A Survey of Social Attention and Related Cognitive Processes

www.surveyx.cn

Abstract

This survey paper explores the interconnected psychological and cognitive processes of social attention, joint attention, gaze cuing, attention mechanisms, social cognition, and visual attention, highlighting their critical roles in shaping social interactions and communication. Social attention is pivotal in interpreting social cues, essential for effective communication, and is explored in various contexts, including educational, therapeutic, and technological domains. Joint attention, a shared focus between individuals, underpins social cognitive development and is crucial in contexts such as autism therapy and human-robot interaction. The gaze cuing effect demonstrates how gaze direction can reflexively shift attention, influencing social cognition and interaction quality. Attention mechanisms, particularly in AI models, enhance cognitive processing by optimizing focus and resource allocation. Visual attention, guided by saliency, plays a key role in processing social cues, facilitating effective communication in complex environments. The survey underscores the importance of these processes in understanding social cognition, with implications for education, therapy, and technology. Future research directions include exploring cultural and emotional influences on social cognition, refining attention mechanisms in AI, and developing interventions for social cognitive deficits. The integration of these insights into cognitive architectures and AI models holds promise for advancing our understanding of human cognition and interaction, emphasizing the need for continued research and innovation in this field.

1 Introduction

1.1 Relevance of Social Attention in Cognitive Processes

Social attention is crucial for understanding cognitive processes, acting as a mechanism for interpreting essential social cues that facilitate effective communication and interaction. Its importance is particularly evident in educational contexts, where the correlation between teacher gaze and student-rated interpersonal behavior highlights gaze's role in enhancing instructional dynamics [1]. Additionally, the interaction of eye contact with objects to direct attentional focus underscores social attention's significance in navigating complex environments [2].

In face-to-face interactions, gaze serves as a foundation for social behavior, bridging gaps in literature regarding gaze dynamics in communication [3]. Cultural variations profoundly influence face perception, emphasizing the role of social context in interpreting facial cues [4]. Within cognitive and psychological frameworks, the precision of image segmentation methods is vital for accurate diagnosis and treatment planning; integrating social attention mechanisms can enhance their efficacy and improve clinical outcomes [5]. Furthermore, gender disparities in social attention reveal biases in attention allocation, affecting the recognition and evaluation of researchers in academic settings [6].

Social attention also significantly impacts asynchronous video job interviews, shaping candidate evaluations through the assessment of non-verbal cues and interpersonal dynamics [7]. In robotics,

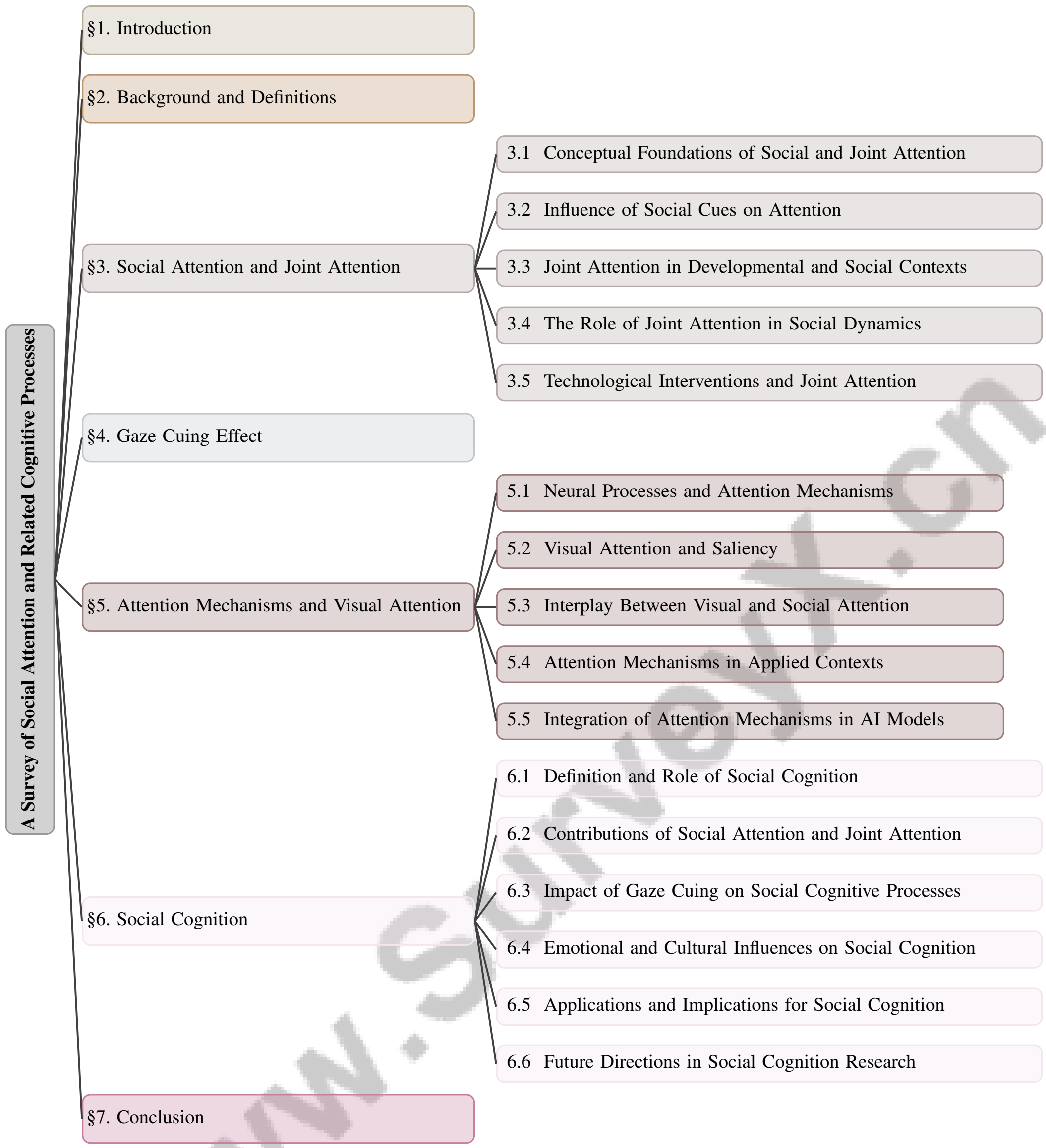


Figure 1: chapter structure

developing socially intelligent interactions requires computational models that incorporate social attention to manage interactions effectively [8]. The integration of social attention into cognitive architectures is essential for enabling autonomous learning agents to incrementally learn from experiences while flexibly relating sensory information to goals [9]. In autonomous driving, attention mechanisms are critical for understanding complex interactions with uncertain human drivers, facilitating effective behavioral planning [10].

Social attention is integral to cognitive processes, influencing decision-making, behavior control, and social interactions. Its implications span various domains, including education and the development of autonomous systems, highlighting the necessity for interdisciplinary collaboration to advance understanding in multiple fields. This evolution of concepts like joint attention reflects a shift from singular domain focus to broader interdisciplinary research, offering valuable insights for policymakers and educators aiming to enhance collaborative learning and knowledge sharing [6, 11, 9, 12, 13].

1.2 Interconnectedness of Key Terms and Concepts

The interrelation among social attention, joint attention, gaze cuing effect, attention mechanisms, social cognition, and visual attention forms a complex network underpinning social interactions and communication. Social attention facilitates the interpretation of social cues necessary for joint attention, where two individuals share focus on an object or event, which is vital for effective communication and often initiated by gaze cues that direct attention and influence social dynamics [11].

Gaze cuing exemplifies the interaction between visual and social attention, where an individual's gaze direction can guide another's focus. This phenomenon underscores the role of visual attention in processing social information by enabling selective focus on relevant stimuli while filtering distractions, essential for navigating social environments [6].

Attention mechanisms, governing the direction and maintenance of focus, are integral to both social and visual attention. They dynamically adjust based on social context, enhancing responsiveness to changing cues. This adaptability is crucial for social cognition, encompassing the mental processes that facilitate social interactions, such as perceiving, interpreting, and predicting others' behaviors. Research indicates that individuals can modify their attentional focus according to the social context of these cues, emphasizing the importance of accurately interpreting social signals to improve interaction and communication [14, 15].

The interconnectedness of these concepts is further illustrated in interdisciplinary research, where the evolution of keywords related to joint attention reflects the convergence of various cognitive processes in understanding social behavior [11]. Integrating these elements into a cohesive framework allows for a comprehensive understanding of how individuals process and respond to social information, highlighting each component's significance in facilitating effective social interactions.

1.3 Objectives and Structure of the Survey

This survey aims to elucidate the multifaceted nature of social attention and its related cognitive processes, guiding researchers in effectively employing attention mechanisms within their models. By categorizing existing techniques, the survey provides a structured framework for integrating social attention concepts across diverse research domains [16].

Organized into key sections, the survey focuses on distinct aspects of social attention and its cognitive implications. The introduction establishes the relevance of social attention in cognitive processes, emphasizing its interconnectedness with concepts like joint attention, gaze cuing effect, and visual attention. The background and definitions section offers comprehensive definitions and explanations of these core concepts, underscoring their significance in cognitive research.

Subsequent sections explore specific themes: the roles of social and joint attention in social interactions and cognitive development; an analysis of the gaze cuing effect, including theoretical perspectives and experimental findings; and a discussion on attention mechanisms and visual attention, particularly their neural underpinnings and applications in processing social cues. The survey also addresses social cognition, detailing its role in understanding interactions and its connections to social attention, joint attention, and gaze cuing.

Each section is meticulously crafted to build upon the previous, offering a coherent narrative that integrates theoretical insights with empirical evidence. This structure enhances the understanding of social attention's impact on cognitive processes and provides practical guidance for researchers seeking to incorporate these insights into their work.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Definitions and Core Concepts

Social attention is integral to interpreting social cues like gaze, head orientation, and gestures, crucial for spatial attention and effective communication [1]. It facilitates level-1 visual perspective-taking, enabling individuals to infer others' visual perspectives from gaze direction [3]. In dynamic contexts,

AI-generated, for reference only.

mutual gaze and gaze following are vital for learning [17], influencing interpersonal dynamics as seen in teacher-student interactions [1].

Joint attention, the shared focus on an object or event, is fundamental for social communication and cognitive development, involving initiation and response to joint attention [5]. This process is supported by brain regions related to perspective-taking and mentalizing, highlighting its role in cognitive and social development [18].

The gaze cuing effect, where gaze direction influences attention, is shaped by social hierarchies and power dynamics [4]. Understanding this effect is crucial for decoding social interactions and responses to social cues.

Attention mechanisms, neural processes directing and maintaining focus, adapt attention based on social contexts, predicting social behaviors and aiding computer vision tasks like soft, hard, and multi-modal attention [7]. Comparisons between human attention and Transformer architectures reveal implications for artificial intelligence [9].

Social cognition involves mental operations for perceiving, interpreting, and predicting behaviors, linked closely to social and joint attention, providing cues for understanding social dynamics. Studies on observation effects and autism underscore its significance [8].

Visual attention, the selective focus on visual stimuli amidst distractions, is crucial for processing social cues in complex environments. Integrating visual with social attention is essential for effective social navigation [18], explored in Collectionless AI prioritizing real-time interactions [9].

These core concepts underpin cognitive processes in social contexts, offering insights into attention and behavior mechanisms, important across research domains like psychotherapy and autonomous systems requiring sophisticated attention models [6].

2.2 Significance in Cognitive Processes

Social attention and related cognitive processes are crucial for understanding human cognition in complex social environments. They facilitate the interpretation of social cues, influencing spatial attention and visual information prioritization, essential for modeling cognitive processes underlying social interactions and crowd behavior predictions [19]. Their integration into deep learning models highlights their relevance in visual tasks crucial for understanding human cognition [16].

Joint attention is vital in human-robot interactions, where aligning attention with humans is key for communication, emphasizing its importance in cognitive systems [20]. Biological factors like oxytocin levels affect visual attention development in children exposed to maltreatment, illustrating biological and environmental interactions in attention processes [21].

Interpersonal coordination, a core aspect of social cognition, impacts psychotherapeutic outcomes. Gaze perception's role in social cognition is significant in clinical settings like autism, where gaze processing abnormalities affect social functioning [22]. Integrating human attention mechanisms into AI models for tasks like image recognition and language understanding remains a research challenge [23].

The distinction between social and nonsocial attention is debated, with studies suggesting unique temporal stability and cognitive characteristics for social attention, though inconsistencies require further investigation [24]. The parallels and distinctions between human attention and Transformer models present significant challenges [12].

Integrating attention models into cognitive architectures, supported by datasets of top-down attention maps, provides a robust framework for simulating human cognitive processes. The cognidynamics framework enhances understanding by framing learning as an optimization problem mirroring natural cognitive processes, offering insights into attention and cognition mechanisms [25].

In recent years, the study of social and joint attention has gained prominence due to its critical role in various fields, including communication, cognitive development, and human-robot interaction. Understanding the intricate dynamics of these concepts requires a comprehensive examination of their underlying structures and contextual influences. As depicted in Figure 2, this figure illustrates the hierarchical structure of social and joint attention, categorizing its conceptual foundations, the influence of social cues, and the relevant developmental and social contexts. Moreover, it emphasizes the role of social and joint attention in social dynamics and highlights the impact of technological interventions. The diagram serves to underscore the integral function of social cues and technological advancements in enhancing joint attention, thereby reinforcing the significance of these elements in educational and therapeutic settings. This visualization not only aids in clarifying the complex interrelations but also enhances our understanding of how these processes operate within varied contexts.

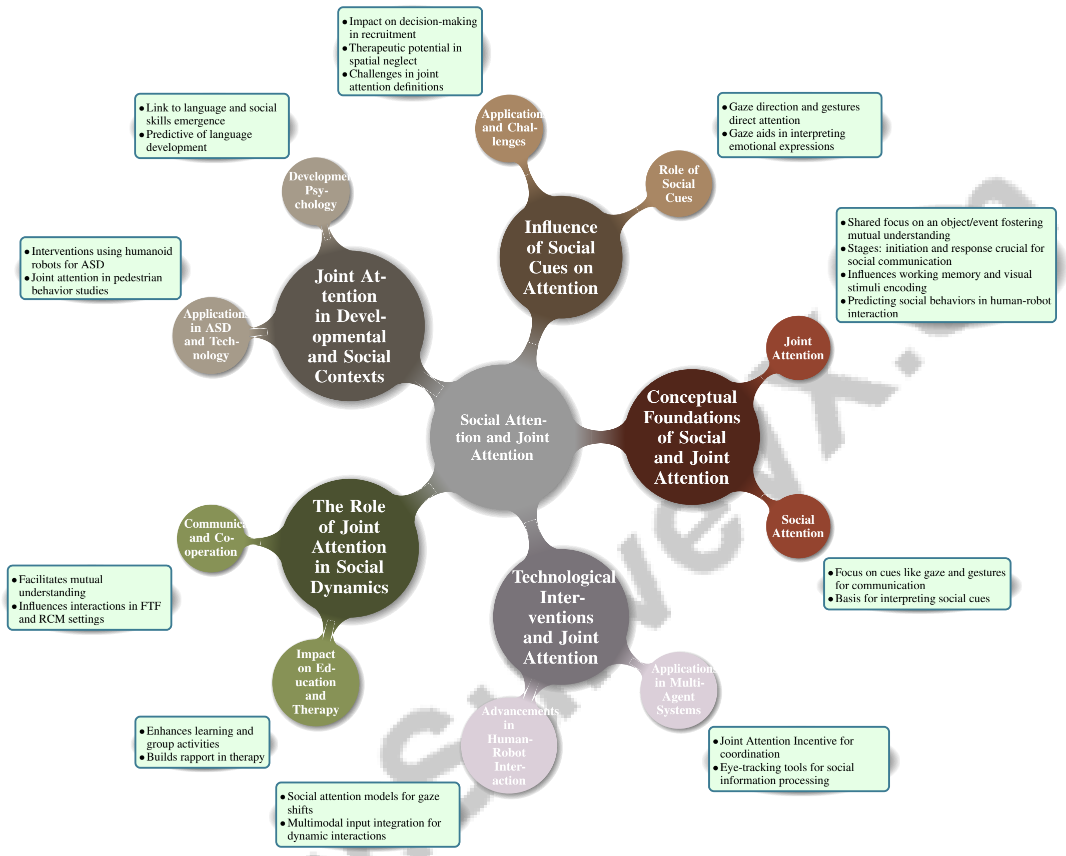


Figure 2: This figure illustrates the hierarchical structure of social and joint attention, categorizing its conceptual foundations, influence of social cues, developmental and social contexts, role in social dynamics, and technological interventions. The diagram highlights the integral role of social and joint attention in communication, cognitive development, human-robot interaction, and educational and therapeutic settings, emphasizing the significance of social cues and technological advancements in enhancing joint attention.

3 Social Attention and Joint Attention

3.1 Conceptual Foundations of Social and Joint Attention

Social and joint attention are integral to cognitive and social development, forming the basis for interpreting social cues. Social attention involves focusing on cues like gaze and gestures, essential for communication [8]. Joint attention, a subset of social attention, entails a shared focus on an object or event, fostering mutual understanding and cooperation [26]. This shared focus aligns cognitive experiences, enhancing interpersonal attunement.

Joint attention development involves key stages, including initiation and response, vital for social communication and cognitive growth [27]. These stages correspond to neural and behavioral correlates, emphasizing joint attention's complexity as a socio-cognitive process. In educational contexts, joint attention influences working memory by affecting how gaze cues impact visual stimuli encoding [28].

In human-robot interaction, predicting social behaviors through joint attention is crucial for systems navigating complex social environments. The Multiagent Group Perception and Interaction (MGPI) network exemplifies this by predicting social actions in group conversations, highlighting social attention's significance [8]. Challenges in estimating joint attention among individuals, due to independent gaze attribute treatment, underscore joint attention dynamics' complexity [26].

Theoretical perspectives suggest joint attention as a foundational social cognitive skill linked with gesture [27]. This interplay of social interactions and cognitive processes in joint attention provides a framework for understanding focus alignment in social situations, crucial for effective communication. In classrooms, teacher-student engagement enhances learning outcomes. Technological advancements, like gaze tracking, offer insights into intentions during interactions, improving human-computer interactions, educational assessments, and therapeutic approaches for attention disorders [29, 5, 30, 31].

3.2 Influence of Social Cues on Attention

Social cues, such as gaze direction and gestures, are fundamental in directing attention and facilitating interactions. These cues serve as sensory inputs, aligning focus and enhancing communicative efficacy [32]. Direct gaze is particularly effective in capturing attention, influencing object allocation and cognitive processes [2]. Gaze aids in interpreting emotional expressions and social signals [3].

Gaze cue following reflects social dynamics and hierarchies [33]. Gaze can direct attention, influencing decision-making in contexts like recruitment [7]. In therapy, gaze direction aids patients with spatial neglect, demonstrating gaze cues' therapeutic potential [34].

In human-robot interactions, social cues impact interaction quality and decision-making. Mutual eye contact frequency during joint attention tasks modulates neural mechanisms, emphasizing gaze's role in coordinating behaviors [26]. Cultural expectations influence gaze's communicative function, shaping interactions [4].

Social cues integration into attention mechanisms is evident in applications like Ego-Attention architecture in autonomous driving [10]. Perspective-taking influences working memory when observers perceive cues can see items, highlighting social cues' cognitive impact [28].

Lack of standardized joint attention definitions leads to inconsistent findings, necessitating clearer understanding of social cues' influence. This variability is relevant in developmental contexts, where gaze fixation patterns link to social-emotional outcomes [33]. Gesture and joint attention may reflect a shared social cognitive skill, crucial for understanding interactions [27].

Social cues like gaze, head direction, and pointing enhance interactions across contexts. Individuals rely on these cues to prioritize attention, identifying locations for social or informational benefits. The gaze cueing effect shows observing others' attention can reflexively shift focus, improving perceptual performance. Contextual interpretations influence social cues' effectiveness, highlighting understanding social cues' interaction with attentional mechanisms for effective communication [14, 35].

3.3 Joint Attention in Developmental and Social Contexts

Joint attention is foundational in developmental and social contexts, underpinning communication and cognitive growth. In developmental psychology, it's linked to language and social skills emergence. Infants' gesture use correlates with joint attention capacity, predictive of language development by 24 months [27]. Joint attention's role in early childhood development facilitates communication skills and social understanding.

As illustrated in Figure 3, the hierarchical structure of joint attention spans various contexts, including developmental psychology, Autism Spectrum Disorder (ASD), and technological and educational domains. This figure highlights the significant aspects and interventions relevant to each area, providing a visual representation of how joint attention operates across different fields.

In ASD, joint attention impairment affects engagement with socially relevant stimuli, hindering socio-communicative development. Interventions using humanoid robots, like NAO, enhance joint attention in children with ASD by providing structured interactions that encourage engagement [36].

These robotic interventions complement traditional therapies, offering new avenues for improving joint attention skills [37].

Joint attention's importance extends to social and technological domains. In pedestrian behavior studies, joint attention ensures safe interactions, facilitating mutual awareness in dynamic environments [38]. In human-computer interaction and robotics, it's essential for systems interpreting and responding to social cues, enhancing interaction quality [11].

Technological advancements have led to innovative joint attention modeling methods. The Position-embedded Joint Attention Transformer (PJAT) models interactions among attributes, generating accurate joint attention heatmaps [26]. Advanced computational models enhance joint attention understanding and application across contexts.

In education, joint attention is pivotal in collaborative learning activities, where shared focus enhances interaction and outcomes. Activities promoting joint attention in classes improve engagement and achievement [13].

Joint attention is critical across contexts, influencing cognitive development, communication, and interaction quality. Multidisciplinary research seeks to deepen understanding and improve joint attention through innovative interventions and methodologies, exploring its evolution across fields, collaborative discourse dynamics, and technological advancements in detecting joint visual attention [13, 11, 5, 6].

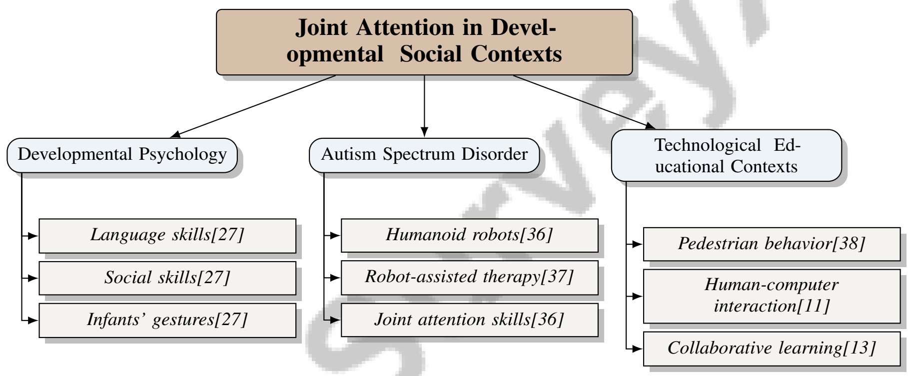


Figure 3: This figure illustrates the hierarchical structure of joint attention across various contexts, including developmental psychology, autism spectrum disorder, and technological and educational domains. It highlights the significant aspects and interventions relevant to each area.

3.4 The Role of Joint Attention in Social Dynamics

Joint attention is pivotal in shaping social dynamics and enhancing communication, serving as a foundational mechanism for synchronizing focus. This shared attentional framework facilitates mutual understanding and cooperation, essential for social exchanges. In face-to-face (FTF) and remote communication (RCM) settings, joint attention significantly influences interactions, particularly in job assessment interviews, where joint attention engagement impacts candidate assessment [39].

In FTF interactions, joint attention is established through direct gaze and shared focus, aligning attention and intentions, fostering communication. In contrast, RCM settings pose joint attention challenges due to the absence of physical presence and non-verbal cues. Technology advancements, like video conferencing, bridge this gap by providing cues that facilitate joint attention and enhance interaction quality [39].

Joint attention's impact on social dynamics extends to education, where it's integral to collaborative learning and group activities. By promoting shared focus, joint attention enhances engagement and fosters supportive environments. In therapy, it builds rapport and enhances communication, especially for individuals with social communication challenges. Joint attention's foundational importance in social cognition facilitates perspective sharing and coordination, contributing to therapeutic outcomes. Research indicates joint attention, developing in infancy, engages a neural network supporting interactions, relevant in addressing autism spectrum disorder impairments [5, 30, 40, 41, 42].

In human-robot interaction, joint attention enables robots to interpret and respond to social cues effectively. This capability is crucial for developing socially intelligent systems navigating intricate environments and engaging intuitively. Frameworks like Joint Attention-Interaction-Creation (AIC) capture collaborative discourse dynamics, enhancing understanding of social cues and interactions. Incorporating insights on gaze following and social competence allows systems to better interpret and respond to behaviors, improving communication and collaboration [13, 43, 11, 44]. Robots engaging in joint attention enhance their utility across applications, from assistive technologies to educational tools.

Joint attention shapes social dynamics by facilitating communication and enhancing interaction quality across contexts. This phenomenon, characterized by focusing on a common target, is essential for understanding social cues, intentions, and goals. Research indicates joint attention enriches classroom learning and has implications in human-computer interaction, educational assessment, and attention disorder treatment. Technological advancements, like algorithms for detecting joint attention, underscore its importance in analyzing behaviors and improving collaborative discourse, highlighting its multifaceted impact on communication [13, 6, 5, 31]. Its role in aligning focus and facilitating understanding underscores its importance in natural and artificial systems, highlighting the need for continued research and innovation.

3.5 Technological Interventions and Joint Attention

Technological advancements have significantly influenced joint attention understanding and enhancement through tools and models facilitating human-computer and human-robot interactions. Social attention models trained on eye-tracking data enable robots to perform gaze shifts mimicking human behavior, enhancing joint attention engagement [45]. This capability is crucial for developing robots that effectively participate in social interactions, interpreting and responding to social cues naturally.

Incorporating deictic gestures and visual attention mechanisms into robotic systems facilitates dynamic interactions, enabling meaningful joint attention tasks [20]. These interactions are enhanced by multimodal input integration, creating saliency maps guiding a robot's attention toward relevant stimuli, improving focus on critical elements [46].

3D gaze tracking frameworks play a pivotal role in enhancing joint attention, particularly in collaborative scenarios. By accurately tracking gaze patterns, these frameworks enable nuanced social dynamics understanding and coordinated efforts among team members [47]. Technological interventions improve interaction quality in educational and professional settings, where joint attention is crucial for collaboration.

In multi-agent systems, Joint Attention Incentive (JAI) encourages agents to align attention, enhancing coordination and social learning [48]. This approach highlights technology's potential to foster joint attention in human-robot interactions and among artificial agents, showcasing joint attention models' versatility across domains.

Tools like eye-tracking enhance joint attention understanding [49]. These tools provide insights into social information processing in group settings, challenging traditional perspectives and offering a deeper understanding of joint attention mechanisms [33].

Integrating human attention modeling into AI applications improves performance across tasks like image recognition and natural language processing, illustrating technological interventions' impact on joint attention [23]. As research advances, technological innovations promise to enhance joint attention capabilities in natural and artificial systems, leading to more effective interactions.

4 Gaze Cuing Effect

Exploring the gaze cuing effect requires an examination of theoretical models that elucidate its mechanisms, enhancing our understanding of gaze as a pivotal social cue. These frameworks provide a foundation for empirical findings that corroborate these theoretical constructs. The following subsections will detail the theoretical approaches and models informing our comprehension of gaze cuing, setting the stage for experimental insights.

4.1 Theoretical Perspectives and Models

The gaze cuing effect, a reflexive attentional shift triggered by gaze direction, is central to social attention studies. Theoretical models emphasize gaze's dual role as an attentional and communicative signal, vital for guiding interactions and interpreting social information. The Interpersonal Gaze Processing model integrates these functions, illustrating how gaze dynamics direct attention and convey social signals in face-to-face interactions [33], with direct gaze capturing more attention due to its rich information content [50].

As illustrated in Figure 4, which highlights the theoretical perspectives on the gaze cuing effect, gaze serves not only as an attentional signal but also finds application in human-robot interaction while being influenced by contextual interpretation. Gaze following is influenced by social context, such as group size, modulating gaze cuing effects [33]. The multi-level meta-analytic framework categorizes research on gaze cuing, emphasizing task and cue features' importance in gaze's effectiveness as a directional signal [51]. This framework highlights gaze cuing's variability across contexts, necessitating consideration of task-specific factors.

Incorporating gaze into computational models, particularly in human-robot interactions, demonstrates gaze cuing's application in artificial systems. A dataset for real-time gaze detection in human-robot interactions exemplifies this, focusing on settings where gaze detection is critical for communication [52]. The gaze cuing effect's reliability has been tested across multiple measurements, emphasizing its consistency as a social cue [53].

Gaze interacts with other social cues, like arrows, activating shared and distinct attentional mechanisms based on social significance. This is crucial in contexts like asynchronous interviews, where hierarchical attention models leverage gaze as a communicative cue [28]. Gaze cues influence perspective-taking and working memory performance, engaging higher-level cognitive processes [28].

Theoretical perspectives on the gaze cuing effect (GCE) offer a nuanced framework for understanding how gaze cues shape social cognition. The influence of gaze shifts on attention is contingent upon contextual interpretations, as demonstrated by research showing individuals may suppress reflexive responses when interpreting gaze cues as indicative of internal focus rather than external attention [14, 51]. These models emphasize gaze's multifaceted nature as a social signal, guiding attentional processes across contexts.

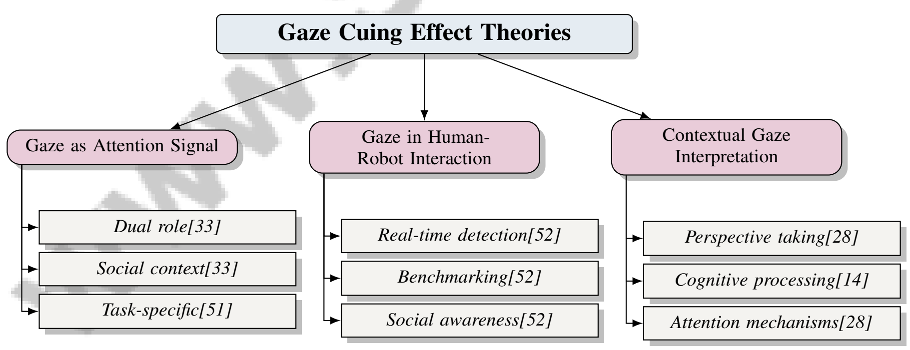


Figure 4: This figure illustrates the theoretical perspectives on the gaze cuing effect, highlighting its role as an attentional signal, its application in human-robot interaction, and the influence of contextual interpretation on gaze perception.

4.2 Experimental Findings on Gaze Cuing

Experimental investigations into the gaze cuing effect (GCE) provide insights into factors influencing attentional shifts elicited by gaze direction. Research indicates gaze cues are more resistant to suppression than arrow cues in counter-predictive scenarios, underscoring gaze's unique role in attentional processes [54]. This finding aligns with studies on GCE under various conditions, such as synchronous and asynchronous presentations of gaze and arrow cues, demonstrating gaze cuing's context-dependent nature [55].

Surveys of gaze-cueing studies in adults reveal task types, cue features, and emotional expressions significantly influence gaze cues' efficacy in directing attention [51]. Joint attention tasks, like categorizing color images, show how gaze cuing facilitates attention orientation in structured settings [56]. Free-looking tasks reveal the potency of gaze cues in guiding attention [57].

Gaze cues have therapeutic potential, aiding patients with unilateral spatial neglect (USN) by enhancing attention orientation, illustrating gaze cuing's clinical applications [58]. Experiments show working memory performance is enhanced for validly cued items when the observer perceives the cue can see the items [28].

In group dynamics, a single individual's gaze can direct attention in smaller groups, while larger groups require more individuals, indicating social context's influence on gaze cuing [33]. Gaze cuing's reliability across measurements reinforces its consistency as a social cue [53]. Gaze detection in human-robot interactions achieves significant accuracy, demonstrating potential for incorporating gaze cues into artificial systems [52].

These experimental findings underscore the intricate relationship between task demands, social contexts, and attentional mechanisms influencing the gaze cuing effect. The GCE is not merely a reflexive response but is modulated by task cognitive demands and interpretive context. For instance, the GCE diminishes when gaze shifts are perceived as cognitive processing rather than overt attention shifts. Task type, such as target localization versus discrimination, affects how emotional expressions enhance the GCE, illustrating cognitive resources' role in integrating social cues and attentional orienting. These insights reveal social attention's flexibility and context-dependence, highlighting the need for a nuanced understanding of gaze cues in various frameworks [2, 14, 59, 60, 51].

4.3 Neural and Psychological Mechanisms

The gaze cuing effect (GCE) is supported by intricate neural and psychological mechanisms facilitating automatic attention orientation in response to gaze direction. Neuroimaging studies, using techniques like fMRI, identify key brain regions involved in gaze processing, notably those associated with attention control and social cognition [55]. Emotional congruency modulates these neural responses, with larger late positive potential (LPP) amplitudes observed when cue and target are emotionally congruent, intensifying the gaze-cuing effect [61].

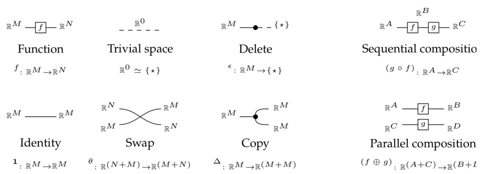
The neuroendocrine system, particularly oxytocin, underscores social attention's biological foundations, affecting gaze cue processing [24]. Mechanisms like the AttentionRNN improve performance and attention mask quality across tasks, refining human gaze perception representation [62].

Psychologically, the gaze cuing effect is shaped by individual differences and social contexts. Social exclusion diminishes the gaze-cueing effect for social cues, illustrating the interplay between exclusion and attentional processes [63]. Social cues maintain their cuing effects over time, contrasting with nonsocial cues' temporal decay, highlighting social cues' robustness in sustaining attention [24]. Comparative analyses reveal variability in gaze and arrow cues' efficacy among patients, emphasizing personalized approaches in clinical settings [58].

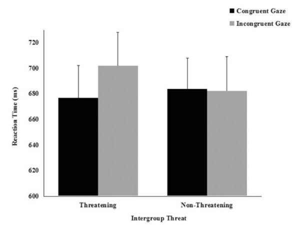
In collaborative augmented reality (AR) environments, gaze-related errors impact user performance, underscoring precise gaze cue processing's importance for enhancing interactions [64]. Real-time gaze detection's effectiveness is validated through pipelines achieving significant accuracy in identifying gaze target objects, demonstrating potential for integrating gaze cues into artificial systems [52].

The innovation of multi-head social attention mechanisms in autonomous driving models enables systems to focus on relevant vehicles, crucial for decision-making in complex scenarios [10]. This highlights gaze-related attention mechanisms' applicability beyond human interactions, extending into autonomous systems.

Overall, the neural and psychological mechanisms underlying the gaze cuing effect provide a comprehensive framework for understanding gaze cues' influence on attention and social cognition. These insights are crucial for advancing theoretical research and practical applications across interdisciplinary fields, including human-computer interaction, where understanding user attention dynamics can enhance interface design; autonomous systems, benefiting from improved decision-making processes informed by attention mechanisms; and social robotics, particularly in therapeutic contexts, where advancements in joint attention interventions can significantly improve outcomes for individuals with autism spectrum disorder [23, 37, 11, 12, 13].



(a) Function Composition in a Graphical Notation[65]



(b) The image shows a bar graph comparing reaction times for two groups of participants, categorized by whether they were looking at a threatening or non-threatening intergroup threat.[66]

Figure 5: Examples of Neural and Psychological Mechanisms

As shown in Figure 5, the "Gaze Cuing Effect; Neural and Psychological Mechanisms" is illustrated through two visual representations that delve into the cognitive and neural processes underlying attention and perception. The first image, "Function Composition in a Graphical Notation," provides a mathematical framework for understanding how functions are composed and manipulated, featuring symbols and diagrams depicting operations on functions. This mathematical visualization sets the stage for analyzing complex cognitive processes. The second image presents a bar graph comparing reaction times between two groups of participants exposed to either threatening or non-threatening intergroup threats. The graph categorizes reactions under "Congruent Gaze" and "Incongruent Gaze" conditions, illustrating how gaze direction influences attention and reaction time in the presence of perceived threats. Together, these images provide a comprehensive overview of the neural and psychological mechanisms involved in gaze cuing, highlighting both abstract, theoretical underpinnings and tangible, observable effects on human behavior [65, 66].

5 Attention Mechanisms and Visual Attention

Category	Feature	Method
Visual Attention and Saliency	Semantic Relevance Centralized Focus	RJAM[18] EA[10]
Interplay Between Visual and Social Attention	Coordinated Focus Mechanisms Sensory Information Integration	JAL[67], HN[7], SCTFA[68] GASP[46]
Attention Mechanisms in Applied Contexts	Sequential and Temporal Focus Collaborative Interaction Enhancement	ACE[54] AIC[13]
Integration of Attention Mechanisms in AI Models	Gaze and Interaction Strategies Human Attention Modeling	PGCS[69] ARNN[62]

Table 1: This table provides a comprehensive summary of various attention mechanisms and their applications across different domains. It categorizes methods based on their focus on visual attention, social attention interplay, applied contexts, and integration in AI models, highlighting key features and the methodologies employed. The table serves as a reference for understanding the diverse approaches and innovations in attention research.

Understanding attention mechanisms is crucial for exploring their role in cognitive processes, particularly visual attention. Table 1 presents a detailed classification of attention mechanisms, illustrating their roles and methodologies across visual, social, and applied contexts, as well as their integration into AI models. Additionally, Table 4 presents a comprehensive comparison of different attention mechanisms, illustrating their spatial and temporal dynamics and approaches to multimodal integration within neural processes, visual attention, and social contexts. The following subsection examines the neural processes that underpin these mechanisms, elucidating how they facilitate focus amidst competing stimuli. This exploration provides a foundational understanding of the interplay between neural activity and attentional control, setting the stage for a comprehensive examination of visual attention mechanisms and their implications in diverse contexts.

Method Name	Spatial Dynamics	Temporal Dynamics	Multimodal Integration
ARNN[62]	Structured Spatial Attention	Sequential Dependencies	-
EA[10]	Spatial Grid Representation	Sequence ON Attention	Multimodal Integration
SCTFA[68]	Spatial-channel Information	Temporal Dimensions	Audio And Visual

Table 2: Comparison of various neural process models highlighting their spatial dynamics, temporal dynamics, and multimodal integration capabilities. The table provides an overview of the methodologies employed by different models, emphasizing their unique approaches to managing attention in complex environments.

5.1 Neural Processes and Attention Mechanisms

Neural processes are fundamental to directing and maintaining attention, ensuring optimal cognitive resource allocation amidst diverse stimuli. These processes are crucial for understanding how attention is dynamically managed across contexts, including social and artificial systems. Models like AttentionRNN highlight the importance of spatial dynamics in attention processes through structured spatial attention [62]. In joint attention, saliency information guides attention mechanisms, enhancing joint attention capture [70]. The Social Attention model, using recurrent neural networks, predicts trajectories by modeling spatial and temporal dynamics [10]. The SCTFA module further enhances spiking neural networks by integrating spatial-channel and temporal information [68].

Table 2 presents a comparative analysis of three neural process models, focusing on their strategies for handling spatial and temporal dynamics, as well as their approaches to multimodal integration. Interactions between gaze and visual cues, such as arrows, reveal complex attentional shifts. Arrows' spatial interference effects depend on temporal separation, while gaze induces a reversed interference effect modulated by visual context [55]. The Ego-Attention method in autonomous driving emphasizes attention's role in decision-making [10]. Therapeutic and interpersonal methods like the Posner-like paradigm use gaze and arrow cuing to measure attentional shifts, demonstrating these cues' robustness [53]. The Recursive Joint Attention Model (RJAM) exemplifies multimodal integration of neural processes, fusing audio and visual features [9].

These insights into neural processes and attention mechanisms enhance our understanding of attention direction and maintenance, offering implications for theoretical research and practical applications. The challenge of acquiring high-quality human gaze data remains critical for training models replicating human-like attention behavior, necessitating continued research and innovation [23].

5.2 Visual Attention and Saliency

Visual attention enables selective focus on relevant stimuli, facilitating efficient information processing in complex environments. This process is intricately linked to saliency, which refers to stimuli's distinctiveness that captures attention. Salient features guide visual attention by drawing focus to areas of high contrast, movement, or color, crucial for interpreting dynamic scenes [16].

The interaction between visual attention and saliency is evident in cognitive and perceptual tasks. In joint attention, saliency directs gaze and focus, enhancing shared attentional states [70]. This is essential in social interactions, where detecting and responding to salient social cues is vital for effective communication [18]. In computational models, saliency simulates human-like attention, allowing systems to prioritize information processing based on visual feature prominence. Integrating saliency into attention models improves visual recognition task performance [9]. Algorithms utilizing saliency maps enhance critical visual cue detection in autonomous systems [10].

The relevance of visual attention and saliency extends to practical applications, such as user interface design, where salient features enhance user experience. Understanding visual attention's operation with saliency enables designers to create intuitive visual environments aligning with human cognitive processes [23]. Exploring visual attention and saliency provides insights into perception and cognition mechanisms, critical for theoretical research and advancing technologies reliant on visual processing. These insights enhance understanding of joint visual attention in images, essential for interpreting social interactions and intentions. Integrating human attention modeling into deep learning frameworks improves image and video processing applications, highlighting the necessity for ongoing research to refine these models [16, 23, 5, 51].

5.3 Interplay Between Visual and Social Attention

The interaction between visual and social attention is crucial for processing social cues, significantly influencing cognitive and perceptual tasks. Visual attention, characterized by selective focus on pertinent stimuli, intersects with social attention, involving interpreting and responding to social signals like gaze and gestures. This interplay is essential for navigating social environments and engaging in meaningful communication, influencing collaborative discourse dynamics and enhancing interpersonal behaviors [1, 13, 11, 71].

In joint attention, semantically meaningful shared attention regions enhance shared representation learning, as demonstrated by the I-SEE method [67]. Visual saliency enhances social attention by guiding focus to relevant social cues, facilitating joint attention. In human-robot interactions, incorporating human gaze data into social attention models enables robots to mimic human behavior, facilitating natural interactions [46]. This practical application of visual and social attention interplay allows for systems engaging effectively with humans in social contexts.

Task demands modulate social attention mechanisms, contributing to understanding gaze cueing. The hierarchical attention model in video interviews incorporates visual cues to enhance candidate response analysis [7]. This highlights attention processes' dynamic nature and tasks' role in influencing social attention allocation. The SCTFA method shows how historical spatial-channel information improves spiking neural networks' feature representation, illustrating visual and social attention's sophisticated interplay in enhancing cognitive processing [68]. The AttentionRNN model captures structural dependencies among attention variables, leading to coherent and informative attention masks [62].

In collaborative augmented reality (AR) environments, accurate target identification using shared gaze cues is essential for effective task performance, as gaze processing errors impact interaction quality [64]. This underscores precise visual attention's importance in facilitating social interactions, particularly in complex collaborative settings. The interaction between visual and social attention enhances cognitive processing, improving social cue interpretation and interpersonal interaction quality. Research indicates individuals rely on gaze shifts as social-attentional strategies, guiding them toward locations likely to yield social rewards. However, gaze cues' effectiveness depends on context; when interpreted as cognitive engagement rather than overt attention, the gaze cueing effect diminishes. Gaze and body movement convergence during social interactions further informs observers' attentional focus, with eye-tracking studies revealing a tendency to prioritize gaze direction over body movement in ambiguous situations. This suggests social attention's flexibility depends on social context interpretation and multimodal cue integration, facilitating effective social engagement [14, 32]. Understanding this interplay is crucial for advancing theoretical research and developing technologies relying on visual and social attention mechanisms.

5.4 Attention Mechanisms in Applied Contexts

Method Name	Application Domains	Functional Benefits	Implementation Challenges
ARNN[62]	Image Classification	Improved Model Performance	Variable OR Noisy
ACE[54]	Human-machine Interactions	Improved Interaction Quality	Individual Responses Variability
AIC[13]	Education	Improved Interaction Quality	Computational Demands

Table 3: Overview of various attention mechanism methods, their application domains, functional benefits, and implementation challenges. The table highlights the practical utility and limitations of methods like ARNN, ACE, and AIC in enhancing model performance and interaction quality across different fields.

Attention mechanisms are increasingly integrated into real-world applications, enhancing social interactions, educational processes, and therapeutic interventions. In social contexts, attention models facilitate real-time decision-making by analyzing peer effects, allowing dynamic focus adjustment in response to social cues, improving interaction quality and adaptability [62].

In educational environments, attention mechanisms foster collaborative discourse and idea generation. The AIC framework highlights joint attention's significance in enhancing interaction quality and facilitating creative processes in group activities. Attention models enable educators to design collaborative environments promoting active engagement and knowledge sharing [6].

Therapeutic interventions, especially autism therapy, benefit from attention mechanisms. Robot-assisted interventions show greater effectiveness than traditional approaches in improving joint attention among children with autism, suggesting robots provide consistent, engaging interactions enhancing therapeutic outcomes, though further research is needed for long-term validation [52].

In multi-agent systems, attention mechanisms enhance coordination and learning efficiency. The Joint Attention Incentive (JAI) method guides agents to focus on relevant environmental elements, reducing exploration costs and promoting coordinated actions [53]. This highlights attention models' utility in optimizing agent-based systems' performance.

Challenges persist in applying attention mechanisms, particularly regarding computational demands and data requirements. Many models are computationally intensive and require large datasets, limiting real-time applicability [54]. Patient response heterogeneity to gaze and arrow cueing in rehabilitation complicates effective strategy identification, necessitating personalized treatment approaches.

Integrating attention mechanisms in real-world applications presents substantial opportunities to improve social interactions, educational practices, and therapeutic outcomes by enhancing collaborative discourse analysis, facilitating joint attention in social settings, and supporting cognitive processes. Frameworks like Joint Attention-Interaction-Creation (AIC) capture collaborative learning dynamics, while advanced joint attention estimation methods optimize human-computer interactions and educational assessments. Human attention modeling informs AI solutions across diverse fields, highlighting these mechanisms' transformative potential in enhancing engagement and learning outcomes [13, 23, 31]. Continued research and innovation are essential to address challenges and fully realize these models' practical utility across diverse domains.

Table 3 provides a comprehensive overview of different attention mechanism methods, illustrating their applications, benefits, and the challenges faced during implementation. As shown in **??**, attention mechanisms extend beyond theoretical constructs to practical applications across various domains. The first example, "Function and its transformations in a mathematical context," visually represents mathematical transformations using graphical elements such as functions, trivial spaces, and operations, highlighting attention's role in abstract mathematical reasoning. In the second example, "Binary Cross-Entropy Loss for Person Re-identification," attention mechanisms are integral to a neural network architecture designed for person re-identification. Here, the attention module selectively focuses on specific regions of an input image, enhancing the network's ability to accurately reconstruct the image through an encoder-decoder framework, optimizing performance using a binary cross-entropy loss function. The third example, "Similarity Scores for Students and Quotes," demonstrates attention's application in educational contexts by graphically representing similarity scores between students and quotes, facilitating an understanding of relational dynamics and preferences. Collectively, these examples underscore attention mechanisms' versatility and significance in enhancing computational models and interpreting complex data across varied fields [65, 72, 13].

5.5 Integration of Attention Mechanisms in AI Models

Integrating attention mechanisms into AI models has significantly advanced the field, offering new avenues for enhancing model performance and generalization. A critical aspect of this integration is using human-derived attention maps, shown to be more diagnostic for object recognition than traditional bottom-up saliency maps. This benchmark dataset enhances model training by providing more contextually rich information, improving AI systems' accuracy and robustness [73].

Recent developments have introduced string diagrams in attention mechanisms, allowing both formal detail and abstraction. This innovation enhances attention models' formal expressivity, enabling more complex representations of attention processes. Such advancements facilitate a deeper understanding of how attention can be modeled and applied within AI systems, contributing to more sophisticated models [65].

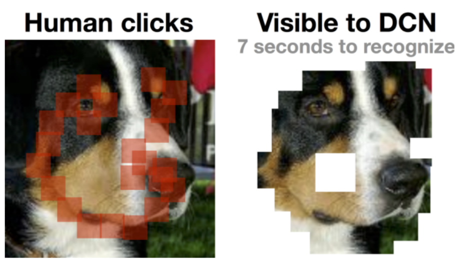
In gaze control, a Planning-based Gaze Control System (PGCS) generates gaze behavior by considering future interactions and conversation dynamics rather than merely reacting to them. This forward-looking approach aligns with human attention's dynamic nature, offering AI models the capability to anticipate and adapt to evolving interactions, improving interactive performance [69]. Technological advances facilitate new methodological approaches to study gaze behavior, crucial for developing AI models emulating human-like attention processes. Leveraging these methodologies,

AI-generated, for reference only.

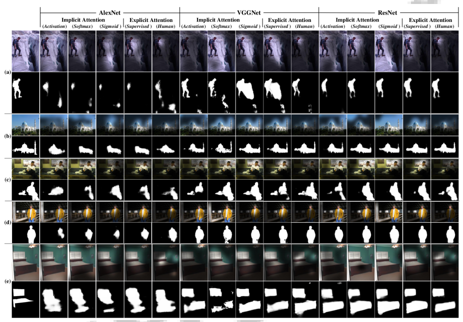
researchers can better understand gaze intricacies and its role in social interactions, informing more effective attention-based model design [29].

A comparative review highlights understanding human attention mechanisms to enhance attention-based model design in AI. This understanding is pivotal for developing AI systems exhibiting more generalized intelligence, capable of adapting to various tasks and environments [12]. The integration of AttentionRNN, combinable with any feed-forward convolutional network and end-to-end trainable, exemplifies modern attention mechanisms' flexibility in AI [62].

The ongoing challenge of acquiring high-quality human gaze data remains critical for training models replicating human-like attention behavior. A survey categorizing research based on application type and integration methods underscores gaze data's importance in enhancing model performance. This framework highlights the need for continued research and innovation to fully leverage human attention data in AI applications, leading to more sophisticated AI systems [23].



(a) Human clicks vs. Visible to DCN: 7 seconds to recognize[73]



(b) Attention-based human pose estimation on diverse scenes[72]

Figure 6: Examples of Integration of Attention Mechanisms in AI Models

As shown in Figure 6, integrating attention mechanisms in AI models advances visual attention, enhancing AI systems' ability to process and interpret complex visual data. In the first example, comparing human interaction and deep convolutional neural network (DCN) processing highlights AI's capability to mimic human attention patterns, albeit needing a brief processing period. The second example showcases attention-based human pose estimation across diverse scenes, comparing various neural network architectures. This comparison underscores attention mechanisms' versatility and effectiveness in enhancing AI's ability to estimate human poses in varied environments, underscoring attention's pivotal role in refining AI model performance [73, 72].

Feature	**Neural Processes and Attention Mechanisms**	**Visual Attention and Saliency**	**Interplay Between Visual and Social Attention**
Spatial Dynamics	Structured Spatial Attention	Saliency-guided Focus	Gaze And Gesture
Temporal Dynamics	Recurrent Modeling	Dynamic Scene Interpretation	Task-dependent Modulation
Multimodal Integration	Audio-visual Fusion	Simulates Human-like Attention	Shared Representation Learning

Table 4: This table provides a comparative analysis of various attention mechanisms, focusing on their spatial and temporal dynamics, as well as their approaches to multimodal integration. It highlights the distinct methodologies employed in neural processes, visual attention, and the interplay between visual and social attention, offering insights into their application in AI models.

6 Social Cognition

6.1 Definition and Role of Social Cognition

Social cognition involves mental processes crucial for interpreting and responding to social stimuli, facilitating effective communication and navigation in complex social environments. It underpins understanding others' intentions, emotions, and behaviors, essential for forming and maintaining social relationships. This cognitive capacity enables individuals to interpret cues, predict actions, and adapt responses contextually. The instinctive following of others' gaze exemplifies this ability, enhancing focus on relevant stimuli and showcasing the flexibility of social attention influenced by contextual factors [74, 14, 75, 24, 43]. The Predictive Coding Model (PCM) illustrates social cognition's role in simulating adaptive and maladaptive behaviors, highlighting its significance in theoretical and practical applications [76].

6.2 Contributions of Social Attention and Joint Attention

Social and joint attention are pivotal in shaping social cognitive processes, enabling perception and interpretation of social cues. Mechanisms like gaze following and eye contact enhance interpersonal communication, illustrated by the gaze-liking effect, which underscores gaze's role in understanding social signals [70]. The influence of gaze direction on engagement impacts attention and interaction dynamics [50]. Joint attention, involving shared focus on objects or events, is crucial for social cognitive development, particularly in early childhood, improving memory for objects observed by others [28]. Distinctions between gesture and joint attention reveal their unique contributions to language development [27].

In Autism Spectrum Disorder (ASD), joint attention impairments correlate with social-cognitive deficits. Technological interventions, including humanoid robots, show promise in enhancing joint attention and social communication skills in children with autism [26]. The reliability of eye gaze as a social cue supports theories of social cognition and joint attention, emphasizing their role in guiding attention and interaction [51]. Integrating activity and interaction awareness enhances joint attention estimation accuracy, improving social dynamics understanding [26]. Applications in artificial systems, such as autonomous decision-making architectures, demonstrate significant performance improvements in behavioral planning, capturing interaction patterns that yield human-like decision-making [10].

6.3 Impact of Gaze Cuing on Social Cognitive Processes

Gaze cuing significantly influences social cognitive processes by facilitating reflexive attentional shifts essential for interpreting social cues. It induces automatic attentional shifts in response to another's gaze direction, playing a critical role in social cognition dynamics [77]. Factors such as prior direct gaze and task type modulate the gaze cuing effect's magnitude [51]. Individual differences in social competence affect susceptibility to gaze cuing, with lower social competence individuals more influenced by irrelevant perceptual changes [43]. In developmental contexts, children process gaze cues from adults and peers similarly, indicating the universality of gaze cuing mechanisms [78]. However, cognitive load can disrupt perspective-taking, influencing gaze cue processing [77].

In clinical populations, such as individuals with ASD, altered gaze processing networks impact social cognitive processes, with differences in neural activation patterns affecting joint attention and gaze cuing capabilities [41]. Understanding eye gaze as a dynamic social signal is crucial for improving therapeutic interventions [74]. The gaze cuing effect is integral to social cognitive processes, influencing interpretations and responses to social cues. Variability in gaze cuing responses, shaped by individual differences, contextual interpretations, and clinical factors, highlights the complexity of social attention processes, necessitating ongoing research to elucidate these dynamics [79, 74, 14, 51, 70].

6.4 Emotional and Cultural Influences on Social Cognition

Emotions and cultural contexts significantly shape social cognition, affecting the perception and interpretation of social cues. Gaze and body cues convey emotional states and intentions, yet cultural differences in interpreting these cues complicate interactions [32]. Cultural norms surrounding gaze influence face perception and social interactions, necessitating culturally sensitive approaches in studying gaze behavior [4]. The COVID-19 pandemic highlights contextual factors' role in social cognition, as face masks alter social interaction dynamics and emotional state recognition [80]. This shift underscores the need to consider environmental influences on gaze behavior and cognitive processes [81]. Many studies conducted in non-interactive settings may not accurately reflect real-life complexities, limiting ecological validity [3].

Individual differences, such as autistic traits, modulate gaze cuing, while self-reported anxiety and depression do not significantly relate to gaze behavior [82]. Increased gaze following correlates with better emotional state recognition and greater conformity to group opinions, indicating that emotional acuity enhances social cognitive processes [79]. However, the mechanisms by which emotional expressions and face types influence gaze cuing remain unclear, highlighting the need for further investigation [51]. Cultural factors play a pivotal role in social cognition, with gaps in understanding how age, gender, and cultural context influence the relationship between trust and gaze cuing [83]. Existing research's focus on specific racial groups may overlook the complexity of gaze following across diverse populations, emphasizing the necessity for inclusive studies that account for a broader range of cultural backgrounds [84].

The interplay between emotional and cultural influences shapes how individuals perceive and respond to social cues. Cultural context alters gaze patterns during interactions, with East Asian and Western participants displaying different eye movement behaviors based on mutual or averted gaze perceptions, suggesting cultural norms critically influence facial expression interpretation and social attention engagement. Individual differences in gaze cuing effects link to emotional recognition and conformity to group norms, underscoring emotional and cultural factors' importance in social cognition. Continued research is essential to unravel these connections, enhancing our understanding of human social interactions [14, 59, 4, 79].

6.5 Applications and Implications for Social Cognition

Research in social cognition offers extensive applications across education, therapy, and technology, providing critical insights into human interactions. In educational contexts, integrating attention mechanisms can enhance cognitive engagement and learning outcomes. Optimizing visual tasks through computational efficiency can improve classroom interactions and address attention disorders [16]. Future research should examine gaze behavior implications in various instructional contexts and explore student-student gaze interactions to refine educational strategies [85].

In therapeutic settings, particularly for individuals with Autism Spectrum Disorder (ASD), interventions targeting gaze processing abnormalities show promise for improving social communication skills. The potential applications of these methods in social robotics for children with ASD highlight the practical implications of social cognition research in real-world settings [17]. Future studies should focus on larger sample sizes and additional factors affecting social attention in autistic individuals, such as perceived social dynamics, to enhance therapeutic interventions [22]. Moreover, research should investigate the gaze-liking effect's implications in social psychology through various experimental conditions [70].

Technological advancements have enabled the integration of social cognition insights into computational models, improving human-robot interaction systems. By leveraging gaze control and attention mechanisms, these systems exhibit increased flexibility and performance in dynamic environments, advancing robotics and artificial intelligence [44]. Future research could explore integrating additional social cues and language signals to enhance gaze target prediction and interaction quality [52]. Recognizing human attention mechanisms' role in enhancing AI performance, particularly in visual and language understanding tasks, underscores the need for robust gaze data collection methods [23].

The implications of social cognition research extend to multi-agent systems, where incorporating human-like attention mechanisms can enhance coordination and decision-making processes [48]. The success of MH-based models in predicting human acceptance behavior indicates practical applications for understanding symbol emergence in social cognition [56]. Future research may explore new applications based on this work, such as synthesizing first-person videos from third-person inputs [67]. Additionally, enhancing the Ego-Attention architecture's adaptability to various driving scenarios has implications for social cognition in autonomous systems [10].

Furthermore, the SCTFA-SNN model's superior performance in event stream classification compared to baseline models indicates its effectiveness in leveraging cognitive mechanisms from the brain [68]. The findings could inform human resource management practices by improving interpersonal dynamics understanding across various interaction modalities [39]. This study's comprehensive approach to examining gesture and joint attention together provides insights into their roles in language development [27].

The applications and implications of social cognition research are vast, offering significant potential for enhancing social interactions, therapeutic outcomes, and technological innovations. Continued exploration in this field is crucial for fully harnessing its benefits across various contexts, deepening our understanding of how individuals interpret social cues, such as gaze shifts, which reflect attentional focus and cognitive processing. This inquiry informs collaborative learning frameworks that leverage joint attention and interaction dynamics, ultimately leading to more effective strategies for fostering collaboration and enhancing perceptual performance in complex environments [14, 13, 9, 11].

6.6 Future Directions in Social Cognition Research

Future social cognition research should emphasize exploring gaze following across diverse populations and contexts to uncover cultural and contextual nuances currently underexplored [84]. Longitudinal studies are essential to tracking the developmental trajectory of gaze-cueing effects, particularly investigating the potential emergence of an own-age bias as children age [78]. In the context of Autism Spectrum Disorder (ASD), research should delve into how female interests impact attention, incorporating dynamic stimuli that reflect these interests to understand and address unique attention patterns in females with ASD [86]. Additionally, developing unified measurement models integrating social orienting and joint attention measures could enhance understanding of the bi-directional nature of social attention, advancing theoretical and practical applications [87].

Exploring joint attention principles in autonomous vehicles presents a promising research avenue, focusing on creating robust communication frameworks that adapt to varying traffic conditions [38]. Addressing gaps in understanding individual characteristics and cultural context roles in gaze-cueing is vital, employing diverse methodologies for a nuanced understanding of these factors' influence on social cognition [83]. Longitudinal research is also needed to investigate neural correlates of joint attention in diverse populations and explore strategies for enhancing joint attention in individuals with ASD [41].

Experimental paradigms that allow real-time gaze dynamics study, alongside investigations into neural mechanisms underlying gaze processing, could advance understanding of social cognition deficits [74]. Exploring gaze manipulation effects in non-face stimuli may yield new insights into social cognition, particularly in understanding gaze processing deficits [88].

Future research should refine NLP techniques in the AIC framework, applying them to diverse discourse contexts and integrating pedagogical designs to enhance collaborative learning [13]. Additionally, studies should investigate gaze and body cues dynamics in more naturalistic settings to improve ecological validity [32]. The dynamic interplay between teacher gaze and interpersonal behavior over time should be explored, potentially using advanced analytical techniques [1]. Furthermore, refining diagnostic tools for assessing cueing effects and exploring neuroanatomical correlates of gaze and arrow cueing in USN patients is crucial [58].

Exploring emerging trends in attention mechanisms, particularly how cognitive psychology insights inform effective AI architecture development, is essential for advancing AI research [12]. Future research should focus on optimizing the SCTFA module and exploring its application in other intelligent systems and neuromorphic computing [68]. Additionally, research should include larger and more diverse samples and explore additional cultural contexts and their effects on gaze behavior in social interactions [4].

Future social cognition research should aim to develop efficient attention mechanisms and explore integrating multi-modal data to improve generalization capabilities [16]. Addressing these areas will enhance understanding of social cognition and its applications across various domains, contributing to a deeper comprehension of human social behavior and interaction.

7 Conclusion

This survey consolidates pivotal insights into the role of social attention within cognitive processes, underscoring its foundational importance in facilitating effective social interactions and communication. The exploration of gaze processing anomalies in schizophrenia highlights the intricate nature of social cognition and paves the way for future research aimed at deciphering these complex dynamics. The efficacy of the MGPI network in modeling social interactions and identifying groups without explicit annotations exemplifies the potential of computational approaches to enhance our grasp of social dynamics.

Furthermore, the survey emphasizes the relevance of these findings for psychological research, particularly in developmental and clinical settings. It calls for empirical studies focused on gaze state transitions and individual differences in gaze behavior, which are crucial for a deeper understanding of social cognition across varied environments. These insights are instrumental in crafting interventions and models tailored to the diverse needs of individuals in both natural and artificial contexts.

The integration of social attention mechanisms into cognitive architectures and AI models offers promising avenues for advancing our understanding of human cognition and interaction. Continued research and innovation in this domain are vital to harness the full potential of social attention studies, ultimately enriching our comprehension of human social behavior and interaction.

References

[1] Nora A McIntyre, Kees Tim Mulder, and M Tim Mainhard. Looking to relate: Teacher gaze and culture in student-rated teacher interpersonal behaviour. *Social Psychology of Education*, 23:411–431, 2020.

[2] Chi Yan, Yunfei Gao, Saisai Hu, Fangxing Song, Yonghui Wang, and Jingjing Zhao. The impact and mechanism of gaze cues on object-based attention. *Acta Psychologica Sinica*, 54(7):748, 2022.

[3] Roy S Hessels. How does gaze to faces support face-to-face interaction? a review and perspective. *Psychonomic bulletin & review*, 27(5):856–881, 2020.

[4] Matthias S Gobel, Alyssa Chen, and Daniel C Richardson. How different cultures look at faces depends on the interpersonal context. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 71(3):258, 2017.

[5] Daniel Harari, Joshua B. Tenenbaum, and Shimon Ullman. Discovery and usage of joint attention in images, 2018.

[6] Sara M. González-Betancor and Pablo Dorta-González. Does society show differential attention to researchers based on gender and field?, 2023.

[7] Léo Hemamou, Ghazi Felhi, Vincent Vandenbussche, Jean-Claude Martin, and Chloé Clavel. Hirenet: a hierarchical attention model for the automatic analysis of asynchronous video job interviews, 2019.

[8] Navyata Sanghvi, Ryo Yonetani, and Kris Kitani. Mgpi: A computational model of multiagent group perception and interaction, 2020.

[9] Hugo Latapie, Ozkan Kilic, Kristinn R. Thorisson, Pei Wang, and Patrick Hammer. Neurosymbolic systems of perception cognition: The role of attention, 2021.

[10] Edouard Leurent and Jean Mercat. Social attention for autonomous decision-making in dense traffic, 2019.

[11] Jian Xu, Yi Bu, Ying Ding, Sinan Yang, Hongli Zhang, Chen Yu, and Lin Sun. Understanding the formation of interdisciplinary research from the perspective of keyword evolution: A case study on joint attention. *Scientometrics*, 117:973–995, 2018.

[12] Minglu Zhao, Dehong Xu, and Tao Gao. From cognition to computation: A comparative review of human attention and transformer architectures, 2024.

[13] Xinran Zhu and Bodong Chen. Understanding idea creation in collaborative discourse through networks: The joint attention-interaction-creation (aic) framework, 2023.

[14] Amit Zehngut and Shlomit Yuval-Greenberg. The interpretation attributed to observed gaze shifts affects their social-cueing effect. *bioRxiv*, pages 2024–07, 2024.

[15] Haoyue Ji, Li Wang, and Yi Jiang. Cross-category adaptation of reflexive social attention. *Journal of Experimental Psychology: General*, 149(11):2145, 2020.

[16] Mohammed Hassanin, Saeed Anwar, Ibrahim Radwan, Fahad S Khan, and Ajmal Mian. Visual attention methods in deep learning: An in-depth survey, 2024.

[17] Jack Hadfield, Georgia Chalvatzaki, Petros Koutras, Mehdi Khamassi, Costas S. Tzafestas, and Petros Maragos. A deep learning approach for multi-view engagement estimation of children in a child-robot joint attention task, 2018.

[18] R Gnana Praveen, Eric Granger, and Patrick Cardinal. Recursive joint attention for audio-visual fusion in regression based emotion recognition, 2023.

[19] Anirudh Vemula, Katharina Muelling, and Jean Oh. Social attention: Modeling attention in human crowds, 2018.

[20] Nick DePalma and Cynthia Breazeal. Towards learning through robotic interaction alone: the joint guided search task, 2016.

[21] Shizuka Suzuki, Takashi X Fujisawa, Nobuko Sakakibara, Toru Fujioka, Shinichiro Takiguchi, and Akemi Tomoda. Development of social attention and oxytocin levels in maltreated children. *Scientific reports*, 10(1):7407, 2020.

[22] Megan Freeth and Patricia Bugembe. Social partner gaze direction and conversational phase; factors affecting social attention during face-to-face conversations in autistic adults? *Autism*, 23(2):503–513, 2019.

[23] Giuseppe Cartella, Marcella Cornia, Vittorio Cuculo, Alessandro D'Amelio, Dario Zanca, Giuseppe Boccignone, and Rita Cucchiara. Trends, applications, and challenges in human attention modelling, 2024.

[24] Wenjie Liu, Xiangyong Yuan, Dingrui Liu, Li Wang, and Yi Jiang. Social attention triggered by eye gaze and walking direction is resistant to temporal decay. *Journal of Experimental Psychology: Human Perception and Performance*, 47(9):1237, 2021.

[25] Marco Gori. An introduction to cognidynamics, 2024.

[26] Chihiro Nakatani, Hiroaki Kawashima, and Norimichi Ukita. Interaction-aware joint attention estimation using people attributes, 2023.

[27] Virginia C Salo, Meredith L Rowe, and Bethany C Reeb-Sutherland. Exploring infant gesture and joint attention as related constructs and as predictors of later language. *Infancy*, 23(3):432–452, 2018.

[28] Samantha EA Gregory and Margaret C Jackson. Barriers block the effect of joint attention on working memory: Perspective taking matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(5):795, 2019.

[29] Mathis Jording, Arne Hartz, Gary Bente, Martin Schulte-Rüther, and Kai Vogeley. The "social gaze space": A taxonomy for gaze-based communication in triadic interactions. *Frontiers in psychology*, 9:226, 2018.

[30] Swethasri Dravida, J Adam Noah, Xian Zhang, and Joy Hirsch. Joint attention during live person-to-person contact activates rtpj, including a sub-component associated with spontaneous eye-to-eye contact. *Frontiers in human neuroscience*, 14:201, 2020.

[31] Omer Sumer, Peter Gerjets, Ulrich Trautwein, and Enkelejda Kasneci. Attention flow: End-to-end joint attention estimation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3327–3336, 2020.

[32] Sonia Betti, Umberto Castiello, Silvia Guerra, Umberto Granziol, Giovanni Zani, and Luisa Sartori. Gaze and body cues interplay during interactive requests. *PLoS One*, 14(10):e0223591, 2019.

[33] Francesca Capozzi, Andrew P Bayliss, and Jelena Ristic. Gaze following in multiagent contexts: Evidence for a quorum-like principle. *Psychonomic bulletin & review*, 25:2260–2266, 2018.

[34] Rindra Narison, Marie de Montalembert, Andrew Bayliss, and Laurence Conty. Measuring gaze and arrow cuing effects with a short test adapted to brain damaged patients with unilateral spatial neglect: A preliminary study. *Frontiers in psychology*, 12:690197, 2021.

[35] Zhifan Lu and Wieske van Zoest. Combining social cues in attention: Looking at gaze, head, and pointing cues. *Attention, Perception, & Psychophysics*, 85(4):1021–1033, 2023.

[36] Wei Cao, Wenxu Song, Xinge Li, Sixiao Zheng, Ge Zhang, Yanting Wu, Sailing He, Huilin Zhu, and Jiajia Chen. Could interaction with social robots facilitate joint attention of children with autism spectrum disorder?, 2018.

[37] Christian Giannetti. Advancing robot-assisted autism therapy: A novel algorithm for enhancing joint attention interventions, 2024.

[38] Amir Rasouli and John K Tsotsos. Joint attention in driver-pedestrian interaction: from theory to practice. *arXiv preprint arXiv:1802.02522*, 2018.

[39] Michela Balconi, Giulia Fronda, Federico Cassioli, and Davide Crivelli. Face-to-face vs. remote digital settings in job assessment interviews: A multilevel hyperscanning protocol for the investigation of interpersonal attunement. *Plos one*, 17(2):e0263668, 2022.

[40] Travis J Wiltshire, Johanne Stege Philipsen, Sarah Bro Trasmundi, Thomas Wiben Jensen, and Sune Vork Steffensen. Interpersonal coordination dynamics in psychotherapy: A systematic review. *Cognitive Therapy and Research*, 44:752–773, 2020.

[41] Peter Mundy. A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience*, 47(6):497–514, 2018.

[42] Dehua Tao, Tan Lee, Harold Chui, and Sarah Luk. Hierarchical attention network for evaluating therapist empathy in counseling session, 2023.

[43] Dana A Hayward and Jelena Ristic. Feature and motion-based gaze cuing is linked with reduced social competence. *Scientific Reports*, 7(1):44221, 2017.

[44] Maria Lombardi, Elisa Maiettini, Vadim Tikhanoff, and Lorenzo Natale. icub knows where you look: Exploiting social cues for interactive object detection learning, 2022.

[45] Di Fu, Fares Abawi, Erik Strahl, and Stefan Wermter. Judging by the look: The impact of robot gaze strategies on human cooperation, 2022.

[46] Di Fu, Fares Abawi, Hugo Carneiro, Matthias Kerzel, Ziwei Chen, Erik Strahl, Xun Liu, and Stefan Wermter. A trained humanoid robot can perform human-like crossmodal social attention and conflict resolution, 2023.

[47] Eduardo Davalos, Yike Zhang, Ashwin T. S., Joyce H. Fonteles, Umesh Timalsina, and Guatam Biswas. 3d gaze tracking for studying collaborative interactions in mixed-reality environments, 2024.

[48] Dennis Lee, Natasha Jaques, Chase Kew, Jiaxing Wu, Douglas Eck, Dale Schuurmans, and Aleksandra Faust. Joint attention for multi-agent coordination and social learning, 2021.

[49] Allison Gabouer and Heather Bortfeld. Revisiting how we operationalize joint attention. *Infant Behavior and Development*, 63:101566, 2021.

[50] Johanna Palcu, Jennifer Sudkamp, and Arnd Florack. Judgments at gaze value: Gaze cuing in banner advertisements, its effect on attention allocation and product judgments. *Frontiers in psychology*, 8:881, 2017.

[51] Kate T McKay, Sarah A Grainger, Sarah P Coundouris, Daniel P Skorich, Louise H Phillips, and Julie D Henry. Visual attentional orienting by eye gaze: A meta-analytic review of the gaze-cueing effect. *Psychological Bulletin*, 147(12):1269, 2021.

[52] Shiva Hanifi, Elisa Maiettini, Maria Lombardi, and Lorenzo Natale. icub detecting gazed objects: A pipeline estimating human attention, 2024.

[53] AL Karreman. The reliability of the gaze cueing effect. B.S. thesis, 2017.

[54] Inka Schmitz, Hanna Strauss, Ludwig Reinel, and Wolfgang Einhäuser. Attentional cueing: Gaze is harder to override than arrows. *Plos one*, 19(3):e0301136, 2024.

[55] Rafael Román-Caballero, Andrea Marotta, and Juan Lupiáñez. Target–background segregation in a spatial interference paradigm reveals shared and specific attentional mechanisms triggered by gaze and arrows. *Journal of Experimental Psychology: Human Perception and Performance*, 47(11):1561, 2021.

[56] Ryota Okumura, Tadahiro Taniguchi, Yosinobu Hagiwara, and Akira Taniguchi. Metropolis-hastings algorithm in joint-attention naming game: Experimental semiotics study, 2023.

[57] Tao Gao, Daniel Harari, Joshua Tenenbaum, and Shimon Ullman. When computer vision gazes at cognition, 2014.

[58] Rindra Narison, Marie De Montalembert, and Laurence Conty. Diagnosing gaze and arrow cueing effects in unilateral spatial neglect. *Neurocase*, 26(1):42–50, 2020.

[59] Zelin Chen, Sarah D McCrackin, Alicia Morgan, and Roxane J Itier. The gaze cueing effect and its enhancement by facial expressions are impacted by task demands: Direct comparison of target localization and discrimination tasks. *Frontiers in Psychology*, 12:618606, 2021.

[60] Takemasa Yokoyama and Yuji Takeda. Gaze cuing effects in peripheral vision. *Frontiers in Psychology*, 10:708, 2019.

[61] Jinbo Zhang, Xiang He, Werner Sommer, and Zhenzhu Yue. Does gaze direction of fearful faces facilitate the processing of threat? an erp study of spatial precuing effects. *Cognitive, Affective, & Behavioral Neuroscience*, 21:837–851, 2021.

[62] Siddhesh Khandelwal and Leonid Sigal. Attentionrnn: A structured spatial attention mechanism, 2019.

[63] Roberta Capellini, Paolo Riva, Paola Ricciardelli, and Simona Sacchi. Turning away from averted gazes: The effect of social exclusion on gaze cueing. *Frontiers in Psychology*, 10:1000, 2019.

[64] Austin Erickson, Nahal Norouzi, Kangsoo Kim, Ryan Schubert, Jonathan Jules, Joseph J LaViola Jr, Gerd Bruder, and Gregory F Welch. Sharing gaze rays for visual target identification tasks in collaborative augmented reality. *Journal on Multimodal User Interfaces*, 14(4):353–371, 2020.

[65] Nikhil Khatri, Tuomas Laakkonen, Jonathon Liu, and Vincent Wang-Maścianica. On the anatomy of attention, 2024.

[66] Yujie Chen, Yufang Zhao, Hongwen Song, Lili Guan, and Xin Wu. The neural basis of intergroup threat effect on social attention. *Scientific reports*, 7(1):41062, 2017.

[67] Huangyue Yu, Minjie Cai, Yunfei Liu, and Feng Lu. What i see is what you see: Joint attention learning for first and third person video co-analysis, 2019.

[68] Wuque Cai, Hongze Sun, Rui Liu, Yan Cui, Jun Wang, Yang Xia, Dezhong Yao, and Daqing Guo. A spatial-channel-temporal-fused attention for spiking neural networks, 2023.

[69] Chinmaya Mishra and Gabriel Skantze. Knowing where to look: A planning-based architecture to automate the gaze behavior of social robots, 2022.

[70] Jason Tipples and Anna Pecchinenda. A closer look at the size of the gaze-liking effect: A preregistered replication. *Cognition and Emotion*, 33(3):623–629, 2019.

[71] Geonsun Lee, Dae Yeol Lee, Guan-Ming Su, and Dinesh Manocha. "may i speak?": Multimodal attention guidance in social vr group conversations, 2024.

[72] Qiuxia Lai, Salman Khan, Yongwei Nie, Jianbing Shen, Hanqiu Sun, and Ling Shao. Understanding more about human and machine attention in deep neural networks, 2020.

[73] Drew Linsley, Dan Shiebler, Sven Eberhardt, and Thomas Serre. Learning what and where to attend, 2019.

[74] Roser Cañigueral and Antonia F de C Hamilton. The role of eye gaze during natural social interactions in typical and autistic people. *Frontiers in psychology*, 10:560, 2019.

[75] Mathis Jording, Denis Engemann, Hannah Eckert, Gary Bente, and Kai Vogeley. Distinguishing social from private intentions through the passive observation of gaze cues. *Frontiers in human neuroscience*, 13:442, 2019.

[76] D. Rudrauf, G. Sergeant-Perthuis, O. Belli, Y. Tisserand, and G. Di Marzo Serugendo. Modeling the subjective perspective of consciousness and its role in the control of behaviours, 2021.

[77] John Michael, Thomas Wolf, Clément Letesson, Stephen Butterfill, Joshua Skewes, and Jakob Hohwy. Seeing it both ways: Using a double-cuing task to investigate the role of spatial cuing in level-1 visual perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance*, 44(5):693, 2018.

[78] Rianne Van Rooijen, Caroline Junge, and Chantal Kemner. No own-age bias in children's gaze-cueing effects. *Frontiers in Psychology*, 9:2484, 2018.

[79] Won-Gyo Shin, Hyoju Park, Sung-Phil Kim, and Sunhae Sul. Individual differences in gaze-cuing effect are associated with facial emotion recognition and social conformity. *Frontiers in psychology*, 14:1219488, 2023.

[80] Han Jia, Qi Wang, Xinghe Feng, and Zhonghua Hu. Face mask reduces gaze-cueing effect. *Scientific Reports*, 13(1):13160, 2023.

[81] Shane L Rogers, Craig P Speelman, Oliver Guidetti, and Melissa Longmuir. Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific reports*, 8(1):4271, 2018.

[82] Sarah D McCrackin and Roxane J Itier. Individual differences in the emotional modulation of gaze-cuing. *Cognition and Emotion*, 33(4):768–800, 2019.

[83] Mariapaola Barbato, Aisha A Almulla, and Andrea Marotta. The effect of trust on gaze-mediated attentional orienting. *Frontiers in Psychology*, 11:1554, 2020.

[84] Max Weisbuch, Kristin Pauker, Reginald B Adams Jr, Sarah A Lamer, and Nalini Ambady. Race, power, and reflexive gaze following. *Social Cognition*, 35(6):619–638, 2017.

[85] Eeva Haataja, Visajaani Salonen, Anu Laine, Miika Toivanen, and Markku S Hannula. The relation between teacher-student eye contact and teachers' interpersonal behavior during group work: A multiple-person gaze-tracking case study in secondary mathematics education. *Educational Psychology Review*, 33(1):51–67, 2021.

[86] Clare Harrop, Desiree Jones, Shuting Zheng, Sallie W Nowell, Brian A Boyd, and Noah Sasson. Sex differences in social attention in autism spectrum disorder. *Autism research*, 11(9):1264–1275, 2018.

[87] Peter Mundy and Jenifer Bullen. The bidirectional social-cognitive mechanisms of the social-attention symptoms of autism. *Frontiers in Psychiatry*, 12:752274, 2022.

[88] Nouchine Hadjikhani, Jakob Åsberg Johnels, Nicole R Zürcher, Amandine Lassalle, Quentin Guillon, Loyse Hippolyte, Eva Billstedt, Noreen Ward, Eric Lemonnier, and Christopher Gillberg. Look me in the eyes: constraining gaze in the eye-region provokes abnormally high subcortical activation in autism. *Scientific Reports*, 7(1):3163, 2017.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.