05$, Bonferroni corrected for time points and electrodes).

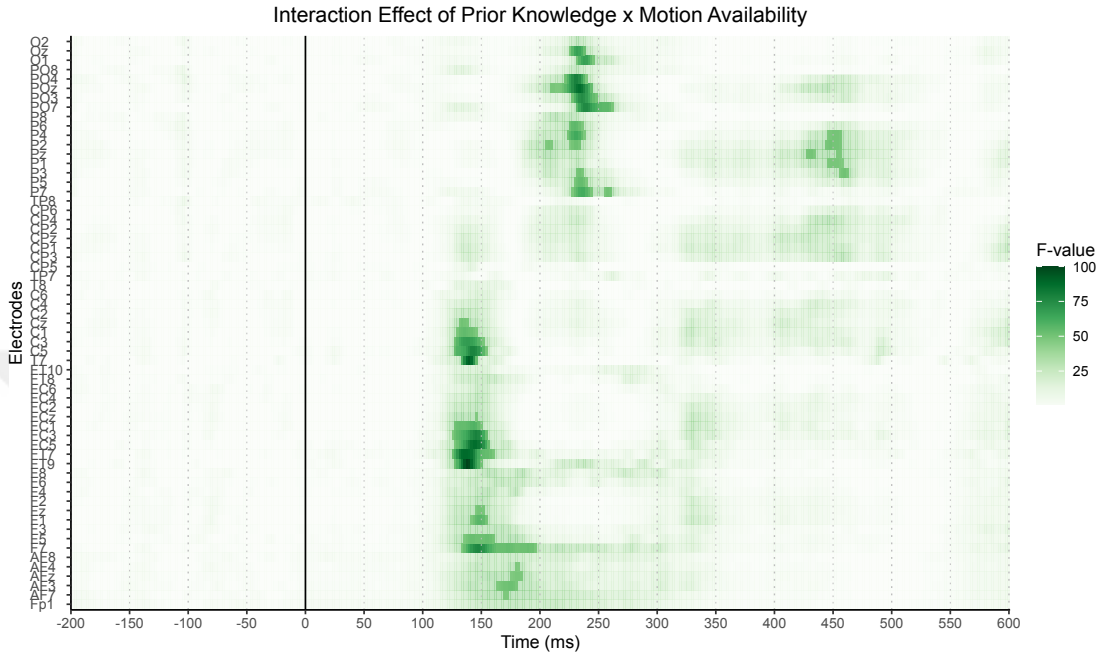

Figure 3.4: Mass-univariate analysis: The Interaction Effect of **Prior Knowledge x Motion Availability**

The heatmap presents the results of a mass-univariate 3x2x2 ANOVA, showcasing the interaction effect of **Prior Knowledge x Motion Availability**. Time, marked in milliseconds along the x-axis, begins at 0, which denotes the initiation of stimulus presentation. The EEG electrodes, depicted on the y-axis, are organized from posterior to anterior based on their scalp location, adhering to the standard 10-20 system. The color variations in the heatmap, ranging from lighter to darker shades of green, correspond to the F-values procured from the ANOVA. These F-values represent the magnitude of the Agent variable's effect at each individual electrode and time point, with darker hues indicating a stronger effect. Notably, only the statistically significant effects (those with $p < 0.05$ after Bonferroni correction for both time points and electrodes) are solidly colored. Non-significant effects, in contrast, are shown with transparency.

We cannot rule out the possibility of low-level visual differences between the stimuli driving the observed effects in the ERPs. Low-level variability in illumination or contrast differences or high-level variability in actions may contribute to the observed ERP differences. The availability of motion information can drive

the difference between the **Still** and **Moving** conditions. Although it is unlikely that all observed differences are driven by low-level visual differences (e.g., motion availability is invariant only in the **Prior Experiment**), we cannot truly disentangle the effects of low-level visual information, motion information, and prior information in the current ERP analysis. Therefore, acknowledging the intricate complexities contributing to the ERP data, we pivoted towards a model-based representational similarity analysis (RSA). This method offers a deeper understanding and addresses the gaps in our current ERP analysis, providing a robust mechanism to disentangle the representations of low-level visual information, motion information, and prior information.

3.3 Representational Similarity Analysis

3.3.1 Whole-Brain RSA

In the whole-brain RSA results, agent identity information peaks around 100 ms in all four conditions (**Naive-Moving**, **Naive-Still**, **Prior-Moving**, and **Prior-Still**); however, only **Naive Experiment** reach statistical significance ($p<0.05$ after Bonferroni correction, refer to Figure 3.5). In **Naive-Moving** condition, EEG RDMs between 118 and 144 ms correlate significantly with Agent Model. In the **Naive-Still** condition, Agent information persists longer with two peaks (70-118 ms and 154-196 ms). **Prior Experiment** had no significant periods after Bonferroni correction; nevertheless, both **Prior-Still** and **Prior-Moving** conditions share similar temporal dynamics with **Naive-Moving** conditions. Previous ERP analysis results indicated that the effects of agent identity varied with electrode location (Figure 3.2). Accordingly, we have decided to examine the effect of agent identity across different electrode sites on the scalp.

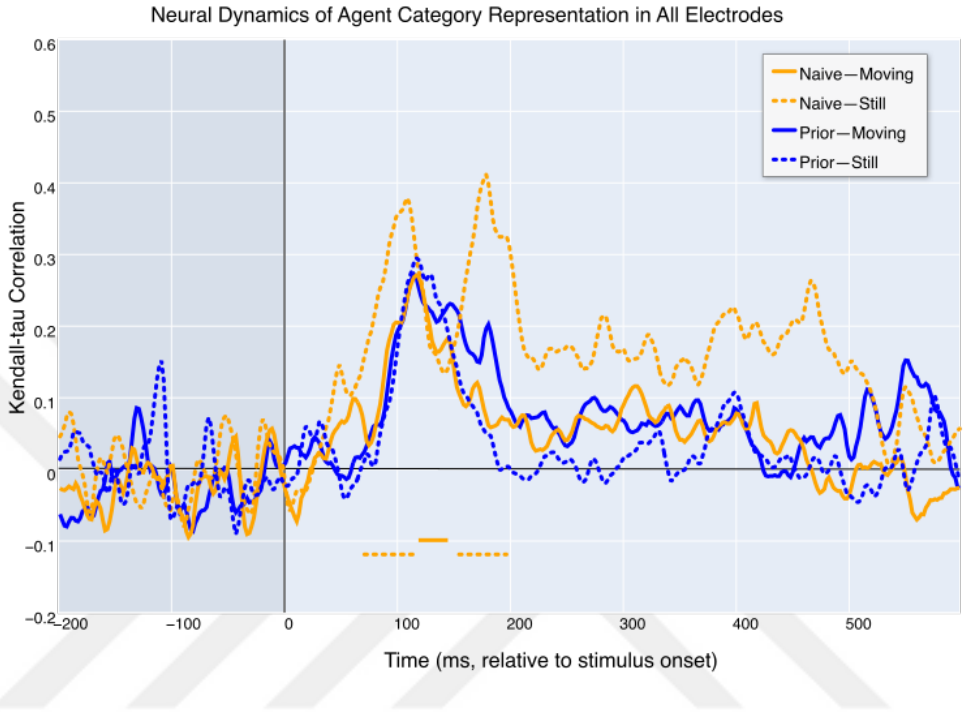

Figure 3.5: Results of the whole brain temporal RSA

The traces depict the time course of Kendall-tau correlation between the EEG representational dissimilarity matrices (RDMs) and the Agent RDM, after regressing out the potential confounding effects of action identity, low-level form and motion information. The significant time windows (after Bonferroni correction) are indicated under the Kendall-tau traces with a horizontal line in the same style corresponding to that condition.

3.3.2 Electrode-specific RSA

Analyzing different electrode sets confirmed our assumptions and revealed prominent differences between electrode sites. In central, frontal, and temporal electrodes, correlation with the Agent model did not reach statistical significance at any time interval after regressing out the effects of other models. On the contrary, occipital and parietal electrode sites have statistically significant time intervals in all four conditions ($p < 0.05$; refer to Figure 3.6).

Correlation traces of both occipital and parietal electrodes share similar temporal dynamics with the whole brain correlation traces, but the specific values

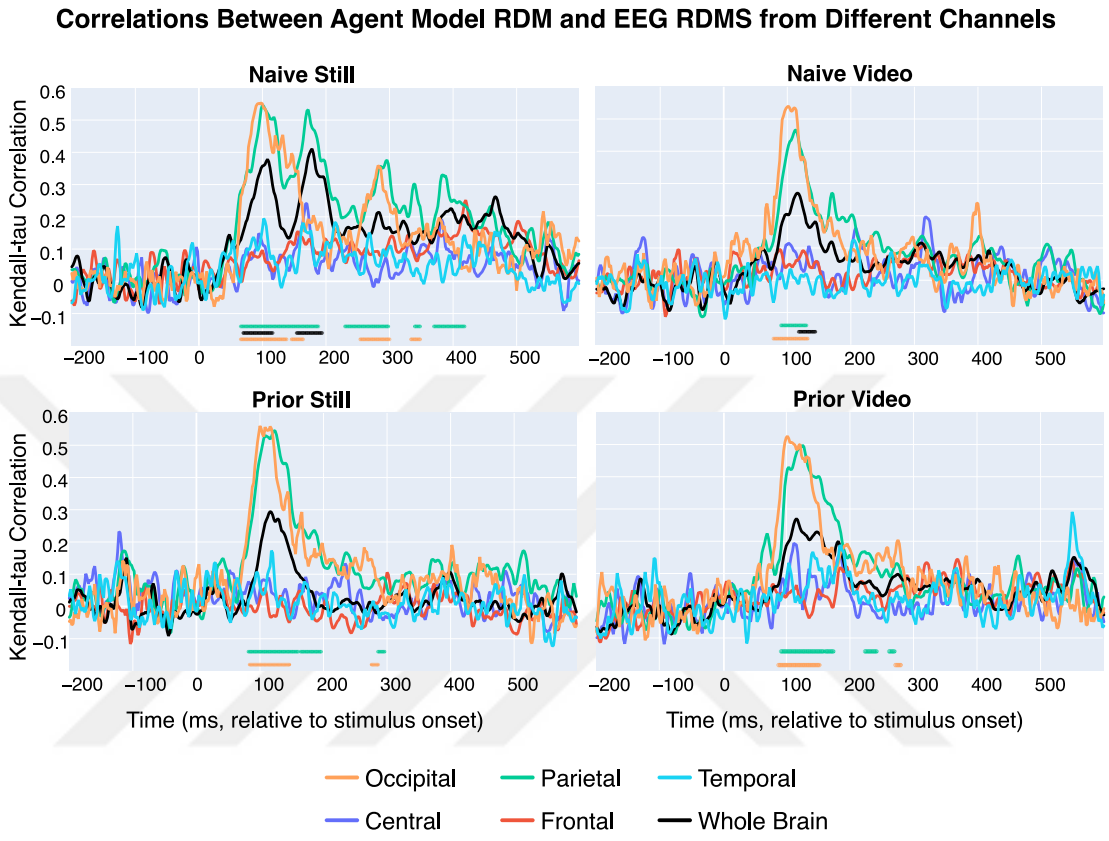

Figure 3.6: Results of the electrode-specific RSA

The traces depict the time course of Kendall-tau correlation between the EEG representational dissimilarity matrices (RDMs) and the Agent RDM, after regressing out the potential confounding effects of action identity, low-level form and motion information. The significant time windows (after Bonferroni correction) are indicated under the Kendall-tau traces with a horizontal line in the same style corresponding to that electrode site.

for occipital and parietal electrodes are higher than the whole set of electrodes, indicating that occipital and parietal channels have a higher signal-to-noise ratio for the current analysis and signals from these electrode sites had driven the previously observed effects using the whole set of electrodes (Figure 3.5). Furthermore, restricting the analysis on occipital or parietal electrodes resulted in statistically significant time intervals for the **Prior Experiment**; these effects were absent in the prior analysis using the whole set of electrodes. Accordingly, we have decided to focus the further examination on the most informative sets of electrodes, namely, occipital and parietal electrodes.

3.3.3 RSA on Parietal and Occipital Electrodes

After removing the influences of action and the low-level visual information (motion and pixel intensity), the emergent EEG response at the parietal electrodes significantly correlated with the Agent model between 66–418 ms in the Naive-Still condition, 90–130 ms in the **Naive-Moving** condition, 82–296 ms in the **Prior-Still** condition and 90–242 ms in the **Prior-Moving** condition (Figure 3.7).

Analysis of occipital electrodes produced agreeing results with the parietal electrodes (Figure 3.6). Parietal electrodes had higher peaks than occipital electrodes in all four conditions. In particular, correlation persisted characteristically after the initial peak for the **Naive-Still** condition. Later peaks in the correlation are better captured in the parietal electrodes; thus, we have opted for presenting the condition comparisons for the parietal electrodes only considering visual clarity.

Correlation traces of **Prior-Moving**, **Prior-Still**, and **Naive-Moving** conditions follow a similar pattern: they climb before the 100 ms mark, peak around 120 ms, then slowly subside before reaching 200 ms. After 200 ms, there is a minor effect (occasionally reaching significance between 200–300 ms in the **Prior Experiment**). **Naive-Moving** condition does not reach statistical significance after the initial peak; regardless, its pattern conforms with the traces for the **Prior Experiment**.

Naive-Still condition diverges from the other conditions: correlation trace has multiple peaks and reaches statistical significance in much later periods. The correlation trace for the **Naive-Still** condition is more extensive than other conditions between 150–450 ms. The latest peak, around 400 ms, is statistically significant.

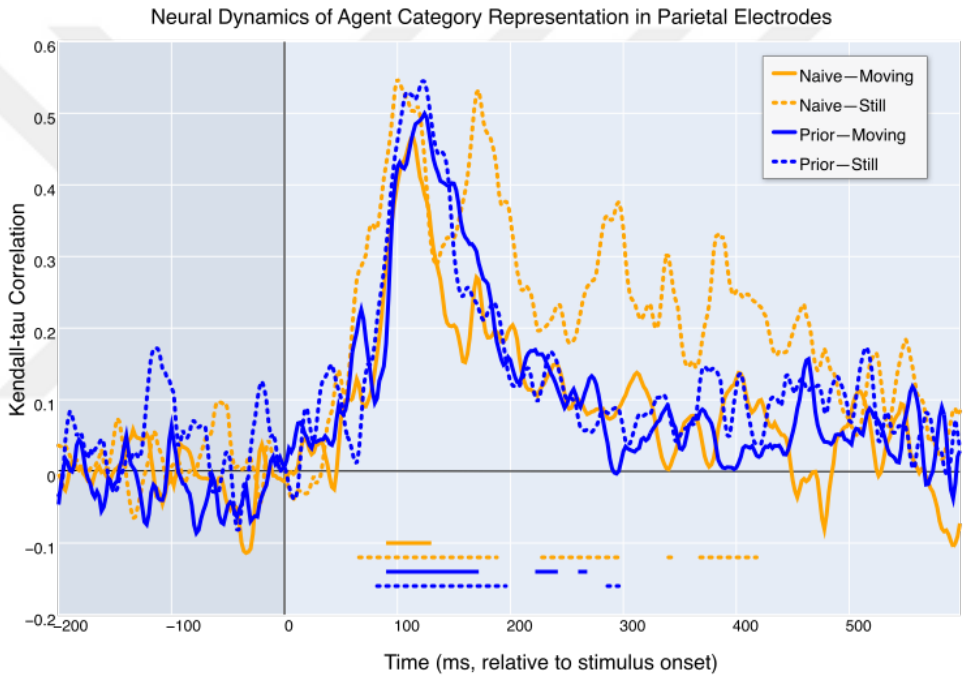

Figure 3.7: Results of temporal RSA over Parietal channels

The traces depict the time course of Kendall-tau correlation between the EEG representational dissimilarity matrices (RDMs) and the Agent RDM, after regressing out the potential confounding effects of action identity, low-level form and motion information. The significant time windows (after Bonferroni correction) are indicated under the Kendall-tau traces with a horizontal line in the same style corresponding to that condition.

Chapter 4

Discussion

Understanding others' actions is a vital skill that humans possess. Comprehending and processing the agents involved in these actions constitutes a critical component of this skill. Consequently, our knowledge about the perception of agents comes from research into Action Observation Network (AON) and the specific nodes of AON that might carry information about agents; however, currently, there is not much research into the temporal characteristics of agent perception and the influence of top-down and bottom-up factors that might alter this dimension.

In this study, we investigated the effects of prior knowledge and motion information on the temporal processing of different agents. We achieved this by manipulating the motion availability (**Still** & **Moving** conditions) and prior knowledge (**Prior** & **Naive** conditions) about human and non-human agents that possess different levels of human-likeness in terms of form and motion. By recording scalp EEG and utilizing ERP analysis and model-based representational similarity analysis (RSA), we aimed to uncover the progression of agent representation in the human brain over time. We hypothesized that the presence of motion-related information and prior information would interact with the temporal attributes of agent processing. Furthermore, we predicted a significant influence of prior information on agent perception when motion information is unavailable due to

the increased ambiguity arising from the absence of motion information.

4.1 The Effects of Motion and Prior Information on ERPs

Our exploration into event-related potentials (ERPs) indicated that both motion and prior information play roles in the temporal processing of different agents. Through visual inspection and ANOVA, we observed effects hinting towards the influence of agent identity and motion availability on the ERPs signals. Our mass univariate tests showed that both agent information and motion availability exerted observable effects in both experiments. However, these effects often interacted and were interdependent in ways that made their individual impacts challenging to isolate. While ERP analysis hinted at agent-specific and motion-specific influences on processing, the technique fell short when attempting to disassociate the role of low-level visual features, motion cues, and form-related information and prior knowledge of the agents. Thus, we turned to a more advanced analysis technique –model-based representational similarity analysis (RSA)– to overcome these shortcomings. This approach provided a more sophisticated method to analyze the complex temporal interplay of form, motion, action information, and prior knowledge of agents, enhancing our capacity to unravel the distinct impact of motion and prior knowledge on agent processing.

4.2 The effect of motion cues without prior information

Our results show that when participants had no prior information about the agents and the stimulus had no motion information, the agent RDM correlated longer with EEG RDMs than in other conditions. Several explanations of this difference include the following: Firstly, motion information is crucial in perceiving

biological motion and action observation [11, 76, 77]. This might also be the case for the perception of agents, as these processes are highly related. Without such essential information, participants might need to process the agent information longer to interpret the stimuli. This heightened cognitive demand and processing could have resulted in stronger and longer processing and representation of agent information. Secondly, the lack of motion information could have led participants to rely more heavily on other cues, such as shape, texture, or other static features of the agents, to form their mental representations. Consequently, agent representations may have had a more pronounced impact on the EEG signals. While we used regression with low-level visual models to prevent low-level details from influencing the correlations with the agent model, it is possible that these models could not account for all low-level visual differences between the agents, such as texture or shape, which might be relied on more in the when there are no motion cues and prior information.

4.3 The effect of prior information without motion cues

While the absence of motion information led to longer significant correlations between Agent RDM and EEG RDMs, our results show this effect only in the **Naive Experiment**, where participants did not have prior information about the agents and their actions. In the **Prior Experiment**, the significance of the correlation traces was similar in both **Still** and **Moving** conditions. This difference suggests that prior information about the agents and their actions reduced the advantage of having motion information.

The reduced advantage of motion information could be attributed to several factors. Firstly, without prior knowledge and motion cues, resolving the ambiguity around agent identities might be more challenging, leading to longer and stronger correlations in the **Naive-Still** condition. This argument aligns with Bayesian inference models, which suggest that ambiguous stimuli result

in increased and prolonged processing due to the prediction errors they cause, prompting recurrent updates to the prior parameters of the Bayesian model [78, 79, 80, 81].

However, it is also possible that prior knowledge of agents might lead to a decrease in prediction error, resulting in reduced activation [82, 80]. Another possibility is that top-down modulation of perception, driven by prior knowledge, could allow participants to process the critical features of agents and their actions more effectively using the available visual information in the **Still** condition.

To effectively compare and disentangle these alternative explanations, further research focusing solely on the effects of prior knowledge and utilizing conditions with no expectations is necessary.

4.4 Efficient Ambiguity Resolution via Motion Cues or Prior Information

In our study, we did not observe any direct effect of either motion availability or prior information in the temporal RSA results. The effect of motion was only present in the **Naive Experiment**, specifically between the **Naive-Still** and **Naive-Moving** conditions. Similarly, the effect of prior information was only present in the **Moving** conditions, specifically between the **Prior-Moving** and **Naive-Moving** conditions.

Interestingly, the temporal signature of agent identity representation was different only for the **Naive-Still** condition. All other conditions (**Prior-Still**, **Prior-Moving**, and **Naive-Moving**) exhibited similar temporal signatures. This suggests that either motion or prior information is sufficient to resolve the ambiguity in agent identities.

It's important to note that the advantageous effects of motion and prior information do not result in stronger representation but rather shorter representation.

This shorter representation could indicate more efficient processing and effective resolution of ambiguity. In other words, the presence of either motion cues or prior information seems to streamline the processing of agent information, allowing the brain to resolve ambiguity more quickly. This finding underscores the importance of both bottom-up cues (like motion) and top-down processes (like prior knowledge) in shaping our perception of agents and their actions.

Furthermore, it is plausible that the longer representation of agent information in the **Naive-Still** condition could be attributed to recurrent processing, which might be necessary to resolve the ambiguity when neither motion cues nor prior information are available. This suggests that efficient processing, as indicated by shorter representation, might be associated with a reduced need for further recurrent processing. However, further research is needed to test this hypothesis and to fully understand the complex interplay between bottom-up and top-down processes in the temporal dynamics of agent perception.

4.5 Implications

Our study provides a clearer insight into the ongoing discussion about agent perception. We examined how prior knowledge and motion cues impact our understanding of agent identity, shedding light on an area not well explored in cognitive science.

As robots become more prevalent in our everyday lives, it becomes crucial to understand more about perception of robotic agents. Our findings can shape the design of robots and androids in settings from caregiving to manufacturing, ensuring effective human-robot interactions.

For industries that depend on animations, virtual reality, or video game design, understanding how humans perceive agents based on motion cues and prior information could inform the creation of more immersive and engaging content.

In the field of education, when utilizing technology such as virtual tutors or avatars, understanding the role of prior knowledge and motion can aid in designing better pedagogical tools. For instance, educators might offer preliminary information about virtual agents to learners, ensuring a smoother and more prolonged engagement with the learning material.

Computational EEG methods applied here, combined with the insights on agent perception, could lead to advancements in therapeutic interventions, especially for individuals with disorders that impede their ability to interpret or anticipate the actions of others, such as certain autism spectrum disorders.

In conclusion, by probing the temporal dynamics of agent perception, this study not only deepens our comprehension of a critical aspect of human cognition but also offers practical insights for various fields, ranging from technology design to education.

4.6 Limitations

While this study provides valuable insights into the interactive effect of availability of motion cues and prior knowledge on the temporal processing of different agents, it is important to acknowledge its limitations.

Firstly, our study relied on the assumption that the regression with low-level visual models would prevent low-level details from influencing the correlations with the agent model. However, it is possible that these models could not account for all low-level visual differences between the agents, such as texture or shape, which might be relied on more in the absence of motion cues and prior information.

Another limitation pertains to the manipulation of form and motion cues in our study. We did not investigate the interaction of top-down prior effects and

bottom-up visual information by gradually and parametrically degrading available form and motion cues. As such, our study may not fully reflect scenarios where both form and motion cues are systematically controlled. In our design, we only controlled the presence of motion by presenting static images or dynamic videos of agents, creating a binary condition where motion cues were either available or not. Similarly, we did not modulate the available form information in the presented stimuli. Instead, we attempted to maintain ambiguity in agent identities in the static stimuli by selecting an agent set with overlapping motion and form characteristics. This approach, while practical, may not fully capture the complexities of agent perception when form cues are degraded or manipulated in a more nuanced manner.

In addition, the agents used in our study were not all biologically or evolutionarily significant. Specifically, the non-biological agents (**Robot**, **Android**) may not be as familiar to participants as human agents, and their evolutionary significance is certainly different. This raises questions about the generalizability of our findings. While our results provide valuable insights into agent perception, they may not fully extend to the observation of different conspecifics or other biologically significant agents. The use of robotic agents, while interesting, introduces unique challenges and considerations in the study of agent perception.

Lastly, our study does not provide specific information on the neural sources of the observed interactive effect between the presence of motion cues and top-down prior effects. While our results suggest an influence of prior knowledge on agent perception, the specific neural mechanisms and pathways underlying this effect remain unclear. For instance, the effect of prior knowledge might be carried in certain oscillatory bands. Furthermore, our results did not inform on information transfer between regions. Understanding the specific neural regions involved, their interconnections, and the nature of their communication (e.g., via oscillations) is crucial for a comprehensive understanding of agent perception. This is an area that warrants further investigation.

4.7 Future Directions

Given the limitations of the current study, several future directions could be pursued to further our understanding of the interactive effect of the availability of motion cues and prior knowledge on the temporal processing of different agents.

Firstly, while our study assumed that regression with low-level visual models would control for low-level visual differences, future studies could employ more sophisticated techniques to control for these differences. One such technique is the "Bubbles technique", which allows for the independent, gradual, and parametric control of available visual information [83, 56]. This technique could be used to systematically control available form and motion information, providing a more nuanced understanding of their roles in agent perception.

Secondly, future research could benefit from the use of parametrically created agents that match visual characteristics and modulate features parametrically. This approach, as demonstrated in Doris Tsao's work on face feature space in macaque monkeys, allows for the creation of parameterized realistic face stimuli [84, 85, 86]. Such stimuli could be used to further investigate the role of form cues in agent perception.

Thirdly, to address the limitation regarding the lack of information on the neural sources of the observed effects, future studies could employ EEG/MEG source reconstruction techniques. These techniques would not only provide insights into the specific brain regions involved but also take advantage of the time-precise response and oscillatory responses provided by EEG/MEG.

Lastly, the strength of bottom-up and top-down signals might be tested by examining specific frequency bands. In the literature, different frequency bands such as gamma, beta, theta, and alpha have been associated with different types of information flow [87]. Although, research indicates that specific frequency bands have distinct functions across the brain. There is sufficient empirical evidence that shows frequency-specific signals are informative on the feedforward and feedback

flow across the hierearchy. In primates, Bastos et al. reported that feedforward influences are carried by theta-band (∼4Hz) and gamma-band (∼60–80Hz) synchronization, and feedback influences by beta-band (∼14–18Hz) synchronization in the primate visual cortex. Similarly, Michalareas et al. show that gamma band carries feedforward signals, whereas alpha-beta band carries feedback signals predominately in the in human visual cortex [88]. These oscillatory signatures are also linked with prediction-related representations in predictive processing theories. Bastos et al. put forward "predictive routing", a neural implementation for predictive processing [89]. According to predictive routing model, predictions carried by lower frequency alpha/beta rhythms inhibit low-level sensory feedforward processing. In short, predicted sensory inputs, predictions (carried by alpha/beta rhythms) and prepare (inhibit) the pathways that would process the predicted input (which is indexed by reduced gamma). On the contrary, unpredicted sensory stimulus, would result in enhanced gamma response because the pathway that would process the unpredicted stimulus is not prepared or inhibited. Their model might be useful to test whether Prior conditions have stronger alpha/beta rhythms and Naive conditions have stronger gamma rhythms.

Examining these frequency bands could provide insights into the nature of the information flow during agent perception. Furthermore, connectivity measures such as Granger causality and theta band coherence could be used to examine information flow between regions. This could provide insights into the source of prior information on agent identities, which might come from the hippocampus or other cortical regions.

Chapter 5

Conclusion

In conclusion, this study investigated the temporal characteristics of the visual processing of agents during action observation by manipulating prior knowledge and motion information. Through time-resolved representational similarity analysis and regression, we revealed that the processing of agents in EEG depends on the availability of motion information and prior information. Specifically, in the naive experiment, agent information was available longer during the still condition than in the moving condition. In contrast, agent information was present for similar durations in still and moving conditions of the prior experiment. These findings suggest that prior knowledge and motion information interactively modulate the duration of the processing of agent information. Our results underscore the critical role of prior knowledge and motion cues in shaping the processing of agent information, highlighting the complex interplay between top-down modulation and bottom-up cues in driving the perception of agents and their actions.

Bibliography

[1] S. Caspers, K. Zilles, A. R. Laird, and S. B. Eickhoff, "Ale meta-analysis of action observation and imitation in the human brain," *Neuroimage*, vol. 50, no. 3, pp. 1148–1167, 2010.

[2] A. Platonov and G. A. Orban, "Action observation: The less-explored part of higher-order vision," *Scientific Reports*, vol. 6, pp. 1–13, 2016.

[3] G. A. Orban, "Action observation as a visual process: Different classes of actions engage distinct regions of human PPC," *Cultural Patterns and Neurocognitive Circuits II: East-West Connections*, pp. 1–32, 2017.

[4] K. Nelissen, E. Borra, M. Gerbella, S. Rozzi, G. Luppino, W. Vanduffel, G. Rizzolatti, and G. A. Orban, "Action observation circuits in the macaque monkey cortex," *Journal of Neuroscience*, vol. 31, no. 10, pp. 3743–3756, 2011.

[5] J. Sliwa and W. A. Freiwald, "A dedicated network for social interaction processing in the primate brain," *Science*, vol. 356, pp. 745–749, 5 2017.

[6] L. Tarhan and T. Konkle, "Sociality and interaction envelope organize visual action representations," *Nature Communications*, vol. 11, pp. 1–11, 2020.

[7] M. F. Wurm, A. Caramazza, and A. Lingnau, "Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity," *Journal of Neuroscience*, vol. 37, pp. 562–575, 2017.

[8] M. Mishkin, L. G. Ungerleider, and K. A. Macko, "Object vision and spatial vision: two cortical pathways," *Trends in neurosciences*, vol. 6, pp. 414–417, 1983.

[9] J. Vangeneugden, M. V. Peelen, D. Tadin, and L. Battelli, "Distinct neural mechanisms for body form and body motion discriminations," *Journal of Neuroscience*, vol. 34, pp. 574–585, 2014. Body form -> EBA
Body motion -> pSTS.

[10] J. Lange and M. Lappe, "A model of biological motion perception from configural form cues," *Journal of Neuroscience*, vol. 26, pp. 2894–2906, 3 2006.

[11] M. A. Giese and T. Poggio, "Neural mechanisms for the recognition of biological movements," *Nature Reviews Neuroscience*, vol. 4, no. 3, pp. 179–192, 2003.

[12] M. Shiffrar and J. J. Freyd, "Apparent motion of the human body," *Psychological Science*, vol. 1, pp. 257–264, 7 1990.

[13] P. Cavanagh, A. T. Labianca, and I. M. Thornton, "Attention-based visual routines: sprites," *Cognition*, vol. 80, pp. 47–60, 6 2001.

[14] I. M. Thornton, R. A. Rensink, and M. Shiffrar, "Active versus passive processing of biological motion," *Perception*, vol. 31, pp. 837–853, 2002.

[15] A. R. Hunt and F. Halper, "Disorganizing biological motion," *Journal of Vision*, vol. 8, no. 9, pp. 12–12, 2008.

[16] A. Jacobs, J. Pinto, and M. Shiffrar, "Experience, context, and the visual perception of human movement.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 30, no. 5, p. 822, 2004.

[17] F. P. De Lange, M. Heilbron, and P. Kok, "How do expectations shape perception?," *Trends in cognitive sciences*, vol. 22, no. 9, pp. 764–779, 2018.

[18] L. M. Vaina, M. Lemay, D. C. Bienfang, A. Y. Choi, and K. Nakayama, "Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: A case study," *Visual neuroscience*, vol. 5, no. 4, pp. 353–369, 1990.

[19] P. McLeod, "Preserved and impaired detection of structure from motion by a'motion-blind" patient," *Visual Cognition*, vol. 3, no. 4, pp. 363–392, 1996.

[20] J. A. Beintema and M. Lappe, "Perception of biological motion without local image motion," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 5661–5663, 4 2002.

[21] T. Jellema and D. I. Perrett, "Cells in monkey sts responsive to articulated body motions and consequent static posture: a case of implied motion?," *Neuropsychologia*, vol. 41, pp. 1728–1737, 1 2003.

[22] S. Gilaie-Dotan, A. P. Saygin, L. J. Lorenzi, R. Egan, G. Rees, and M. Behrmann, "The role of human ventral visual cortex in motion perception," *Brain*, vol. 136, p. 2784, 2013.

[23] M. Bar, K. S. Kassam, A. S. Ghuman, J. Boshyan, A. M. Schmidt, A. M. Dale, M. S. Hämäläinen, K. Marinkovic, D. L. Schacter, B. R. Rosen, and E. Halgren, "Top-down facilitation of visual recognition," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 449–454, 2006.

[24] N. T. Markov, J. Vezoli, P. Chameau, A. Falchier, R. Quilodran, C. Huissoud, C. Lamy, P. Misery, P. Giroud, S. Ullman, P. Barone, C. Dehay, K. Knoblauch, and H. Kennedy, "Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex," *Journal of Comparative Neurology*, vol. 522, pp. 225–259, 1 2014.

[25] A. A. Sokolov, M. Erb, A. Gharabaghi, W. Grodd, M. S. Tatagiba, and M. A. Pavlova, "Biological motion processing: The left cerebellum communicates with the right superior temporal sulcus," *NeuroImage*, vol. 59, pp. 2824–2830, 2 2012.

[26] D. Pitcher and L. G. Ungerleider, "Evidence for a third visual pathway specialized for social perception," *Trends in Cognitive Sciences*, vol. 25, pp. 100–110, 2021.

[27] T. Allison, A. Puce, and G. McCarthy, "Social perception from visual cues: role of the sts region," *Trends in Cognitive Sciences*, vol. 4, pp. 267–278, 7 2000.

[28] B. Deen, K. Koldewyn, N. Kanwisher, and R. Saxe, "Functional organization of social perception and cognition in the superior temporal sulcus," *Cerebral Cortex*, vol. 25, pp. 4596–4609, 11 2015.

[29] B. Deen, R. Saxe, and N. Kanwisher, "Processing communicative facial and vocal cues in the superior temporal sulcus," *NeuroImage*, vol. 221, p. 117191, 11 2020.

[30] T. Serre, A. Oliva, and T. Poggio, "A feedforward architecture accounts for rapid categorization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 6424–6429, 4 2007.

[31] H. Jhuang, T. Serre, L. Wolf, and T. Poggio, "A biologically inspired system for action recognition," *Proceedings of the IEEE International Conference on Computer Vision*, 2007.

[32] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, "How does the brain solve visual object recognition?," *Neuron*, vol. 73, p. 415, 2 2012.

[33] R. C. O'Reilly, D. Wyatte, S. Herd, B. Mingus, and D. J. Jilk, "Recurrent processing during object recognition," *Frontiers in Psychology*, vol. 4, p. 124, 2013.

[34] R. S. van Bergen and N. Kriegeskorte, "Going in circles is the way forward: the role of recurrence in visual inference," *Current Opinion in Neurobiology*, vol. 65, pp. 176–193, 12 2020.

[35] K. Kar, J. Kubilius, K. Schmidt, E. B. Issa, and J. J. DiCarlo, "Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior," *Nature Neuroscience 2019 22:6*, vol. 22, pp. 974–983, 4 2019.

[36] R. S. van Bergen and N. Kriegeskorte, "Going in circles is the way forward: the role of recurrence in visual inference," *Current Opinion in Neurobiology*, vol. 65, pp. 176–193, 12 2020.

[37] J. Kornmeier, K. Bhatia, and E. Joos, "Top-down resolution of visual ambiguity - knowledge from the future or footprints from the past?," *PLoS ONE*, vol. 16, 10 2021.

[38] S. Ondobaka, F. P. D. Lange, M. Wittmann, C. D. Frith, and H. Bekkering, "Interplay between conceptual expectations and movement predictions underlies action understanding," *Cerebral Cortex*, vol. 25, pp. 2566–2573, 9 2015.

[39] K. C. Schenke, N. A. Wyer, and P. Bach, "The things you do: Internal models of others' expected behaviour guide action observation," *PLOS ONE*, vol. 11, p. e0158910, 7 2016.

[40] P. O. Jacquet, A. C. Roy, V. Chambon, A. M. Borghi, R. Salemme, A. Farnè, and K. T. Reilly, "Changing ideas about others' intentions: updating prior expectations tunes activity in the human motor system," *Scientific Reports 2016 6:1*, vol. 6, pp. 1–15, 5 2016.

[41] P. Bach and K. C. Schenke, "Predictive social perception: Towards a unifying framework from action observation to person knowledge," *Social and Personality Psychology Compass*, vol. 11, p. e12312, 7 2017.

[42] V. Chambon, P. Domenech, P. O. Jacquet, G. Barbalat, S. Bouton, E. Pacherie, E. Koechlin, and C. Farrer, "Neural coding of prior expectations in hierarchical intention inference," *Scientific Reports 2017 7:1*, vol. 7, pp. 1–16, 4 2017.

[43] T. C. Kietzmann, C. J. Spoerer, L. K. Sörensen, R. M. Cichy, O. Hauk, and N. Kriegeskorte, "Recurrence is required to capture the representational dynamics of the human visual system," *Proceedings of the National Academy of Sciences*, vol. 116, no. 43, pp. 21854–21863, 2019.

[44] R. P. Rao and D. H. Ballard, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects," *Nature Neuroscience 1999 2:1*, vol. 2, pp. 79–87, 1 1999.

[45] K. Friston, "A theory of cortical responses," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, pp. 815–836, 4 2005.

[46] K. Friston and S. Kiebel, "Predictive coding under the free-energy principle," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, pp. 1211–1221, 5 2009.

[47] P. Kok, J. F. M. Jehee, and F. P. D. Lange, "Less is more: expectation sharpens representations in the primary visual cortex," *Neuron*, vol. 75, pp. 265–270, 2012.

[48] N. Rungratsameetaweemana, S. Itthipuripat, A. Salazar, and J. T. Serences, "Expectations do not alter early sensory processing during perceptual decision-making," *Journal of Neuroscience*, vol. 38, pp. 5632–5648, 2018.

[49] P. Sterzer, A. Kleinschmidt, and G. Rees, "The neural bases of multistable perception," *Trends in cognitive sciences*, vol. 13, no. 7, pp. 310–318, 2009.

[50] A. Pantle and K. Turano, "Visual resolution of motion ambiguity with periodic luminance-and contrast-domain stimuli," *Vision Research*, vol. 32, no. 11, pp. 2093–2106, 1992.

[51] J. H. Salge, S. Pollmann, and R. R. Reeder, "Anomalous visual experience is linked to perceptual uncertainty and visual imagery vividness," *Psychological research*, vol. 85, pp. 1848–1865, 2021.

[52] M. W. Flounders, C. González-García, R. Hardstone, and B. J. He, "Neural dynamics of visual ambiguity resolution by perceptual prior," *Elife*, vol. 8, p. e41861, 2019.

[53] R. Hardstone, M. Zhu, A. Flinker, L. Melloni, S. Devore, D. Friedman, P. Dugan, W. K. Doyle, O. Devinsky, and B. J. He, "Long-term priors influence visual perception through recruitment of long-range feedback," *Nature Communications 2021 12:1*, vol. 12, pp. 1–15, 11 2021.

[54] J. Bill, S. J. Gershman, and J. Drugowitsch, "Visual motion perception as online hierarchical inference," *Nature communications*, vol. 13, no. 1, p. 7403, 2022.

[55] P. Neri, M. C. Morrone, and D. C. Burr, "Seeing biological motion," *Nature*, vol. 395, no. 6705, pp. 894–896, 1998.

[56] S. M. Thurman and E. D. Grossman, "Temporal "bubbles" reveal key features for point-light biological motion perception," *Journal of Vision*, vol. 8, pp. 1–11, 2008.

[57] S. M. Thurman, M. A. Giese, and E. D. Grossman, "Perceptual and computational analysis of critical features for biological motion," *Journal of Vision*, vol. 10, no. 12, pp. 15–15, 2010.

[58] A. P. Saygin, T. Chaminade, H. Ishiguro, J. Driver, and C. Frith, "The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions," *Social cognitive and affective neuroscience*, vol. 7, no. 4, pp. 413–422, 2012.

[59] B. A. Urgen, M. Kutas, and A. P. Saygin, "Uncanny valley as a window into predictive processing in the social brain," *Neuropsychologia*, vol. 114, pp. 181–185, 2018.

[60] V. Gazzola, G. Rizzolatti, B. Wicker, and C. Keysers, "The anthropomorphic brain: the mirror neuron system responds to human and robotic actions," *Neuroimage*, vol. 35, no. 4, pp. 1674–1684, 2007.

[61] A. Wykowska, T. Chaminade, and G. Cheng, "Embodied artificial agents for understanding human social cognition," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, 5 2016.

[62] C. Press, "Action observation and robotic agents: Learning and anthropomorphism," *Neuroscience & Biobehavioral Reviews*, vol. 35, pp. 1410–1418, 5 2011.

[63] Y. F. Tai, C. Scherfler, D. J. Brooks, N. Sawamoto, and U. Castiello, "The human premotor cortex is 'mirror' only for biological actions to-be-grasped object were also included (see experi-mental procedures). our prediction arises from previous neuroimaging," *Current Biology*, vol. 14, pp. 117–120, 2004.

[64] N. Miura, M. Sugiura, M. Takahashi, Y. Sassa, A. Miyamoto, S. Sato, K. Horie, K. Nakamura, and R. Kawashima, "Effect of motion smoothness on brain activity while observing a dance: An fmri study using a humanoid robot," *Social neuroscience*, vol. 5, no. 1, pp. 40–58, 2010.

[65] J. Hogeveen and S. S. Obhi, "Social interaction enhances motor resonance for observed human actions," *Journal of Neuroscience*, vol. 32, pp. 5984–5989, 4 2012.

[66] L. M. Oberman, J. P. McCleery, V. S. Ramachandran, and J. A. Pineda, "Eeg evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots," *Neurocomputing*, vol. 70, no. 13-15, pp. 2194–2203, 2007.

[67] E. S. Cross, R. Liepelt, A. F. Antonia, J. Parkinson, R. Ramsey, W. Stadler, and W. Prinz, "Robotic movement preferentially engages the action observation network," *Human Brain Mapping*, vol. 33, no. 9, pp. 2238–2254, 2012.

[68] A. Kupferberg, M. Iacoboni, V. Flanagin, M. Huber, A. Kasparbauer, T. Baumgartner, G. Hasler, F. Schmidt, C. Borst, and S. Glasauer, "Frontoparietal coding of goal-directed actions performed by artificial agents," *Human Brain Mapping*, vol. 39, pp. 1145–1162, 3 2018.

[69] M. Mori, "The uncanny valley: the original essay by masahiro mori," *IEEE Spectrum*, vol. 6, 1970.

[70] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robotics & automation magazine*, vol. 19, no. 2, pp. 98–100, 2012.

[71] N. Kriegeskorte, M. Mur, and P. Bandettini, "Representational similarity analysis - connecting the branches of systems neuroscience," *Frontiers in Systems Neuroscience*, vol. 2, no. NOV, pp. 1–28, 2008.

[72] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[73] J. Lopez-Calderon and S. J. Luck, "Erplab: an open-source toolbox for the analysis of event-related potentials," *Frontiers in human neuroscience*, vol. 8, p. 213, 2014.

[74] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

[75] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019.

[76] R. Blake and M. Shiffrar, "Perception of human motion," *Annu. Rev. Psychol.*, vol. 58, pp. 47–73, 2007.

[77] J. O. Garcia and E. D. Grossman, "Necessary but not sufficient: Motion perception is required for perceiving biological motion," *Vision research*, vol. 48, no. 9, pp. 1144–1149, 2008.

[78] D. Kaiser, R. Stecher, and K. Doerschner, "Eeg decoding reveals neural predictions for naturalistic material behaviors," *bioRxiv*, pp. 2023–02, 2023.

[79] B. A. Urgen and A. P. Saygin, "Predictive processing account of action perception: Evidence from effective connectivity in the action observation network," *cortex*, vol. 128, pp. 132–142, 2020.

[80] C. Summerfield and F. P. De Lange, "Expectation in perceptual decision making: neural and computational mechanisms," *Nature Reviews Neuroscience*, vol. 15, no. 11, pp. 745–756, 2014.

[81] D. Kersten, P. Mamassian, and A. Yuille, "Object perception as bayesian inference," *Annu. Rev. Psychol.*, vol. 55, pp. 271–304, 2004.

[82] C. Press, P. Kok, and D. Yon, "The perceptual prediction paradox," *Trends in Cognitive Sciences*, vol. 24, no. 1, pp. 13–24, 2020.

[83] F. Gosselin and P. G. Schyns, "Bubbles: a technique to reveal the use of information in recognition tasks," *Vision Research*, vol. 41, pp. 2261–2271, 8 2001.

[84] D. Y. Tsao, W. A. Freiwald, R. B. Tootell, and M. S. Livingstone, "A cortical region consisting entirely of face-selective cells," *Science*, vol. 311, pp. 670–674, 2 2006.

[85] W. A. Freiwald, D. Y. Tsao, and M. S. Livingstone, "A face feature space in the macaque temporal lobe," *NATURE NEUROSCIENCE VOLUME*, vol. 12, 2009.

[86] L. Chang and D. Y. Tsao, "The code for facial identity in the primate brain," *Cell*, vol. 169, pp. 1013–1028.e14, 6 2017.

[87] A. K. Engel, P. Fries, and W. Singer, "Dynamic predictions: Oscillations and synchrony in top–down processing," *Nature Reviews Neuroscience*, vol. 2, pp. 704–716, 10 2001.

[88] G. Michalareas, J. Vezoli, S. van Pelt, J. M. Schoffelen, H. Kennedy, and P. Fries, "Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas," *Neuron*, vol. 89, pp. 384–397, 2016.

[89] A. M. Bastos, M. Lundqvist, A. S. Waite, N. Kopell, and E. K. Miller, "Layer and rhythm specificity for predictive routing," *Proceedings of the National Academy of Sciences*, vol. 117, pp. 31459–31469, 12 2020.

Appendix A

Non-Significant Effects in Mass-Univariate ANOVA

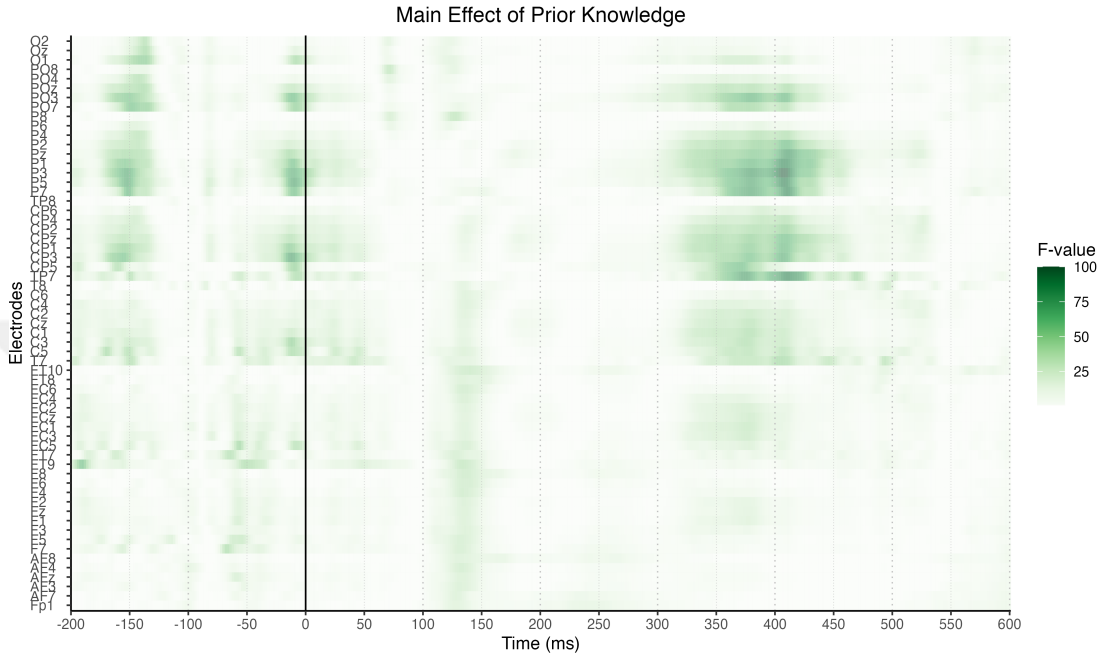

Figure A.1: Prior Knowledge

Figure A.2: Prior Knowledge x Agent

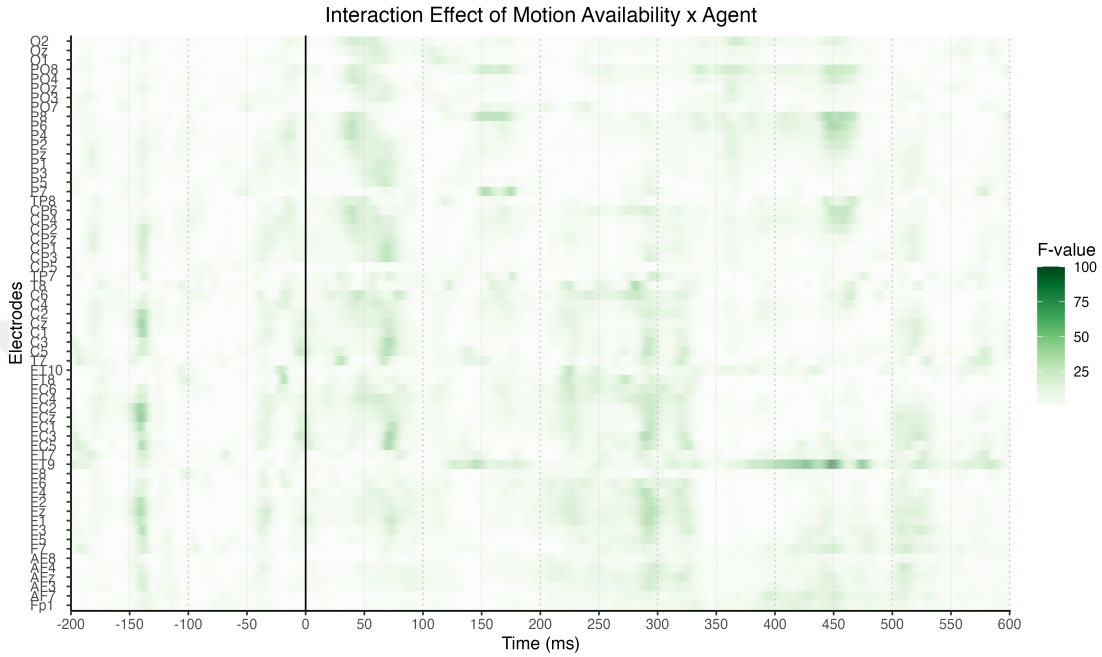

Figure A.3: Motion Availability x Agent

Figure A.4: Prior Knowledge x Motion Availability x Agent

Appendix B

Permissions

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 22, 2023

This Agreement between Sena Er Elmas ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5614111497364
License date	Aug 22, 2023
Licensed Content Publisher	Elsevier
Licensed Content Publication	NeuroImage
Licensed Content Title	ALE meta-analysis of action observation and imitation in the human brain
Licensed Content Author	Svenja Caspers,Karl Zilles,Angela R. Laird,Simon B. Eickhoff
Licensed Content Date	Apr 15, 2010
Licensed Content Volume	50
Licensed Content Issue	3
Licensed Content Pages	20
Start Page	1148
End Page	1167
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	The Interplay of Prior Information and Motion Cues in Resolving Visual Ambiguity in Agent Perception
Institution name	Bilkent University
Expected presentation date	Aug 2023
Portions	Figure 1
Requestor Location	Sena Er Elmas Basinevleri 15. Cadde 44/8 Kecioren Ankara, 06010 Turkey Attn: Sena Er Elmas
Publisher Tax ID	GB 494 6272 12
Customer VAT ID	TR21797281298
Total	**0.00 USD**
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment

terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your RightsLink account and that are available at any time at https://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given. The material may not be reproduced or used in any other way, including use in combination with an artificial intelligence tool (including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool), or to create any derivative work and/or service (including resulting from the use of artificial intelligence tools).

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier's permissions helpdesk here). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from

an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:

- immediately
    - via their non-commercial person homepage or blog
    - by updating a preprint in arXiv or RePEc with the accepted manuscript
    - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
    - directly by providing copies to their students or to research collaborators for their personal use
    - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
    - via non-commercial hosting platforms such as their institutional repository
    - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions
You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.
Terms & Conditions applicable to all Open Access articles published with Elsevier:
Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.
The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.
If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.
Additional Terms & Conditions applicable to each Creative Commons user license:
CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.
CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.
CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.
Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.10

Questions? E-mail us at customercare@copyright.com.