# WHAT CAPTURES OUR ATTENTION? HOW COMMUNICATIVE AND NON-COMMUNICATIVE ACTIONS BY HUMAN AND ROBOT AGENTS INFLUENCE VISUAL ATTENTION UNDER PERCEPTUAL LOAD

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

Tuvana Dilan Karaduman

August 2025

What Captures Our Attention? How Communicative and Non-Communicative Actions by Human and Robot Agents Influence Visual Attention under Perceptual Load
By Tuvana Dilan Karaduman
August 2025

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 Urgen(Advisor)

Zahide Pamir Karatok

Didem Kadıhasanoğlu

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

# ABSTRACT

## WHAT CAPTURES OUR ATTENTION? HOW COMMUNICATIVE AND NON-COMMUNICATIVE ACTIONS BY HUMAN AND ROBOT AGENTS INFLUENCE VISUAL ATTENTION UNDER PERCEPTUAL LOAD

Tuvana Dilan Karaduman
M.S. in Neuroscience
Advisor: Burcu Ayşen Urgen
August 2025

This study investigated how agent type (human vs. robot), communicativeness (communicative vs. non-communicative), and perceptual load (low vs. high) interactively influence visual attention and task performance. A 2 (Agent) $\times$ 2 (Communicativeness) $\times$ 2 (Perceptual Load) within-subjects design was employed with 34 participants. To examine attentional allocation under perceptual load, Areas of Interest (AOIs) were defined for the central letter array and the peripheral agent videos. Behavioral performance (reaction time, accuracy) and multiple eye-tracking metrics (total dwell time, dwell time per fixation, first fixation latency, saccadic velocity, pupil dilation) were analyzed.

Behavioral results confirmed the effect of the perceptual load manipulation, with participants being significantly less accurate and slower in high-loaded conditions. In the presence of a distractor, reaction times and accuracy were further modulated by significant interactions between agent, perceptual load, and communicativeness. Eye-tracking analyses revealed that initial orienting was driven by the communicativeness and perceptual load, when communicative cues captured attention faster only under high load. Dwell time per fixation remained stable across the conditions. On the other hand, total dwell time revealed a critical three-way interaction. Under high load, communicative actions performed by humans increase total dwell time, whereas if the action is performed by a robot, dwell time decreases. Pupil data supported this, indicating that under high load, non-communicative actions required more cognitive effort than the communicative ones. Initial orienting of eye metrics was guided by task demand; overall engagement with the nature of the action is a product of a complex and sensitive cognitive process sensitive to perceptual load. Robotic social cues are processed dissimilar compared to human ones.

*Keywords:* Visual Attention, Perceptual Load Theory, Biological Motions, Human-Robot Interaction, Eye Tracking, Distraction .

# ÖZET

## DIKKATIMIZI NE ÇEKIYOR? ROBOT VE INSAN TARAFINDAN YAPILAN ILETISIM IÇEREN VE IÇERMEYEN HAREKETLERIN FARKLI ALGISAL YUK KOŞULLARINDA GORSEL DIKKAT INCELEMESI

Tuvana Dilan Karaduman
Nörobilim, Yüksek Lisans
Tez Danışmanı: Burcu Ayşen Ürgen
Ağustos 2025

Çalışmamızda bir harf arama görevi sırasında görsel dikkatin nasıl etkilendiğini incelemek amacıyla aktör türü (insan veya robot), hareketin iletişimselliği (iletişimsel veya iletişimsel olmayan) ve algısal yük (düşük veya yüksek) faktörlerinin etkileşimini hem davranışsal (tepki süresi, isabet oranı) hem de göz izleme (toplam bakış süresi, her bir fiksasyon başına düşen bakış süresi, sakkadik hız, ilk fiksasyon gecikmesi, göz bebeği büyümesi) ölçümleriyle araştırılmıştır. Göz hareketleri, EyeLink 1000 Plus sistemi (SR Research, Kanada) kullanılarak 1000 Hz örnekleme hızıyla kaydedilmiştir. İlgilenilen alanlar, algısal yük teorisi kapsamında merkezi harf dizisi ve çevresel aktör videoları etrafında tanımlanmıştır. Çalışma, 2 (Aktör) $\times$ 2 (İletişimsellik) $\times$ 2 (Algısal Yük) dizaynı ile yürütülmüş ve beş okülomotor değişken üzerinde tekrarlı ölçüm ANOVA'ları yapılmıştır.

Dikkat dağıtıcı içermeyen denemelerin tepsi süresi ve isabet oranları, algısal yük manipülasyonunun etkisini doğrulamıştır, katılımcılar yüksek yük koşulunda istatistiki olarak anlamlı bir şekilde daha yavaş tepki vermiş ve daha az isabetli yanıt vermişlerdir. Dikkat dağıtıcıların varlığında, reaksiyon süreleri ve doğruluk oranları, aktör, algısal yük ve iletişimsellik arasındaki anlamlı etkileşimlerle şekillenmiştir. Göz izleme analizleri ilk fiksasyon gecikmesinin iletişimsellik ve algısal yük tarafından etkilendiğini, iletişimsel aksiyonların sadece yüksek yük koşulunda dikkatin daha hızlı yakalandığını ortaya çıkartmıştır. Fiksasyon başına bakış süresi, koşullar arası sabit kalmıştır. Toplam bakış süresi, anlamlı bir üçlü etkileşimi ortaya çıkartmış; yüksek yük altında insanlar tarafından

gerçekleştirilen iletişimsel eylemlerin toplam bakış süresini arttığı durumda, eylem robot tarafından gerçekleştirildiğinde bu süre azalmıştır. Göz bebeğinde yapılan büyüme analizleri bu bulguları desteklemiş, yüksek yük altında iletişim içermeyen eylemlerin iletişim içeren eylemlere kıyasla daha fazla bilişsel çaba gerektirdiği gözlemlenmiştir. Dikkat algısal yük tarafından yönlendirilmektedir ve aktör ile olan etkileşim, algısal yüke göre değişebilen karmaşık bir bilişsel sürecin ürünüdür. Robotların sergilediği aksiyonların, insanlardan gelen sosyal ipuçlarından farklı şekilde işlenmesi söz konusudur.

*Anahtar sözcükler*: Görsel Dikkat, Algısal Yük Teorisi, Biyolojik Hareket, İnsan-Robot Etkileşimi, Goz İzleme, Dikkat Dağıtıcı .

# Acknowledgement

First and foremost, I would like to express my heartfelt gratitude to my dear advisor, Burcu Ayşen Ürgen, who has taught me what it truly means to be a scientist. She has shown me that academia and science are not merely about publishing papers, but about curiosity, creativity, and resilience. Over the years, her guidance has shaped me in countless ways—academically and personally. I am deeply thankful for her unwavering support and for showing me that academic life can be both colorful and full of joy. I feel incredibly lucky to have shared this journey with her.

I am also sincerely grateful to Zahide Pamir and Didem Kadıhasanoğlu for their valuable insights and constructive feedback on my thesis. Their contributions played an essential role in shaping and finalizing this work.

To my beloved family, Dilek, Kadir, İclal, and Sakine, thank you for your boundless love and endless trust. I carry the strength of your support with me always, just as I did on my very first day of elementary school. Knowing you are still by my side today gives me courage beyond words.

To my second family, Barış, thank you for being my constant source of support throughout this process. Without you, everything would have felt a little less complete. I am endlessly grateful for your presence, your encouragement, and your love.

Lastly, to my dearest friends, Feza, Sezan, Tuğçe, and Şeyda, our paths crossing on this journey has been one of the greatest gifts. Knowing that scientists like you exist fills me with hope for the future. I'm truly grateful that this adventure brought us together.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Every day, humans encounter enormous amounts of incoming information from the environment, regardless of their complexity, most of which requires attention. Attention refers to the process of selecting stimuli while inhibiting other information from the environment. The amount of information coming from the environment is more than a person's ability to process. With this amount of information, our perceptual system will be overloaded without a selection system for attention. The environment is full of potential distractors, ranging from other humans to inanimate objects. Although eliminating all distractions is unfeasible, trying to understand how humans select where to direct their attention and how attention is distracted remains a question for attention research. To explore how this process works, the visual domain of the research can be easily focused on, where the examination of the mechanism is much more feasible than other perceptions.

Visual attention is a cognitive process that enables individuals to focus on and process visual information within their environment selectively. This selective process determines which visual stimuli are consciously perceived and acted upon, while unattended inputs remain unprocessed or suppressed from awareness. This selection can be active or passive, but the information that enters the visual system is still huge [1]. Sixty-seven years ago, Broadbent [1] proposed a model for perception called the Selective Filter Theory, which argues that basic physical features of the stimuli are temporarily stored in the human brain, but the abstract meaning of the stimuli has a limited but not predetermined amount of capacity. Because of this limitation, filtering the selected stimuli and regressing out others to match capacity is named early selection models for attention, which either filter or attenuate the unintended stimuli [1–3]. The early models provided a good base for explaining unattended information that remains unknown to the subjects. Most of the evidence for the early selection model comes from the dichotic listening paradigm, where the left and right ears are presented with different stimuli, and the subject has to report information only coming from the attended ear, which cancels out the unattended ear [4]. In contrast, rather than these findings, galvanic skin response increased if the participant encountered a sound previously presented with an electrical shock [5]. This indicates that unattended stimuli are not fully canceled out by perception. In 1959, Moray found that the unattended information may proceed beyond the simple physical level, and the subject can identify their names when they hear it from the unattended ear [6]. Whereas late selection models alternate the hypothesis that selection occurs after getting all the semantic and physical input [2]. Pashler argued that there is no filtering process of the whole unattended stimuli, but there is a difference in the way these two are processed [7]. Kahneman criticized the previous perspectives through the attention as a selective mechanism and argues that attention has limited resources, and current research at that time gave participants enormous amounts of information at one time, relevant and irrelevant information is not discriminated [8]. After the debates about the early selection and late selection models, combined models were studied with simpler tasks and fewer stimuli, and it was suggested that the filter can be a shifting system rather than a constant filter due to the task demands [9].

## 1.1 Perceptual Load Theory

Bottleneck theories, named early and late selection models, are important for understanding the nature of attention, what is being unattended, and the processes behind it, as examined by differences in physical characteristics of the attended and unattended channels. The research shifts from the central problem to the physical distinctions in research, especially the characteristics of the stimuli as their relation to the central attention and relevance of the stimulus, rather than the amount of load [10, 11]. Color, size, and location of the stimuli are found to have an effect on the selection process [12]. In the famous 1995 paper, Nilli Lavie argues that this physical distinction retains its importance, but relevance is essential too. More relevant stimuli may help individuals to process more, but it does not mean that less relevant stimuli remain unprocessed. Perceptual processing has a capacity, and in order not to exceed this capacity, selective processing comes into play. If the demand of the stimuli stays below the level of the capacity, the remaining cognitive resources may be used to process irrelevant stimuli too [11]. In between the early and late selection models, in perceptual load theory, the process never stops until the capacity of the attentional resources is no longer available to process more information coming from the environment. In that sense, both internal (cognitive load of a person) and external (physical characteristics of the stimuli) work simultaneously. At the behavioral level, a load of the task can be differentiated in three ways: changing the task, the number of tasks, or the relevance of the stimuli. The option of changing tasks is beneficial in situations where stimuli stay constant since the task is changing not according to anything related to the stimuli [13]. Where the number of items manipulated, low-load conditions have fewer items than high-load conditions [14], and third, the target and non-items' similarity may be manipulated; when the similarity increases, the load of the task increases. Trying to find the letter "X" surrounded by similar angular letters, such as "V" or "Z" is considered to require a higher perceptual capacity than trying to find the letter "X" surrounded by a dissimilar letter "O" [13]. Information and procedure about the letter search paradigm are explained in the next section, Letter Search Task.

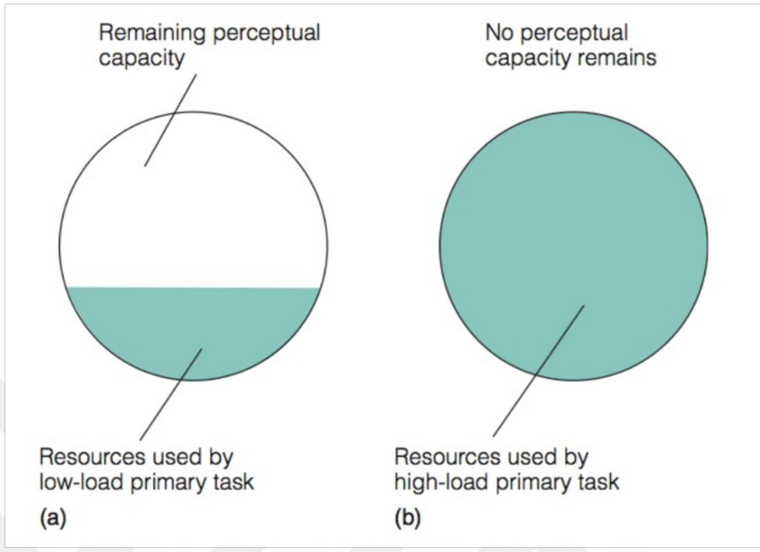



Figure 1.1: Resources in (a)Low-load and (b)high-load tasks according to perceptual load theory. The complexity of the task increases the amount of allocated space on cognitive resource capacity. Reprinted from [15] with permission.

## 1.2 Letter Search Task

A widely adopted method for examining selective attention and distraction is the letter search task, a form of visual search in which participants identify a predefined target letter (e.g., ”X” or ”N”) presented among nontarget distractors. The task requires participants to scan an array of letters and report the presence or absence of the target as quickly and accurately as possible. It allows attentional resources to be used on different perceptual load (low-load vs. high-load) intercepts. In low-load conditions, nontarget letters are homogeneous and easily distinguishable from the target letter, allowing cognitive resources to process other stimuli rather than the main task. In contrast, high-load conditions have heterogeneity among distractors, increasing task difficulty and reducing the likelihood of processing irrelevant information [11, 15]. This paradigm has been instrumental in testing models of attentional capture, especially in dual-task scenarios. Peripheral or centrally presented distractors, visually salient but irrelevant, can be introduced during the letter search to examine the degree to which attention is involuntarily drawn away from the target, depending on factors such as perceptual load, individual differences, or semantic relevance [16,17]. Crucially, the letter search task enables manipulation not only of perceptual difficulty but also of top-down control settings. Depending on whether attention is allocated based on the features of the stimuli or the spatial conditions, the impact of the distractors may change, and the aim is to evaluate how task-irrelevant stimuli can capture attention [9, 18]. In the current study context, the letter search task served as a controlled attentional baseline against which the influence of noncentral, socially meaningful distractors, such as robots or human agents, could be assessed. We aim to clarify how attentional prioritization unfolds under different cognitive demands by varying perceptual load and the communicative nature of peripheral stimuli.

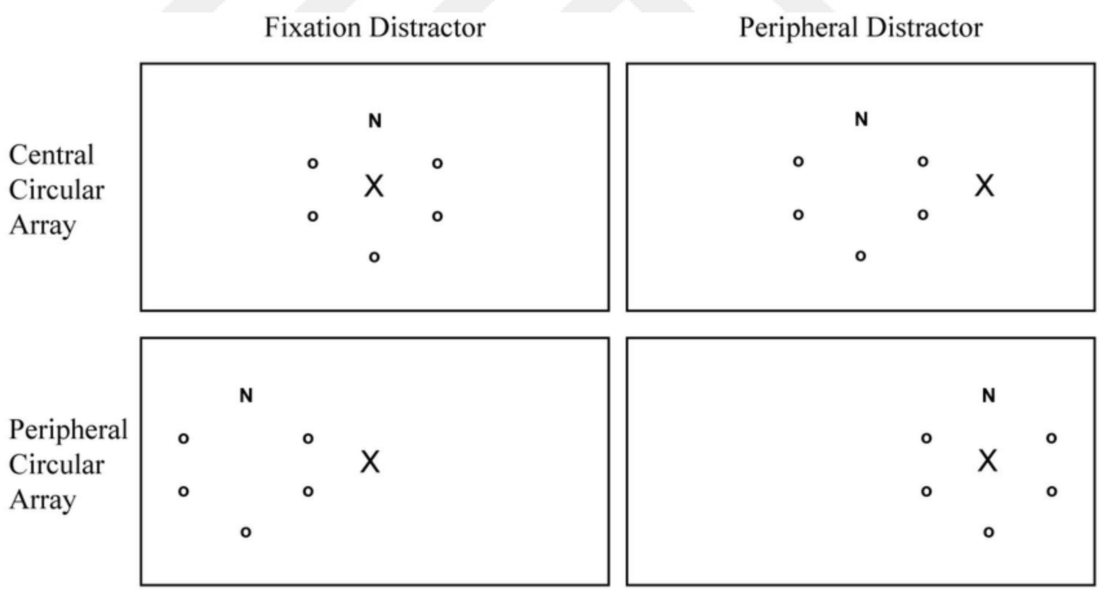



Figure 1.2: Fixation and Peripheral Distractors with Central Circular Array and Peripheral Circular Array. A central circular array with a peripheral distractor was used in the current study. Reprinted from [15] with permission.

## 1.3 Attention: Top-down and Bottom-up

What we see and what we can act upon are shaped by attentional mechanisms. According to Posner, attention can be endogenous or exogenous. In endogenous attention, goal-driven attention was applied, with the subject in control of the channeling of the attention [19], also mentioned in the literature as "top-down" attention [20]. A student who is focused on the board while listening to the teacher uses a top-down attention process. In contrast, exogenously driven attention is an external force that draws the subject's attention in a "bottom-up" manner [21, 22]. If a colleague drops his keys and suddenly our gaze is directed to the location of the sound, it has a bottom-up nature. Even though they are different mechanisms behind them, they interact to understand the surroundings, too [23, 24]. Also, the neural correlates of those mechanisms have quite a distinctive pattern. Bottom-up attention originates in early visual cortices from V1-V4 and is presented in parietal (PPC, LIP) and prefrontal (PFC, FEF) areas. Saliency feature-specific maps for color, orientation, etc, are associated with bottom-up attention. On the other hand, top-down attention originates from the PFC and modulates the sensory feedback mechanism. Inhibition of the distractors, filtering, and selection of targets are associated with the top-down process. Both attention types have an effect on overlapping frontoparietal circuits and are intertwined rather than two distinct systems [25]. Regardless of the environment, internal control of where to attend as top-down processes determine which information will be prioritized and which is not [9, 26]. Unattended stimuli may allocate attention due to remaining capacity. In that sense, late attention theories come into play because inhibition of a stimulus likely occurs in later stages of perception [27, 28]. There are three different components to perform a task: effort, capacity, and allocation procedure. His arguments guide us to examine the voluntary (top-down) and reflexive (bottom-up) processes of attention by emphasizing the role of arousal [29]. These foundation models of attention emphasize different mediums for selection, limits of attention, and allocation strategies, shaping our understanding of visual attention dynamics and how its multi-layered nature functions in real-world situations based on stimulus, internal resources, and goals. Eye tracking metrics as an objective measure are a crucial tool for understanding these mechanisms by providing real-time insight into overt attention shifts and possible cognitive resource allocation.

Distractor interference in perceptual load is a well-established phenomenon in the literature. Due to increased task difficulty, high-load tasks have more errors and longer reaction times when compared with low-load tasks [14]. In contrast, Yeshurun and Marciona did a study with target and distractors, just distractors and just target, and results reveal that low and high distractors were found to have similar effects [30]. Another aspect of the distractors is distractor inhibition, a result of negative priming in low-load tasks, the negative priming effect, but when the load was increased, the effects diminished [28, 31, 32].

Functional magnetic resonance imaging (fMRI), electroencephalography (EEG) and studies that used simultaneous EEG and fMRI showed neural correlations modulated by the load level of the stimulus [33–36]. In a study with moving distractors, the number of syllables of the stimuli was a high load, or the decision between upper- and lower-case words presented, while there were moving distractors in the background. In high load conditions, compared to low load, where distractors in the background evoked potentials in the areas MT, V1, V2, and V5, responses were reduced in the brain, supporting the load theory. In the same manner, in high-load parahippocampal activity was reduced for distractors compared to low-load trials. Since the parahippocampal gyrus is related to memory, and after many trials, repetition of the distractors occurs, it can be said that the brain is affected in low-load trials because of repetition of the stimuli [37].

Relevant and irrelevant distractors showed a similar effect for load theory. In 2008, Forster and Lavie used cartoon characters as distractors as they argued that ignoring meaningful distractors is hard compared to meaningless (cartoon characters), using relevant and irrelevant distractors has similar effects, both can be modulated by perceptual load [38]. The faces were then used, as they have high social information, and neuroimaging studies support this suggestion [39]. In the same sense, perceptual load theory applies to celebrity faces, whereas non-face and animal distractors are applied [40, 41]. Because of these contradictory results, researchers suggest that the effect might be a product of familiarity rather than the importance of human faces. To investigate that, car, bird, and music experts examined whether these different results about the faces come from the social relevance ıf faces or humans' expertise in human faces, under a high-load perceptual task, familiar items have an effect for experts [42, 43].

## 1.4 Communicative Actions

Allocation of attention conditioned by social stimuli is named social attention [44]. Social attention examples can be found through attending to a socially meaningful stimulus or attending to a socially cued stimulus. This communication can of course be spoken, but an unspoken version of a communication may carry the meaning too [45]. For example, a friend's gaze direction includes a social meaning. Observable cues about gaze or pointing behavior were well studied in current literature because of the importance of others' attentional directions, which is an important cue for human beings [46]. Those cues are important for social organisms like humans since they may provide fruitful information about the environment, avoiding danger, collaborating with others, and humans' ability to achieve their goals [47]. Perception of a person, social discrimination, and social structures can be significantly influenced by the way allocation of attention works [48]. Orientation of attention refers to the direction of another person's movement, often conveyed by the pointing behavior [49]. Attentional engagement refers to the use of attentional resources by looking at something or someone in order to get information about it [50, 51]. Attentional communication refers to signaling another entity by gaze directions and movements [52, 53]. In addition to eyes, other non-verbal communicative cues such as pointing can affect how humans' attention is allocated and perform under different conditions [54]. In 2016, Admoni proposed that non-verbal behavior and memory performance increased, and completion time decreased when the task was challenging. In contrast, in easy tasks, nonverbal communicative cues have a decreased effect. Communicative actions increase attentional value with cognitive demand, and the modulation of attention in perception can be influenced by the cognitive demand of the context.

The well-studied area of communicative actions, explored in cognitive level psychology, is how the gaze direction of others influences one's attentional system, both at the behavioral and psycho-physical levels. Previous studies suggest that a directional gaze is processed differently than other directional cues, such as arrow signs [55, 56]. The studies conducted with functional magnetic resonance imaging, eyes, and other social cues activate the brain in different pathways [56].

Understanding the meaningful exchanges between humans and robots in every medium is crucial [57]. Robot nonverbal cues can be divided into subcategories such as body movements, the usage of social space, and eye gaze of a robot [58]. Within these different media, specific information can be conveyed through communicative actions, such as pointing, representing an action, or mimicking an abstract concept, which can be an intentional design [59]. Also, the eye gaze of a robot can signal interpersonal intent [60]. A growing body of research demonstrates that using these communicative behaviors influences how humans frame the robot and affects the understanding of emotional capability while improving the collaborative nature between human and robot [61]

## 1.5 Robots as Distractors

The presence of robots has been a fascinating branch of technology and science fiction, and since the Industrial Revolution, it has been driven by the rapid development of technological breakthroughs. This presence has occupied humanity's agenda and has now also begun to be a part of people's daily lives in different forms, from the eighteenth century to today's machines. The evolution of robots aligns with human invention. From steam-powered machines to robots, human horizons revolved while horizons revolutionized them, too. Day by day, robots' inclusion in daily life increases in every form, whether social or mechanical robots, regardless of where they are going to be used in healthcare, logistics, education, and domestic settings. A wide range of robotics systems is used, ranging from surgical assistants to vacuum cleaners [62,63]. This inclusion raises a fundamental question for the future of human-kind robot interactions. Deviating from original

Isaac Asimov's thoughts about the presence of robots preserves its importance and drives the research [64] by questioning the trust in robots and their relationship with humans. Answers to the questions will shape our future, not only devices in the home but also the research agenda for engineers, designers, and scientists as well. In this sense, human-robot interaction and its application are a highly interdisciplinary aspect of the future of robots. Human-robot interaction is not driven by one aspect, and it is cumulative through knowledge of engineering, computational methodology, psychological insight, and social aspects.

Physical and social capabilities are a distinctive way to distinguish traditional robots from socially active robots. Building or manipulating a physical being is named conventional robotics, whereas HRI investigates the role of the robot in the real world in a social context. HRI differs from the mechanical way of monitoring robots and gives room for understanding the psychological and sociological aspects. As an interdisciplinary work field, HRI also affects human-computer interaction, philosophy, design of technology, and artificial intelligence. Collaborative development of all these methodologies enables researchers to tailor new methods and theoretical frameworks.

What truly distinguishes HRI is its emphasis on embodied social agents—robots that people perceive not merely as tools but as interactive partners endowed with cultural significance. The physical form of a robot imposes constraints on sensing and action, yet it also affords rich opportunities for social engagement. When a robot's appearance and motion mimic human gestures or expressions, people naturally apply their existing knowledge of social interaction, but they may also become frustrated if the robot fails to meet those expectations. Ultimately, HRI is concerned with designing robotic systems that operate effectively across diverse real-world settings.

Most of the robot's design features are made from an anthropocentric point of view; robots resemble the nature of the human body and the human mind. As a result of the growing nature of human-robot interaction, technical and design-related qualities should be examined in these embodied agents in order to understand how humans react to these agents and strive to be more humanoid day by day.

As robots evolve beyond mere instruments into collaborators, friends, tutors, and daily assistants, HRI research must address not only the physical and social design of these technologies but also their cultural integration to the world, organizational implementation, and broader societal impact questions that transcend the scope of any single parent disciplines such as the view from engineering and psychology.

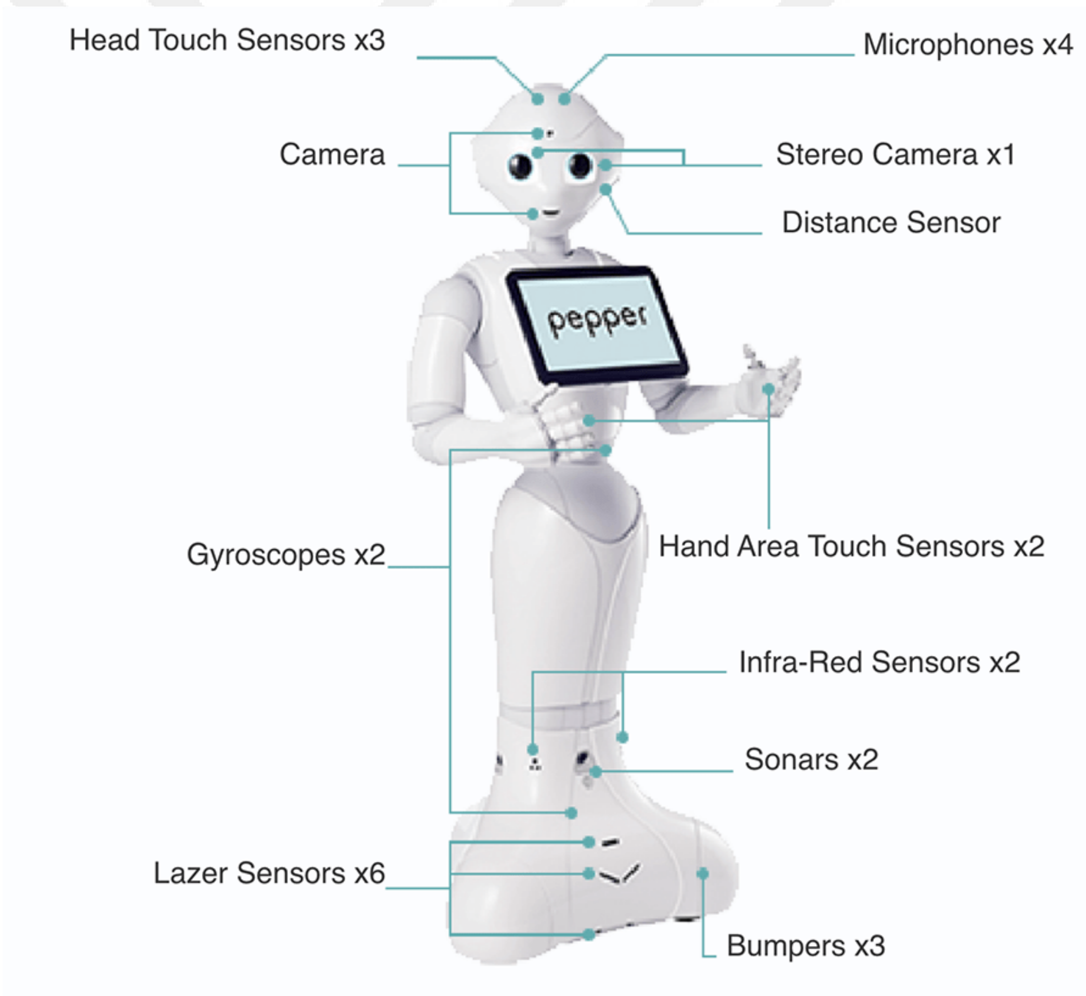



Figure 1.3: Image of Pepper robot (2014-present), 121 cm tall humanoid robot. Interaction with humans is enabled by the robot's voice and tablet.

## 1.6 Eye-tracking as a Tool for Measuring Visual Attention

Eye tracking is a widely used method in behavioral and cognitive sciences that detects oculomotor data during various tasks, primarily visual. The cognitive effort, attentional mechanisms, and perceptual load can be examined through real-time eye tracking. The allocation of visual resources between different settings is objectively quantified by tracking the different oculomotor activities of human gaze data. Eye tracking research uses different metrics in oculometric research: pupilometry, fixations, saccadic velocity, total dwell time, dwell time per fixation, and first fixation onset.

Differences in pupilometry are often associated with mental workload. Task difficulty has been found to have an effect on pupil size, and this effect can be seen across the different levels of difficulty and individual differences [65]. The strength of the effort can be understood better by pupil size, but the origin of the effort remains unclear; perceptual and cognitive demands should be distinguished by the different loads. To solve this problem, researchers have begun to use fixation-based oculomotor metrics instead of pupilometry. Findings are contradictory; some studies suggest that high load fixation duration increases [66], and others report a decrease in fixation duration [67].

Fixations refer to brief relative pauses in eye movements, typically lasting more than 100ms, and encounters with active processing of visual information of the gaze direction [68, 69]. This gaze direction indicates where attention is directed in the current visual field and where it is observing information of attention. Reflecting the perception of information and active cognitive processing makes fixations a main indicator of where overt attention is focused [70, 71].

Saccadic velocity is the peak speed of fast eye movements between fixations called saccades. Search behavior and attentional engagement can be examined by these fast movements of the gaze. Fast saccadic velocity often indicates the task is demanding, indicating search behavior in the visual field and attentional shift [72, 73]. Peak saccadic velocity increases when task difficulty increases. Earlier findings suggest that when the task is more effortful and increases the attentional load, faster saccades are made [74]. Saccadic eye metrics, especially velocity of the saccades, reveal the efficiency of the visual search and physiological effort for the task and possible attentional shift toward main tasks to distractors [70, 71].

Total dwell time is the cumulative duration of the fixations and all gaze movements in a given specified area called the Area of Interest (AOI). Defining AOIs is an important part of the analyses; focusing on specific regions of a given stimulus enhances the quality of the research. Which AOI is visited first, and fixation duration gives valuable information about the prioritization of the stimuli and also which part is suppressed [75]. A study with experts revealed that when compared to radiologists and naive participants, fixations are faster and directed onto lesions, whereas naive participants have different facial areas with less correct identification [76]. Using total dwell time, measure the overall attentional investment [77, 78]. General cognitive engagement and sustained attention are reflected by total dwell time spent in a specified region of interest [70, 79].

Dwell time per fixation, also named mean fixation duration in some papers, is found by dividing the total dwell time by the number of fixations. In that sense, information about each fixation can be understood [79]. The duration of fixation was found to increase in some studies, whereas some studies show different trends, such as decreasing fixation times with various loads. By manipulating the fall speed of the tetris blocks, Mallick and colleagues found that fixation duration decreased as load increased, and the Tetris game became challenging [80]. Under high load conditions, dwell times increase while the frequency of the fixation duration decreases. The nature of individual fixation information can be understood by analyzing how each fixation is allocated during visual intake [81].

The first fixation onset, the time of the first fixation, is measured by the time interval that passes between stimulus onset and the participant's first fixation on target AOI(s). It gives insight into how quickly attention is oriented to the task or predetermined target, which can be salient or task-relevant [82]. A shorter first fixation latency indicates a faster orientation of a response, often influenced by bottom-up salience of the stimuli or the top-down attentional goals by prioritizing a specific stimulus [70, 71].

Altogether, understanding the nature of human gaze metrics allows for attention research. Within an AOI, where the first fixation happens, the number of visits and total dwell time within the AOI through the fixation are fine-grained measures for how attention is directed in a task. Moreover, attention and usage of cognitive resources can be explored by oculomotor metrics and give a stable ground for application in neuroscience, particularly in terms of how different perceptual loads affect engagement with human-robot interaction. In addition to this engagement, theories of cognition and real-life encounters of human-robot interaction can be examined by how, when, and why attention is directed or suppressed.

Figure 1.4: Image of EyeLink 1000 Plus device in our research center, UMRAM (National Magnetic Resonance Research Center, Ankara, Turkey).

## 1.7 Motivation and Hypothesis of Current Study

### 1.7.1 Main Effect of Perceptual Load

It is hypothesized that low perceptual load leads to faster reaction times, increased accuracy, and more fixations on distractor stimuli compared to high-load conditions. This aligns with perceptual load theory [15]. Since this hypothesis is grounded in previous research, it is expected to replicate similar results. In low-load trials, attention is expected to be spilled over to distractors when compared to high-load trials.

### 1.7.2 Main Effect of Agent Type

Human distractors are expected to impair task performance more than robot distractors, resulting in slower reaction times, reduced accuracy, and increased number and duration of fixations on human agents due to their enhanced social relevance [83, 84]. Previous studies have shown human faces have a tendency to attract more attention; our hypothesis was derived through these results, but it is a relatively novel hypothesis since we compare humans and robots under different perceptual loads, and our stimuli are biological motions, not just faces.

### 1.7.3 Main Effect of Action Type

Communicative actions are hypothesized to draw more attention than non-communicative ones, leading to longer reaction times and decreased accuracy. These actions are also expected to elicit more saccadic movements and prolonged dwell times due to their social and intentional salience. These hypotheses are derived from the literature [85].

### 1.7.4 Interaction between Perceptual Load and Agent Type

It is anticipated that under high perceptual load, attentional resources will be more narrowly focused, reducing distractor processing, particularly for robot agents. Human distractors may still elicit some interference due to facial and social relevance for the human faces, but overall fixation time on distractors is expected to decrease as load increases. Literature for perceptual load is expected to interact with our novel hypotheses for communicative actions.

### 1.7.5 Interaction between Perceptual Load and Action Type

Under low perceptual load, communicative distractors are expected to interfere with task performance more than non-communicative distractors, as their intentional and social content competes for limited cognitive resources. This is derived from the perceptual load theory and our novel hypothesis about the action type.

### 1.7.6 Interaction between Agent Type and Action Type

Communicative actions performed by human agents are predicted to elicit the most excellent attentional capture, resulting in the highest fixation counts. In contrast, non-communicative actions by robot agents are expected to have the least influence on task performance and oculomotor behavior. Human agent are inherently salient stimuli for the human visual system due to their biological appearance and natural movements [86,87]. This natural predisposition to human actions may contribute to the prediction that communicative actions performed by humans may elicit the most prominent capture of attention. On the other side, non-communicative actions performed by robot agents are expected to have the least effect on behavioral and oculomotor behavior since they have decreased salience when compared to human agents.

# Chapter 2

# Materials and Methods

In line with best research practices and to ensure complete transparency, our study was reviewed and approved by the Ethics Committee at İhsan Doğramacı Bilkent University. Before beginning data collection, we preregistered the experiment on the Open Science Framework, where the complete raw dataset is freely accessible [88].

## 2.1 Participants

Thirty-four healthy young adults (21 female) aged between 18 and 41 years (M =24.67, SD = 2.86) participated in the study and were all native Turkish speakers or equally proficient. All of the participants reported normal or corrected-to-normal vision and had no known history of neurological or visual impairments; individuals with color or vision impairments were excluded, though those using soft or hard contact lenses were admitted. Participants were screened for co-comorbidities and major neurological disorders via a consent form questionnaire. The number of participants was not determined by a formal power analysis but instead informed by previous eye-tracking and attention-related studies employing similar within-subject designs. To ensure robust statistical power and account for variability, we recruited a larger sample size than several comparable studies in the eye-tracking field. Informed consent was obtained and signed by all participants prior to the experiment.

## 2.2 Stimuli and Experimental Design

The experiment consists of two different kinds of stimuli: a main letter search task and peripheral distractor stimuli. The task stimuli consisted of six letters, and the structure of the experiment is detailed in 2.3.Procedure section. The distractor stimuli were either a video of a human or a video of a robot. All of the videos were selected from the HR-ACT (Human-Robot Action) Database: Communicative and Noncommunicative Action Videos Featuring a Human and a Humanoid Robot [89]. Actions consist of two types of action: communicative and noncommunicative. For communicative and noncommunicative actions, four actions for each of them were selected for the peripheral distractor video, both performed by a human and a robot, for a total of 8 videos (See Table 2.2 and 2.3 ). All of the stimuli were presented on a 1920×1080 resolution computer screen at 60 Hz. The experiment was programmed in MATLAB 2024b using Psychtoolbox extensions. Visual stimuli (main letter search task and distractors) were shown against a grey background. Peripheral video stimuli were displayed using pre-loaded .mp4 files.

Oculomotor data were recorded using the EyeLink 1000 Plus system (SR Research, Canada) at a sampling rate of 1000 Hz. Raw data were stored in the EDF (EyeLink Data File) format. Only the right eye was tracked, and analyses focused on fixation-based metrics (total dwell time, dwell time per fixation, fixation onset latency) and saccadic velocity and pupil dilation. The eye-tracking system was connected to a dedicated computer for data acquisition, while the experimental task was administered via a separate computer using MATLAB (version 2024B) [90].

Participants were seated approximately 123 cm from the display screen, with head movements minimized using a chin and forehead rest. Eye-tracking calibration and validation were conducted using a standard 3×3 point grid for calibration and validation of eye metrics. Eye movement data were recorded under both distractor-present and distractor-absent conditions. In distractor presence conditions, these stimuli depicted a human or robot performing either a communicative or noncommunicative gesture. For each trial, accuracy and response were recorded separately with gaze data. The letter search task and distractors remained on the screen for one second; participants had one second to respond by pressing a key during this one second. There was a one-second interstimulus interval between each trial.

In the low-load condition (See Figure 2.2), each trial began with a central fixation cross presented for 1000 ms. Following a 100 ms delay after the fixation onset, a circular array of letters appeared for 1000 ms. Once the fixation point changed color (black to white), participants were instructed to identify and respond to the target letter (either X or N). If no response was recorded within 1000 ms of fixation offset, the trial was automatically terminated, and the next trial commenced. After each trial, regardless of accuracy, the fixation cross reappeared in white, initiating the subsequent trial. The same procedure applies to the high-loaded condition (See Figure 2.1).

In the low-load conditions, the target letter is surrounded by the letter "O," whereas in the high-load conditions, the target letter is surrounded by letters more similar to target angular letters ( H, W, V, M, Z; see 2.1). The higher similarity of the non-target letters in the high-load task is intended to increase the perceptual load and make it affordable relative to the low-load task.

## 2.3 Data Collection

After entering the room, participants were asked to sit in front of a 23-inch monitor and place their head on the chinrest to ensure the distance of 123 cm from the screen with their head stabilized using a chinrest, is fixed for every participant and place their hand at the keyboard on the letters “X” and “N”. At the beginning of the session, the EyeLink eye-tracking system was calibrated using a standard 9-point calibration procedure. Each trial began with a fixation display (white crosshair at center), followed by a 1-second presentation of a target letter (“X” or “N”) among distractor letters at one of six fixed locations surrounding the center. The letters “X” or “N” are always presented in the main task, and there is no trial type where “X” or “N” is simultaneously absent or present. Participants were maintained in a fixation on central fixation and asked to respond as accurately and as fast as possible, whether the target letter was “X” or “N,” by pressing the corresponding key on the keyboard. Simultaneously, a 1-second video stimulus was shown on either the left or right side of the main letter search task, depending on the condition. Distractors can be present or absent in each trial in a pre-determined sequence, with a ratio of 50% of the trials consisting of any distractor called no-distractor trials. The experimental procedure comprised a total of 768 trials, with these trials systematically distributed across 10 distinct blocks (specifically, eight blocks contained 77 trials and two blocks contained 76 trials, totaling 768 trials). These distractor trials were equally distributed among the eight experimental conditions (2 Agent Type x 2 Communicativeness x 2 Perceptual Load), resulting in approximately 6-7 trials per specific distractor condition within each block.

The experiment was designed based on a within-subjects factorial design, supplemented with baseline trials. The primary design consisted of three factors: 2 (Agent: Human vs. Robot) $\times$ 2 (Communicativeness: Communicative vs. Non-communicative) $\times$ 2 (Perceptual Load: Low vs. High). In addition to these eight distractor conditions, baseline trials with no distractor were included under both low and high perceptual load to measure task performance in the absence of a social agent.

The design was fully balanced. The total number of trials and their distribution across all conditions are detailed in Table 2.1. In total, participants completed 768 trials, with each of the eight distractor conditions appearing 48 times and each of the two baseline (no distractor) conditions appearing 192 times.

## 2.4 Descriptives and Analysis

Table 2.1: Presentation frequency and percentage of all experimental conditions.

| Condition | Trials Shown (%) | N Trials |
|---|---|---|
| **No Distractor (Null)** | | |
| Low Load | 25.0% | 192 |
| High Load | 25.0% | 192 |
| **Human Distractor** | | |
| Communicative – Low Load | 6.25% | 48 |
| Communicative – High Load | 6.25% | 48 |
| Non-Communicative – Low Load | 6.25% | 48 |
| Non-Communicative – High Load | 6.25% | 48 |
| **Robot Distractor** | | |
| Communicative – Low Load | 6.25% | 48 |
| Communicative – High Load | 6.25% | 48 |
| Non-Communicative – Low Load | 6.25% | 48 |
| Non-Communicative – High Load | 6.25% | 48 |
| **Total** | **100%** | **768** |

The experiment followed a 2 (Agent: Human vs. Robot) $\times$ 2 (Communicativeness: Communicative vs. Noncommunicative) $\times$ 2 (Load: Easy vs. Hard) within-subjects design. Across 10 blocks, each block consisted of 77 trials. Participants completed trials from all eight experimental conditions in a distraction presence task in a randomized order. Breaks were given between blocks. Each block lasted approximately 4.5 minutes, and participants had to press the space button when they were ready to start the next block. Gaze data were recorded via the EyeLink host computer, starting from the stimulus onset to the key press entry from the participants. The fixation and saccade-related oculomotor activities were extracted from this gaze data. Total dwell time, dwell time per fixation, first fixation onset, and saccadic velocity, pupil dilation were analyzed separately in this 2 x 2 x 2 ANOVA design.

Raw eye-tracking data were recorded at 1000 Hz with an EyeLink 1000 Plus (SR Research, Canada). We defined Areas of Interest (AOIs) for the central search array and the peripheral agent video areas. From these AOIs, the fixations

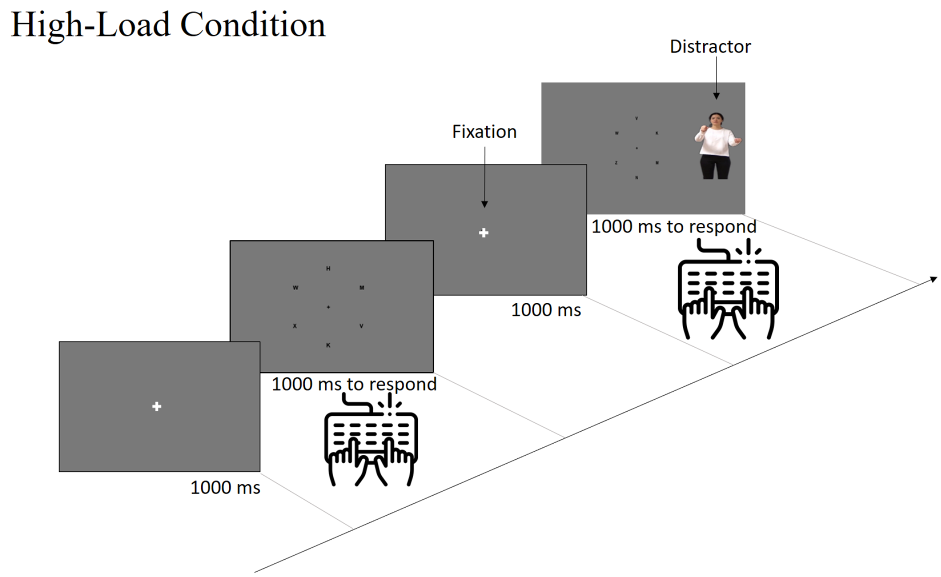



Figure 2.1: Timeline of the experimental setup. The bottom arrow shows time, and the chronology of the high-load blocks is displayed. At the beginning of each trial, a white fixation point is presented for 1000 ms, and a target stimulus is presented in one of the six pre-determined places around the black fixation point. The participant has 1000 ms to detect and press a key according to their detection. After a key press, the answer is recorded, and the next trial begins with a white fixation point again. If the participant did not press a key or missed the key, the trial ended in 1000 ms.

on distractors were analyzed by excluding the central main task AOI from the analyses. We extracted five different variables for eye metrics: total dwell time, dwell time per fixation, saccadic velocity, first fixation onset, pupil dilation, and two behavioral metrics: reaction time and accuracy.

For trials including a distractor, we adopted a 2 (Agent: Human vs. Pepper robot) × 2 (Perceptual Load: Low vs. High) × 2 (Communicativeness: Communicative vs. Noncommunicative) within-subject repeated measures design. This yielded eight experimental conditions: Human–Low-Communicative, Human–High-Communicative, Human–Low-NonCommunicative, Human–High-NonCommunicative, Robot–Low-Communicative, Robot–High-Communicative,

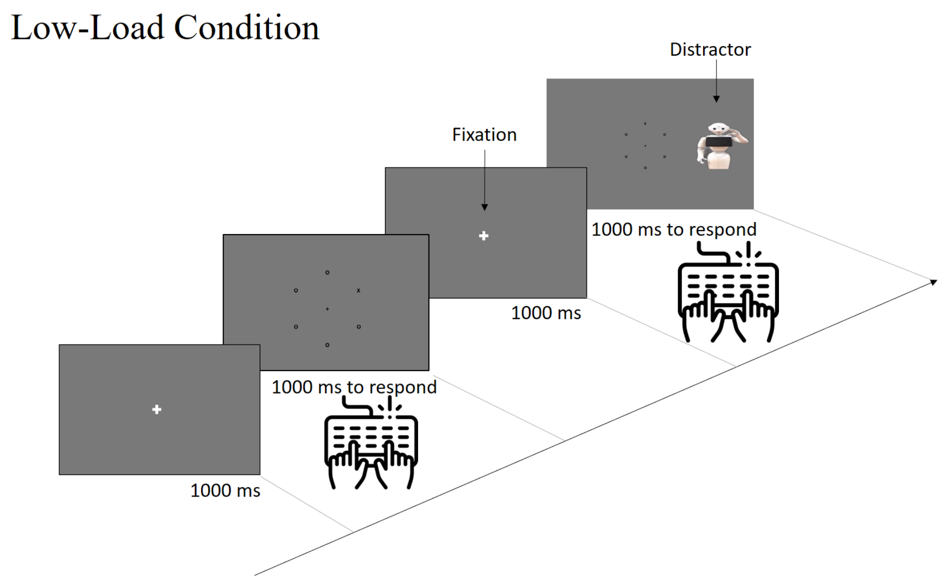
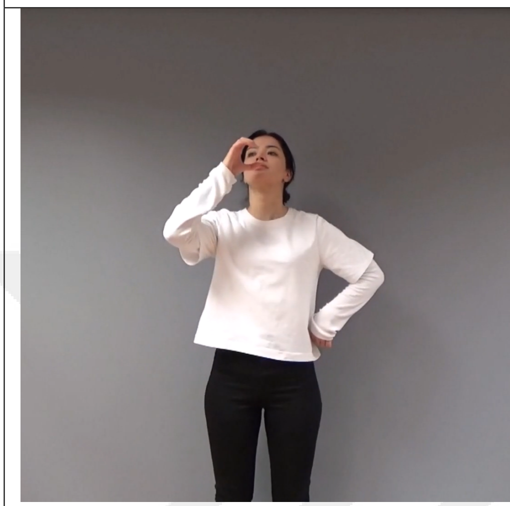
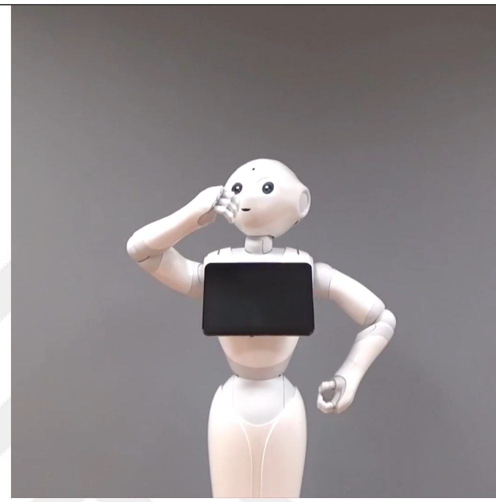
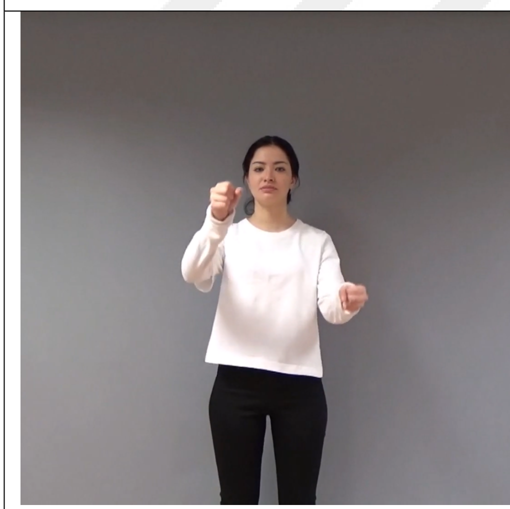
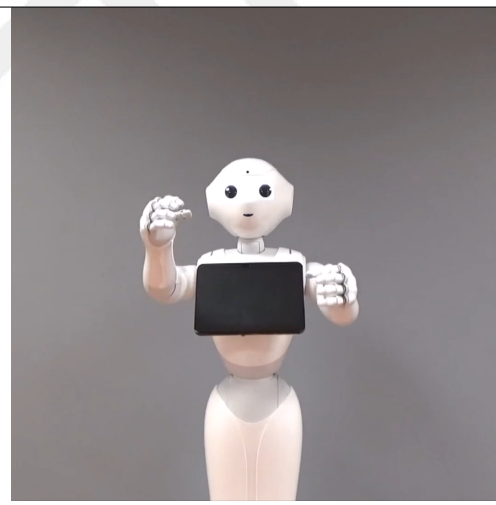
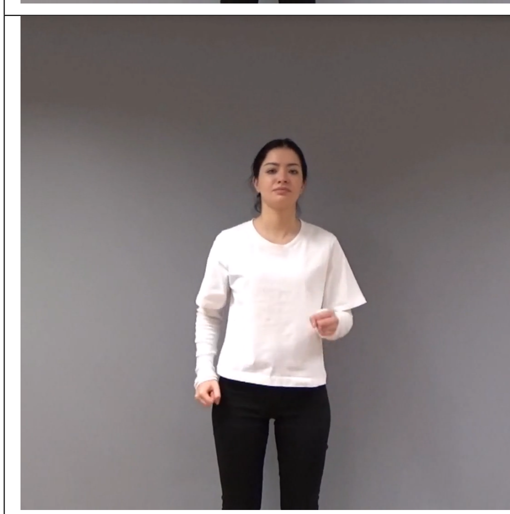
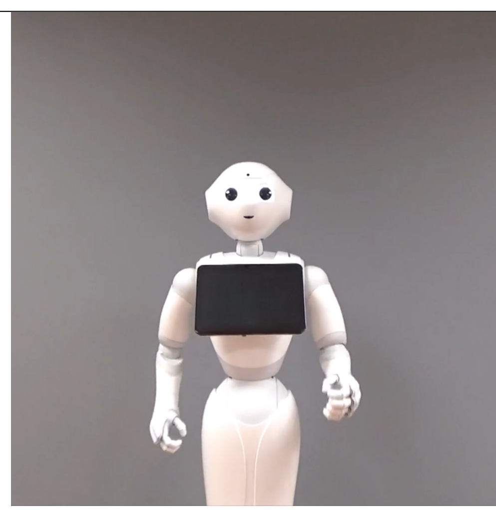



Figure 2.2: Timeline of the experimental setup. The bottom arrow shows time, and the chronology of the low-load blocks is displayed. At the beginning of each trial, a white fixation point is presented for 1000 ms, and a target stimulus is presented in one of the six pre-determined places around the black fixation point. The participant has 1000 ms to detect and press a key according to their detection. After the key press, the answer is recorded, and the next trial begins with a black fixation point again. If the participant did not press a key or missed the key, the trial ended in 1000 ms.

Robot–Low-NonCommunicative, and Robot–High-NonCommunicative. In addition to these conditions, the experiment included baseline trials where no distracting agent was present. These baseline trials were analyzed separately to assess task performance (i.e., accuracy and reaction time) as a function of perceptual load alone.

All statistical analyses and visualization were conducted in R (Version 4.5.0; R Core Team, 2025), primarily using the ez, tidyverse, rstatix, and patchwork packages. The primary analyses consisted of a series of 2 (Agent: Human vs. Robot) $\times$ 2 (Communicativeness: Communicative vs. Noncommunicative) $\times$ 2 (Perceptual Load: Low vs. High) within-subjects repeated measures ANOVAs. These were performed on several dependent variables, including total dwell time, dwell time per fixation, first fixation onset, saccadic velocity, and pupil dilation. Significant interactions were followed up with paired-samples t-tests with Bonferroni correction. Additionally, separate paired-samples t-tests were used to analyze accuracy and reaction time data from the baseline (no distractor) conditions.

Table 2.2: Human and Robot Actions Noncommunicative Stimulus Set

| Human | Robot |
|---|---|
|  |  |
|  |  |
|  |  |

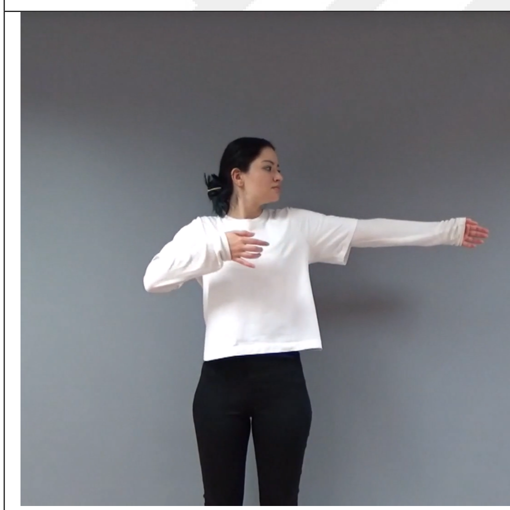
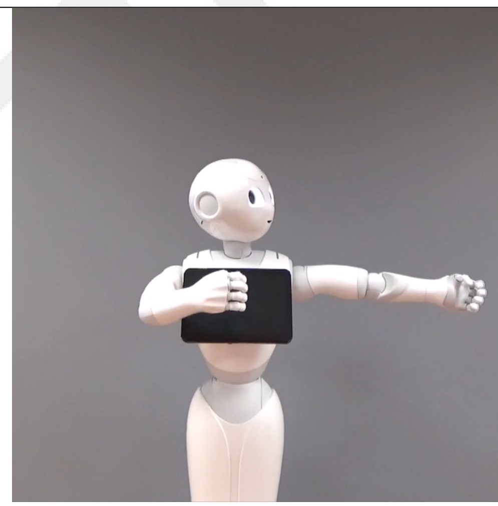
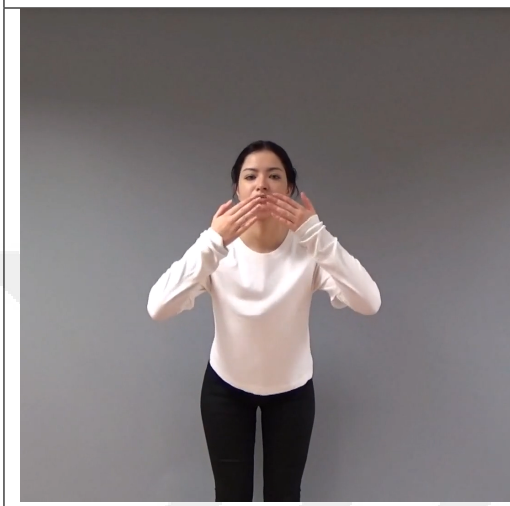
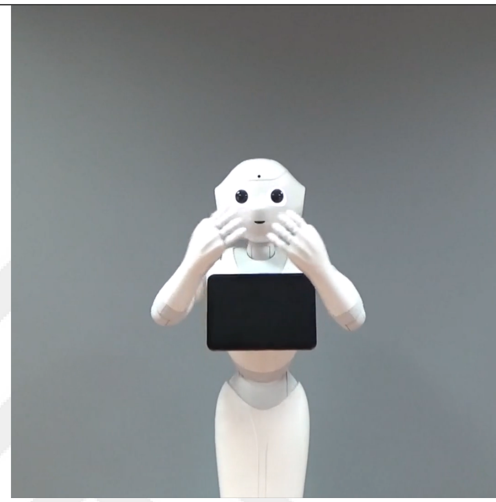
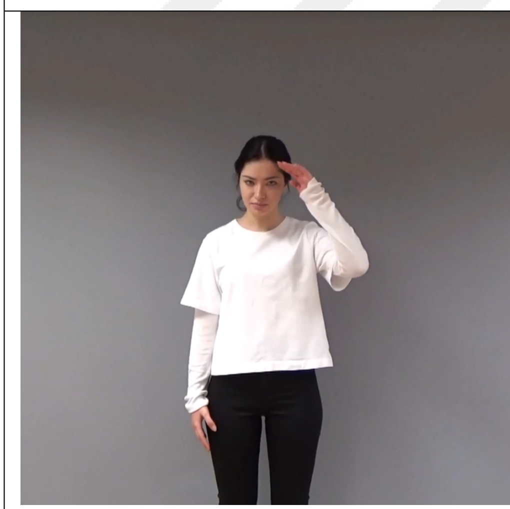
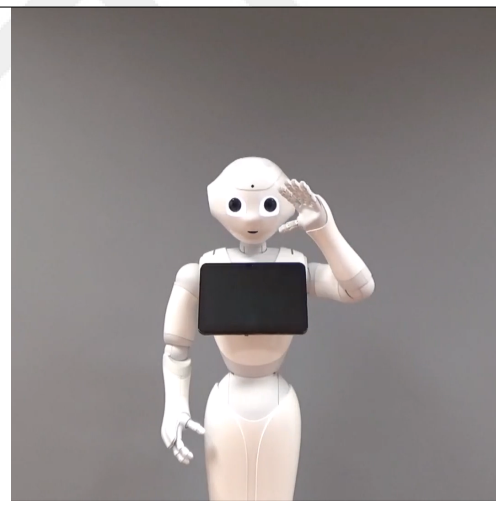
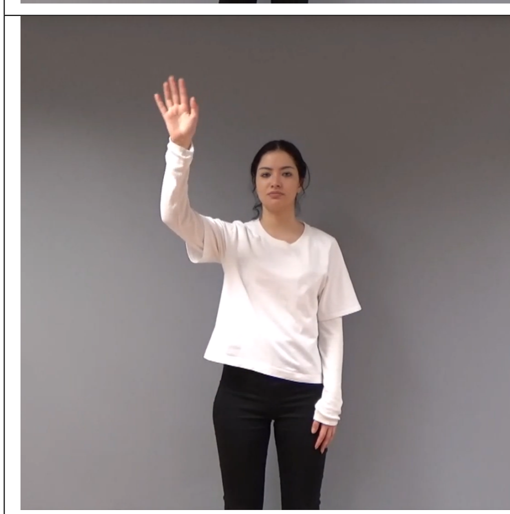
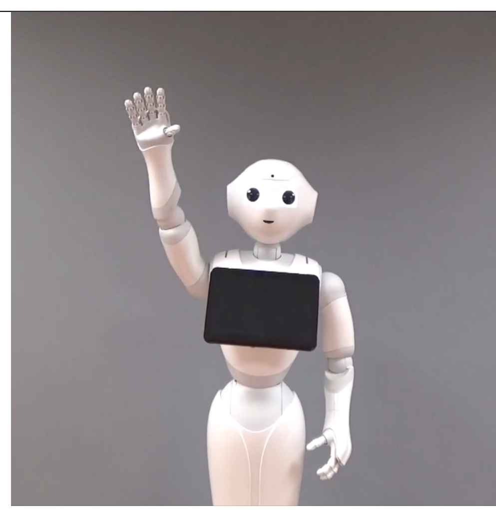
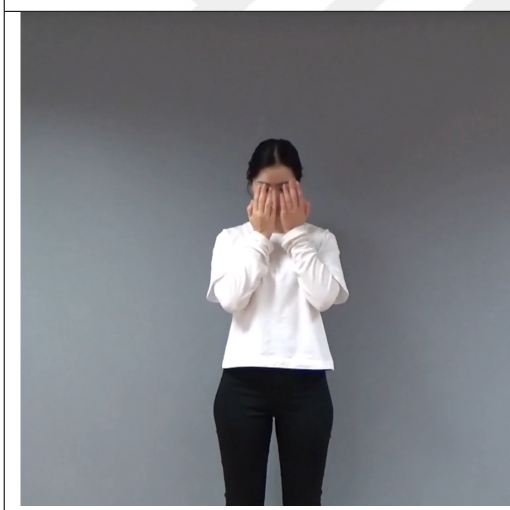
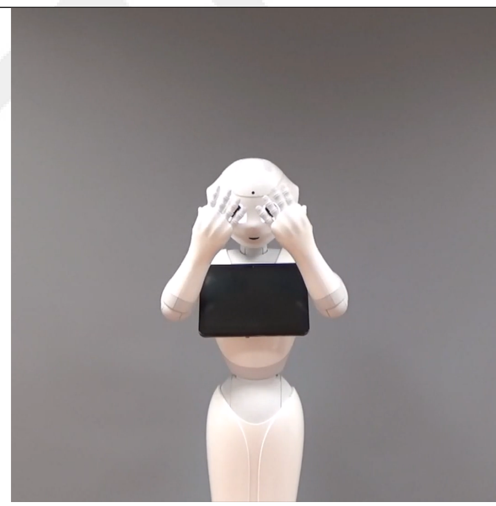
| Human | Robot |
| --- | --- |
|  |  |

Table 2.3: Human and Robot Actions Communicative Stimulus Set

| Human | Robot |
|---|---|
|  |  |
|  |  |
|  |  |

| Human | Robot |
|---|---|
|  |  |

# Chapter 3

# Results

## 3.1 Fixation Counts

In distractor presence trials, the total number of fixations and the mean number of fixations per subject were calculated for all eight experimental conditions. Summary of these statistics provided in the Table 3.3.

The highest number of fixations per subject was observed in the Human-Communicative-Hard conditions (M = 9.28, SD = 14.44), whereas the lowest mean scores were observed in Human-Communicative-Easy conditions (M = 6.69, SD = 10.80). Descriptively, high-load conditions have a tendency to elicit more fixations on distractors when compared to low-load conditions.

Descriptively, the highest mean number of fixations per subject was observed in the Human-Communicative-Hard condition (M = 9.28, SD = 14.44), while the lowest mean was observed in the Human-Communicative-Easy condition (M = 6.69, SD = 10.80). A general trend in the data suggests that high-load conditions tended to elicit more fixations on the distractor compared to low-load conditions.

Descriptive analysis of fixation distribution across all trials was conducted to understand the overall allocation of visual attention (see Table 3.2). The results

Table 3.1: Total and Mean Fixation Counts Per Subject for Each Condition

| Condition | Total Fixations | Mean Fixations | SD |
|---|---|---|---|
| Human-Communicative-Hard | 269 | 9.28 | 14.44 |
| Human-Communicative-Easy | 194 | 6.69 | 10.80 |
| Human-Non-communicative-Hard | 241 | 8.31 | 12.42 |
| Human-Non-communicative-Easy | 239 | 8.24 | 12.48 |
| Robot-Communicative-Hard | 213 | 7.34 | 10.96 |
| Robot-Communicative-Easy | 223 | 7.69 | 12.33 |
| Robot-Non-communicative-Hard | 266 | 9.17 | 13.22 |
| Robot-Non-communicative-Easy | 203 | 7.00 | 11.51 |

show that participants directed the majority of their visual attention to the primary task. On average, 74.26% (SD = 14.52%) of all fixations were made on the main letter search task, while the remaining 25.74% (SD = 14.52%) were captured by the peripheral distractors. This gap between the distributions indicates participants primarily engage in the main letter search task; however, distractors still can capture a substantial amount of overt visual attention in the experiment.

Table 3.2: Means of fixation numbers and their percentage scores across trials

| **Area of Interest** | **Mean Fixation (Sd)** | **Mean Percentage (SD)** |
|---|---|---|
| Main Letter Search Task | 1386.78 (438.79) | 74.26% (14.52%) |
| Distractors | 503.06 (335.54) | 25.74% (14.52%) |

## 3.2 Behavioral Results

### 3.2.1 Accuracy

As expected, the perceptual load manipulation strongly affected accuracy on trials without a distractor. Participants were significantly more accurate in the Low-Load condition (M = 0.95, SD = 0.03) compared to the High-Load condition (M = 0.72, SD = 0.11). A paired-samples t-test confirmed this difference was highly significant and represented a very large effect, $t(33) = 11.3$, $p < .001$, Cohen's $d = 1.93$.

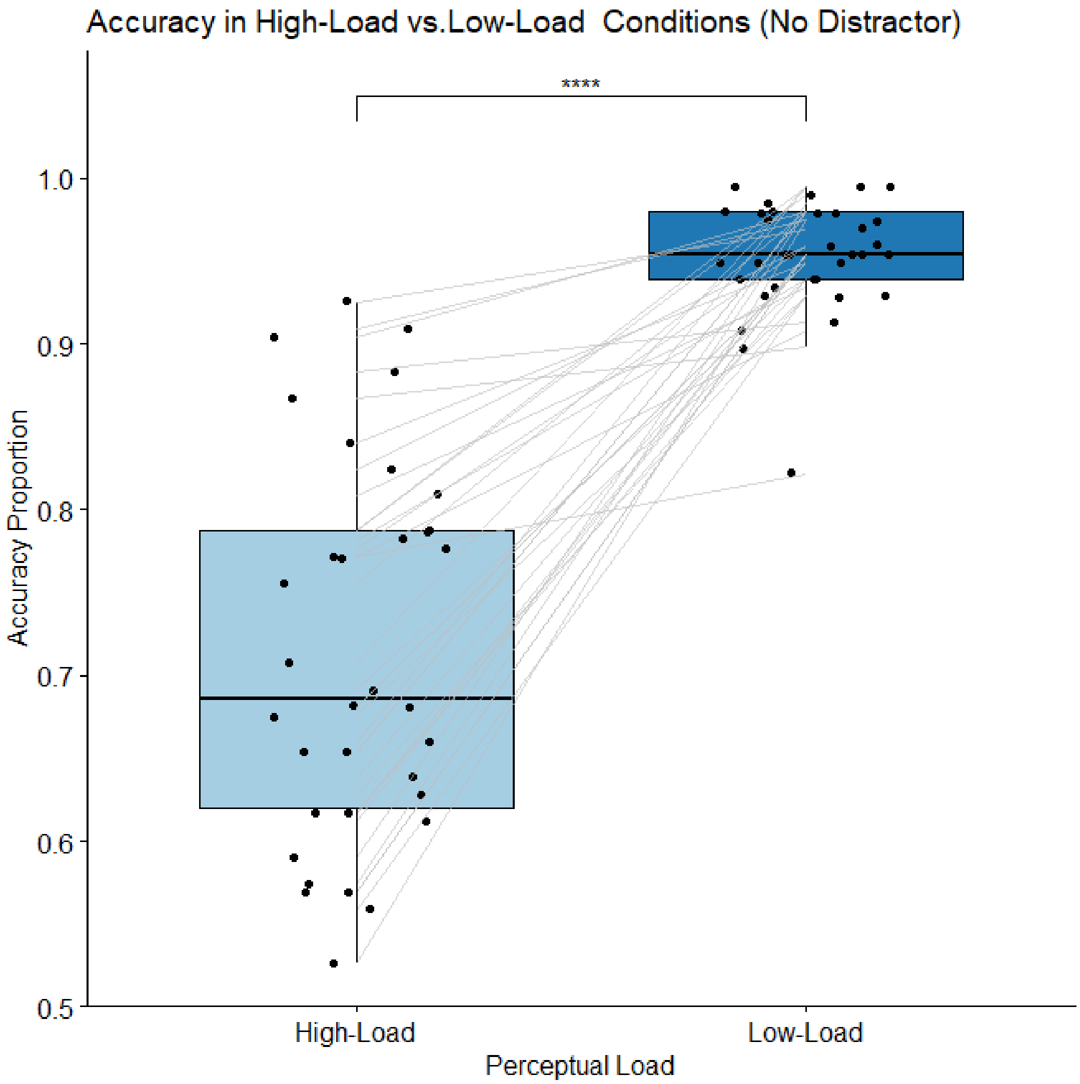



Figure 3.1: Accuracy in no-distractor trials.

### 3.2.2 Effects of Distractors on Accuracy

Next, to understand the effects of the distractor's social properties on accuracy, a 2 (Agent: Human vs. Robot) × 2 (Communicativeness: Communicative vs. Non-communicative) × 2 (Perceptual Load: Low vs. High) within-subjects repeated measures ANOVA was conducted.

There is a significant main effect of Perceptual Load, $F(1, 33) = 41.69$, $p <$

.001, where accuracy was higher in low-load trials. The main effects for Agent (p = .646) and Communicativeness (p = .124) were not significant.

The three-way interaction was not significant, F(1, 33) = 0.09, p = .772. However, the analysis revealed a significant two-way interaction between Communicativeness and Perceptual Load, F(1, 32) = 20.07, p < .001. This interaction indicates that while accuracy was higher in the low-load condition for both action types, this performance drop in the high-load condition was significantly bigger for communicative actions compared to non-communicative actions.

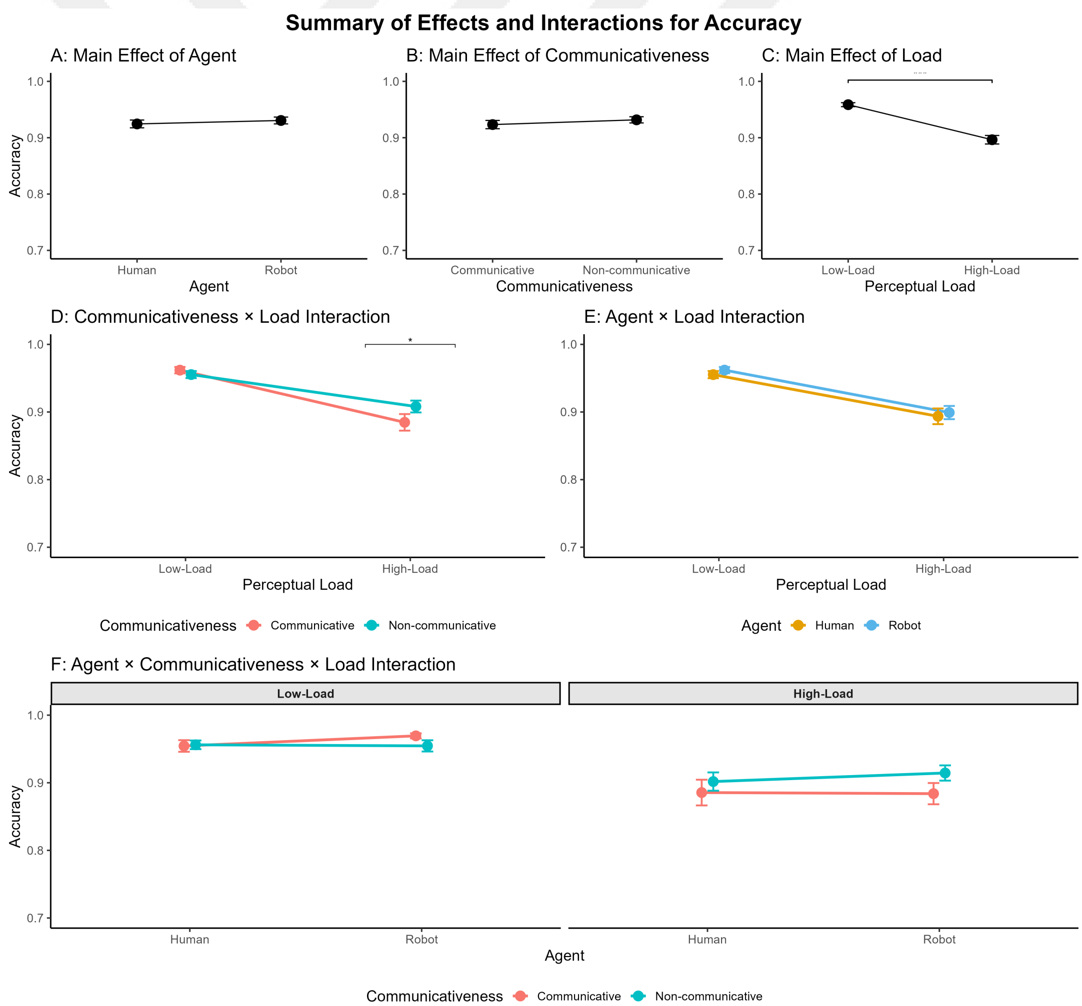



Figure 3.2: Distractor Conditions on Accuracy Scores

### 3.2.3 Reaction Time

A paired-samples t-test was conducted to compare mean reaction times between the Low-Load and High-Load conditions. The analysis revealed a statistically significant difference, with participants responding significantly slower in the High-Load condition (M = 0.73, SD = 0.05) than in the Low-Load condition (M = 0.53, SD = 0.06), $t(34) = -32.7$, $p < .001$. This finding indicates that an increase in perceptual load led to a significant increase in reaction time.

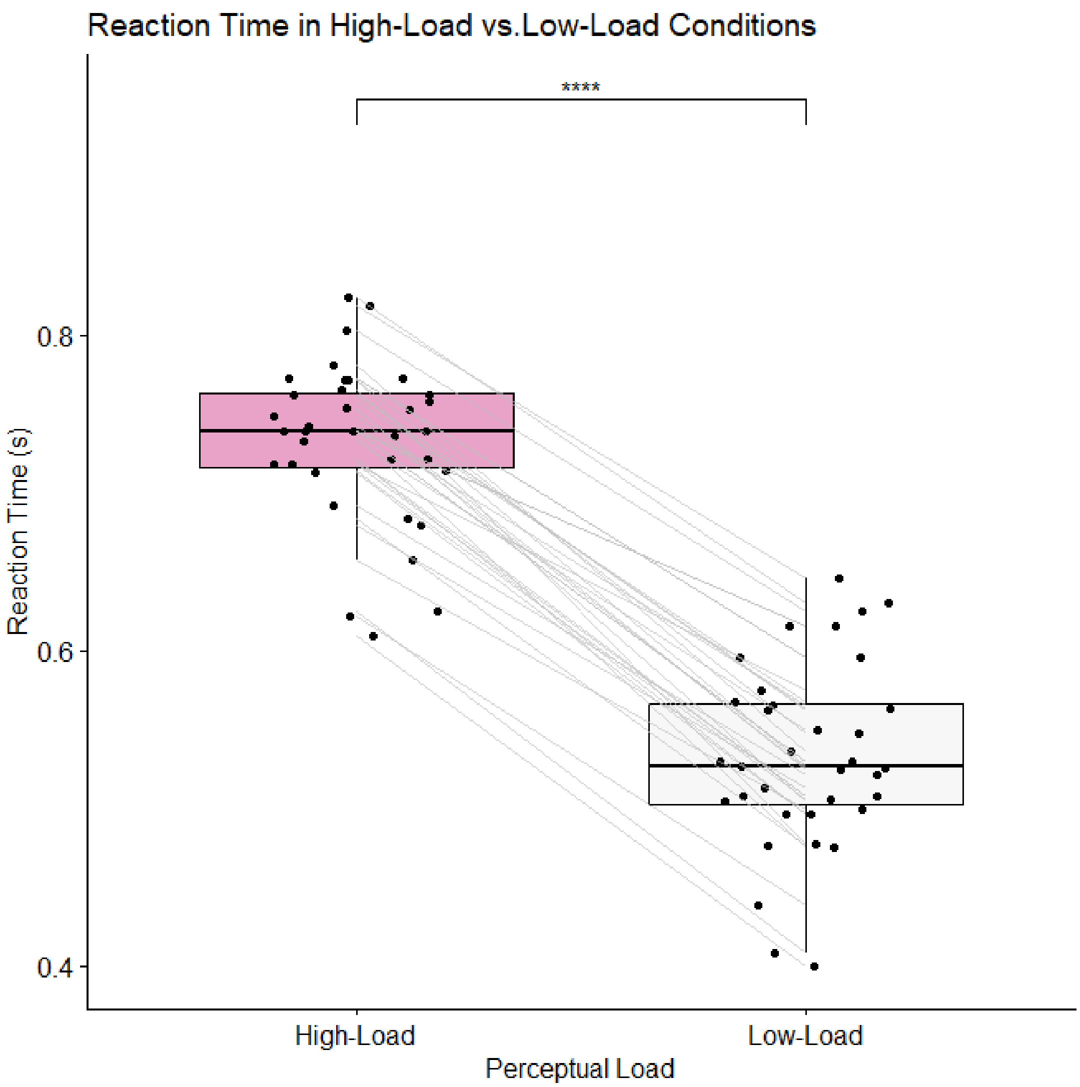



Figure 3.3: Reaction time in no-distractor trials.

### 3.2.4 Effects of Distractors on Reaction Time

Next, to understand how the presence of a distractor modulated reaction times, a 2 (Agent: Human vs. Robot) $\times$ 2 (Communicativeness: Communicative vs. Non-communicative) $\times$ 2 (Perceptual Load: Low vs. High) within-subjects repeated measures ANOVA was conducted.

There is two significant main effects. A large main effect of Perceptual Load was found, $F(1, 33) = 787.38$, $p < .001$, confirming that reaction times were significantly longer in high-load conditions. There was also a significant main effect of Agent, $F(1, 33) = 6.33$, $p = .017$. The main effect of Communicativeness was not significant ($p = .135$).

The analysis revealed that the three-way interaction was not statistically significant ($p = .482$). However, two significant two-way interactions were found. There was a significant interaction between Agent and Perceptual Load, $F(1, 33) = 11.38$, $p = .002$, indicating that the effect of load on reaction times differed between the human and robot agents. There was also a significant interaction between Agent and Communicativeness, $F(1, 33) = 4.29$, $p = .047$, suggesting that the effect of the action's communicativeness on reaction times depended on the agent's identity.

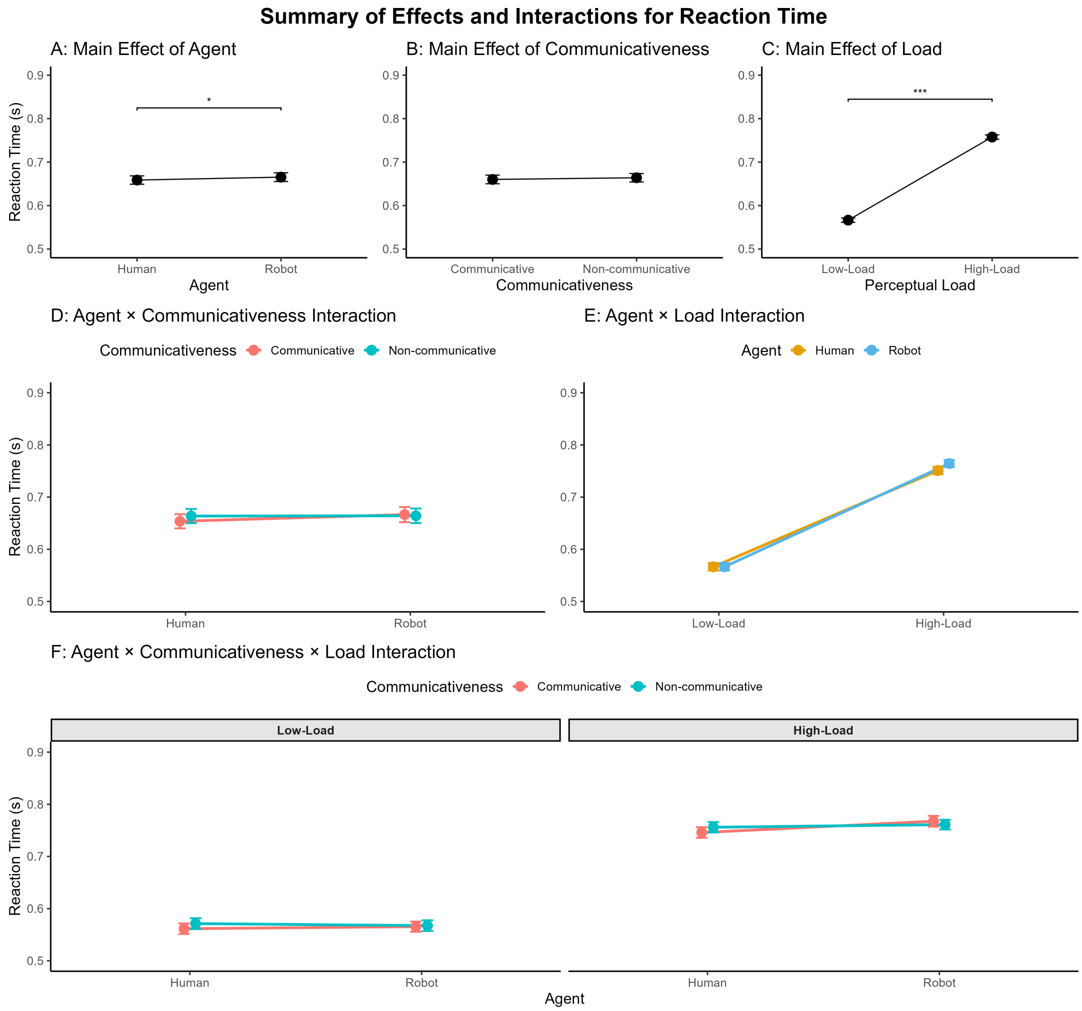



Figure 3.4: Distractor conditions on Reaction Times

## 3.3 Eye-Tracking Results

### 3.3.1 Fixation Counts

The total number of fixations directed at the distractor and the mean number of fixations per subject were calculated for each of the eight experimental conditions. Table 3.3 provides a summary of these descriptive statistics.

Descriptively, the highest mean number of fixations per subject was observed in the Human-Communicative-Hard condition (M = 9.28, SD = 14.44), while the lowest mean was observed in the Human-Communicative-Easy condition (M = 6.69, SD = 10.80). A general trend in the data suggests that high-load conditions tended to elicit more fixations on the distractor compared to low-load conditions. To statistically test these observations, a 2 (Agent) × 2 (Communicativeness) × 2 (Perceptual Load) repeated measures ANOVA was performed on the fixation count data.

Table 3.3: Total and Mean Fixation Counts Per Subject for Each Condition

| Condition | Total Fixations | Mean Fixations | SD |
|---|---|---|---|
| Human-Communicative-Hard | 269 | 9.28 | 14.44 |
| Human-Communicative-Easy | 194 | 6.69 | 10.80 |
| Human-Non-communicative-Hard | 241 | 8.31 | 12.42 |
| Human-Non-communicative-Easy | 239 | 8.24 | 12.48 |
| Robot-Communicative-Hard | 213 | 7.34 | 10.96 |
| Robot-Communicative-Easy | 223 | 7.69 | 12.33 |
| Robot-Non-communicative-Hard | 266 | 9.17 | 13.22 |
| Robot-Non-communicative-Easy | 203 | 7.00 | 11.51 |

### 3.3.2 Total Dwell Time

To analyze participants' total dwell time on the distractors, a 2 (Agent: Human vs. Robot) × 2 (Communicativeness: Communicative vs. Non-Communicative) × 2 (Perceptual Load: Low vs. High) within-subjects repeated measures ANOVA was conducted.

A significant main effect of Perceptual Load was found, $F(1, 33) = 21.39$, $p < .001$, indicating that, overall, participants dwelled longer on distractors during high-load trials compared to low-load trials. The main effects for Agent ($p = .212$) and Communicativeness ($p = .764$) were not statistically significant.

Furthermore, the ANOVA yielded a significant two-way interaction between Agent and Communicativeness, $F(1, 33) = 16.38$, $p < .001$, which is qualified by the superordinate three-way interaction described above (See Graph 3.5 Panel D.). The other two-way interactions were not significant (Agent $\times$ Perceptual Load, $p = .235$; Communicativeness $\times$ Perceptual Load, $p = .498$)

Finally, the analysis revealed a statistically significant three-way interaction between Agent, Communicativeness, and Perceptual Load, $F(1, 33) = 12.10$, $p = .002$. To interpret this complex interaction, the simple two-way interactions at each level of Perceptual Load were examined. Under Low Load, there was no significant interaction between Agent and Communicativeness. However, under High Load, an interaction emerged: participants' dwell time was significantly longer for communicative actions from the human agent compared to non-communicative actions. Conversely, for the robot agent, this pattern was reversed, with communicative actions eliciting shorter dwell times than non-communicative actions (See Graph 3.5 Panel E.).

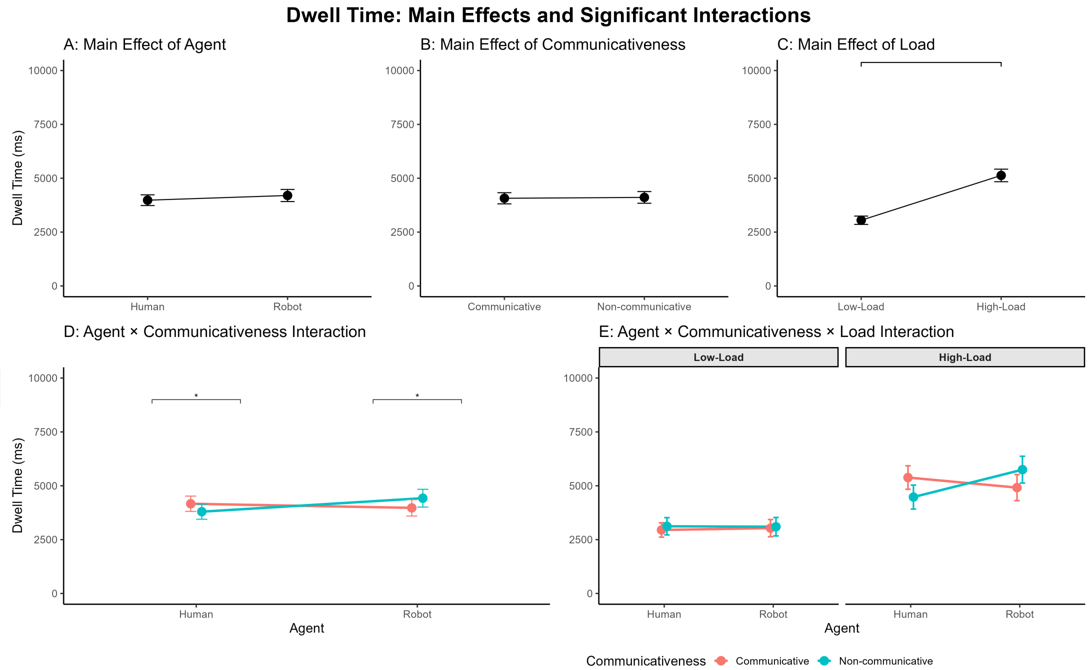



Figure 3.5: Panel A. Main effects of Agent, Panel B. Communicativeness and Panel C. Perceptual Load, Panel D. Significant Interaction Between Agent x Communication, Panel E. Significant Interaction Between Agent x Communicativeness X Perceptual Load

### 3.3.3 Dwell Time Per Fixation

To analyze the mean dwell time per fixation on distractors, a 2 (Agent: Human vs. Robot) × 2 (Communicativeness: Communicative vs. Non-communicative) × 2 (Perceptual Load: Low vs. High) repeated measures analysis of variance (ANOVA) was conducted.

The analysis did not reveal a significant main effect for Agent, $F(1, 33) = 1.01$, $p = .324$, Communicativeness, $F(1, 33) = 0.38$, $p = .540$, or Perceptual Load, $F(1, 33) = 1.51$, $p = .230$.

Furthermore, no significant interactions were found. The three-way interaction between Agent, Communicativeness, and Load was not significant, $F(1, 33) =$

0.94, p = .340. None of the two-way interactions reached significance: Agent × Communicativeness, $F(1, 33) = 0.60$, $p = .445$; Agent × Load, $F(1, 33) = 2.03$, $p = .166$; and Communicativeness × Load, $F(1, 33) = 0.58$, $p = .453$.

In summary, these findings indicate that the mean duration of a single fixation on a distractor was not significantly influenced by the agent's type, communicativeness, the level of perceptual load, or any combination of these factors.

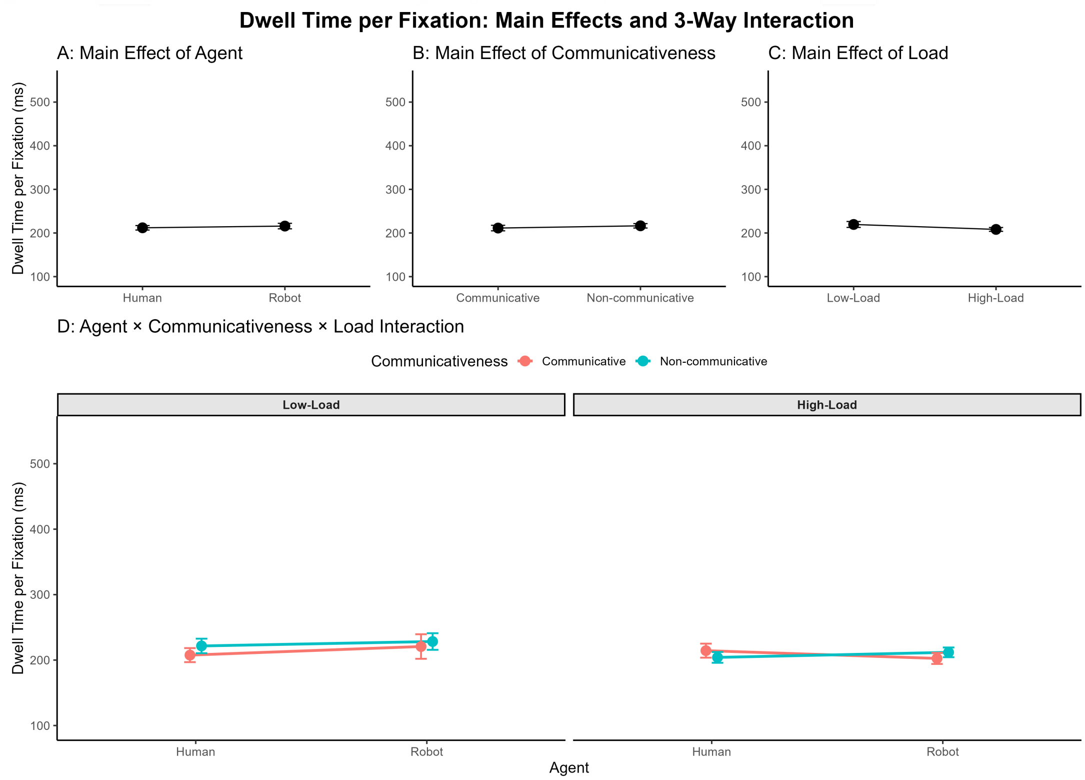



Figure 3.6: Dwell Time Per Fixation: Panel A. Main effects of Agent, Panel B. Communicativeness and Panel C. Perceptual Load, Panel D. Agent x Communicativeness X Perceptual Load

### 3.3.4 First Fixation Latency

To analyze the first fixation onset time on the distractors, a 2 (Agent: Human vs. Robot) × 2 (Communicativeness: Communicative vs. Non-communicative)

$\times$ 2 (Perceptual Load: Low vs. High) repeated measures ANOVA was conducted.

The ANOVA results yielded a significant main effect of Perceptual Load, $F(1, 33) = 144.15$, $p < .001$. This indicates that, overall, participants were slower to fixate on the distractor during high-load trials compared to low-load trials. The main effects for Agent, $F(1, 32) = 2.77$, $p = .106$, and Communicativeness, $F(1, 33) = 2.50$, $p = .124$, were not statistically significant.

Furthermore, the analysis revealed that the three-way interaction was not significant, $F(1, 33) = 0.01$, $p = .978$. However, there was a significant two-way interaction between Communicativeness and Perceptual Load, $F(1, 33) = 13.27$, $p < .001$. Simple effects analysis was performed to break down this interaction. Under Low Load, there was no significant difference in fixation onset time between communicative and non-communicative actions. In contrast, under High Load, communicative actions led to significantly faster fixation onsets compared to non-communicative actions. This suggests that the attention-capturing advantage of a communicative action only emerges under high cognitive demand. The other two-way interactions were not significant.

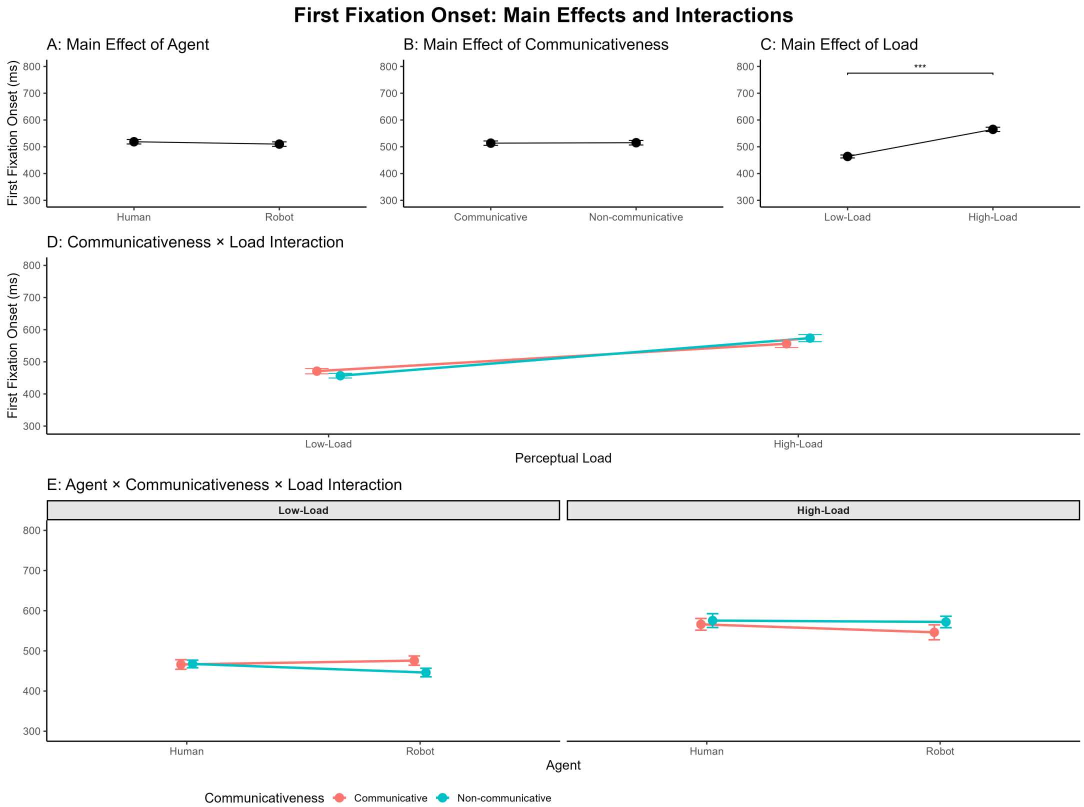



Figure 3.7: First Fixation Latency: Panel A. Main effects of Agent, Panel B. Communicativeness and Panel C. Perceptual Load, Panel D. Significant Interaction Between Communicativeness X Perceptual Load, Panel E. Agent x Communicativeness X Perceptual Load

### 3.3.5 Saccadic Velocity

A 2 (Agent: Human vs. Robot) × 2 (Communicativeness: Communicative vs. Non-communicative) × 2 (Perceptual Load: Low vs. High) repeated measures ANOVA was conducted on mean saccadic velocity to the distractors.

There was a significant main effect of Agent, $F(1, 33) = 5.53$, $p = .029$, indicating that saccadic velocities differed depending on whether the agent was a human or a robot. Additionally, a highly significant main effect of Perceptual Load was observed, $F(1, 33) = 26.99$, $p < .001$, showing that saccade speeds were modulated by task difficulty. The main effect of Communicativeness was not significant, p = .329. These findings suggest that saccadic velocity is independently influenced by the agent's type and the ongoing perceptual load.

Furthermore, none of the two-way or three-way interactions reached statistical significance (all ps > .05), including the Agent × Perceptual Load interaction, which was marginally non-significant, F(1, 33) = 4.15, p = .054.

In summary, these results indicate that the mean saccadic velocity was not significantly influenced by the agent's type, the action's communicativeness, the level of perceptual load, or any combination of these factors.

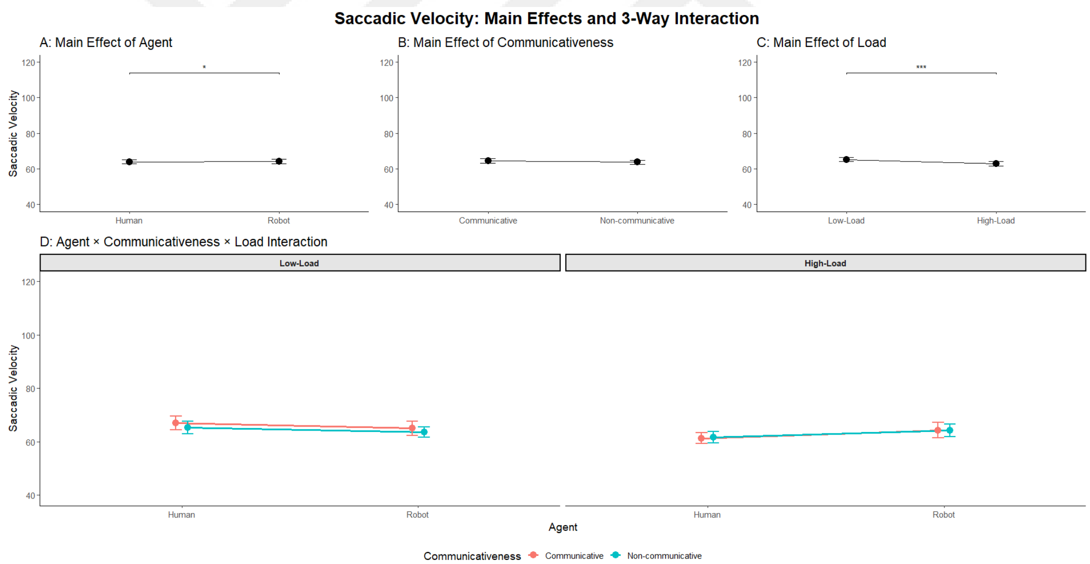



Figure 3.8: Saccadic Velocity on Perceptual Load : Panel A. Main effects of Agent, Panel B. Communicativeness and Panel C. Perceptual Load, Panel D. Agent x Communicativeness X Perceptual Load

### 3.3.6 Pupilometry

To assess cognitive effort, a 2 (Agent: Human vs. Robot) × 2 (Communicativeness: Communicative vs. Non-communicative) × 2 (Perceptual Load: Low vs. High) repeated measures ANOVA was conducted on mean pupil dilation.

The ANOVA yielded a highly significant main effect of Perceptual Load, F(1, 33) = 37.54, $p < .001$, indicating that pupil dilation was, as expected, greater in high-load conditions compared to low-load conditions. The main effects for Agent (p = .170) and Communicativeness (p = .571) were not statistically significant. Furthermore, the interaction between Agent and Communicativeness was marginally non-significant, F(1, 33) = 3.54, p = .069. The other two-way interactions were not significant.

The analysis did not reveal a significant three-way interaction (p = .323). However, a significant two-way interaction between Communicativeness and Perceptual Load was found, F(1, 33) = 4.41, p = .044. Post-hoc comparisons to explore this interaction revealed that under Low Load, there was no significant difference in pupil dilation between communicative and non-communicative actions. In contrast, under High Load, non-communicative actions elicited significantly greater pupil dilation than communicative actions. This suggests that processing simple, non-social cues required more cognitive effort than processing socially meaningful actions when the main task was difficult.

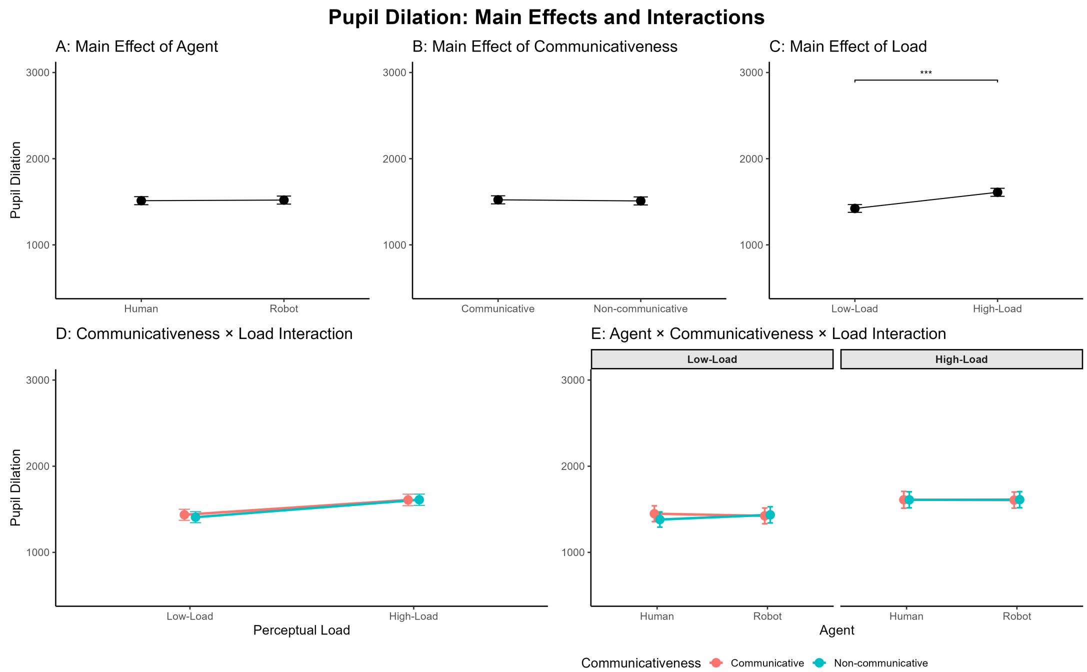



Figure 3.9: Pupil Dilation: Panel A. Main effects of Agent, Panel B. Communicativeness and Panel C. Perceptual Load, Panel D. Significant Interaction Between Communicativeness X Perceptual Load, Panel E. Agent x Communicativeness X Perceptual Load

### 3.3.7 Summary

To examine how the interaction between agent type (Human vs. Robot), communicativeness (Communicative vs. Non-communicative), and perceptual load (High vs. Low) influences visual attention and cognitive processing, we analyzed five key eye-tracking and behavioral metrics: dwell time per fixation, first fixation onset, pupil dilation, saccadic velocity, and total dwell time. Across all metrics, a three-way interaction (Agent $\times$ Communicativeness $\times$ Load) revealed nuanced patterns (See Figure 3.10)

For first fixation latency, participants fixated faster on communicative human distractors under low-load conditions, while under high-load conditions, onset times were relatively similar across conditions. When cognitive resources are available, communicative actions take more sudden attention. Under both load conditions, communicative distractors tended to elicit larger pupil dilation, particularly for human agents in high-load conditions. This pattern suggests greater cognitive load or arousal in response to communicative behaviors, especially when attention is already taxed. Saccadic velocity was generally higher in robot distractor conditions, especially under high-load and for non-communicative stimuli. This may reflect a more effortful or disengaged visual search pattern in these cases.Communicative actions lead to significantly higher total dwell times in high load conditions for the robot agent. These findings suggest that, when the task is demanding, communicative actions performed by the robot can hold visual attention longer. Altogether, these results indicate that the agent type, communicative intent, and perceptual load modulate attention. Communicative behaviors have a stronger effect when cognitive resources are constrained, and robot distractors draw prolonged attention under such conditions.

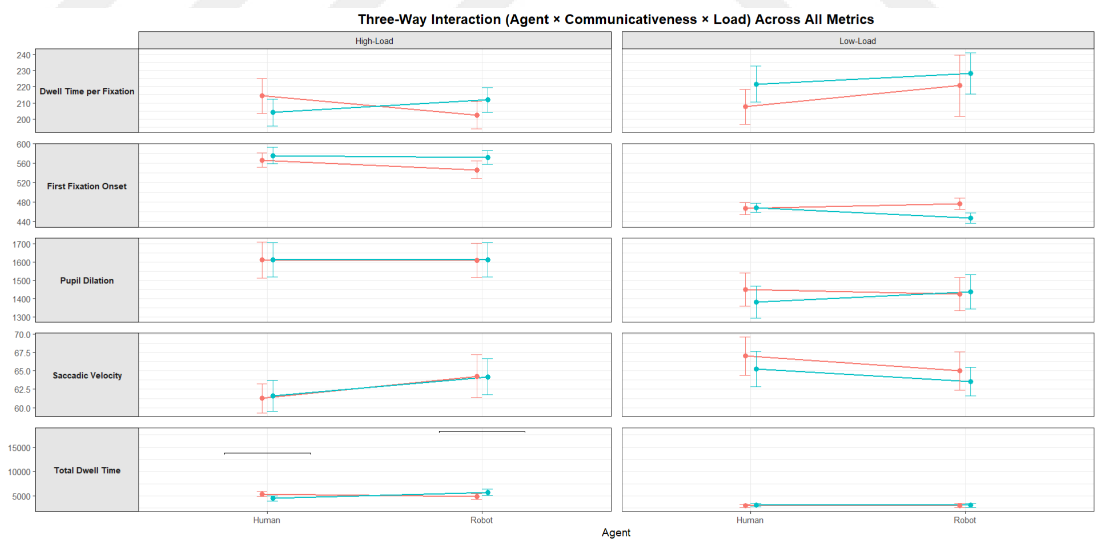



Figure 3.10: The Agent x Communicativeness X Perceptual Load Interactions on Total Dwell Time per Fixation, First Fixation Onset, Pupil Dilation, Saccadic Velocity, and Total Dwell Time can be observed.

# Chapter 4

# Discussion

This study aimed to examine how different types of agents (human vs. robot), the communicative nature of actions (communicative vs. noncommunicative), and perceptual load (low vs. high) influence visual attention using central letter search tasks and peripheral distractor videos. Perceptual load theory is a grounded, well-established theoretical framework used in attention literature, and here we examined how socially meaningful distractors, such as humans and robots, across different actions interfere with attention by depending on the different cognitive demands of the primary task. Nuanced contributions to the interaction between perceptual load and socially relevant distractions were found to have several key implications. Demonstration of the association between different processes of attention revealed nuanced contributions to the interaction between socially meaningful distractions and perceptual load.

First, the initial orienting of attention, which is measured by first fixation latency, is modulated by the demands of the tasks rather than the agent. The significant interaction effect of communicativeness and perceptual load revealed that communicative cues capture attention faster than non-communicative ones if only the main task is hard (high-load). This can be explained by the fact that the brain may prioritize efficiency by using socially meaningful stimuli as a way to quickly analyze the peripheral distraction. The agent's identity did not influence this initial eye movement, but the communicative nature of an action and the difficulty of a man's task indicate the deployment of attention. Thus, the initial orienting did not influence the agent's identity; the initial deployment of the attention allocation is strongly modulated by task demands rather than social meaning.

The more precise information during the gaze does not fully reflect the true nature of eye movements, as the difference between each fixation can be huge; the dwell time per fixation can be a good indication when it's used in free-viewing studies. But since here, the time constraints are fixed, it may fluctuate and not be a reliable source for the studies that expect a rapid response from the participants. Dwell per fixation did not reveal a significant effect, indicating that the mean duration of a single fixation is not affected by communicativeness or task difficulty. In a similar manner, saccadic velocity is independently influenced by agent type and perceptual load, and complex interactions cannot be found, which can be an indication of the straightforward effect of oculomotor mechanisms overriding higher-order cognitive processes. By taking these results together, the single eye movement parameters are susceptible to complex manipulation in this study but they are robust in less complex metrics.

The analysis of Dwell Time per Fixation yielded no significant effects, suggesting that the duration of a single glance is a stable, low-level mechanism not modulated by the social context or task difficulty. Taken together, these results suggest that while the parameters of a single eye movement can be influenced by basic factors, they are robust against the more complex social manipulations in this study, in stark contrast to metrics of overall engagement.

Total dwell time and pupil dilation were used to investigate how attentional processes and task performance are key to understanding how higher social cognition becomes evident in this study. The statistically significant three-way interaction for total dwell time is the study's most interesting finding. Given that the duration of each fixation is stable, participants spend less time on the robot performing communicative action, suggesting that a robot's social inclusion under high cognitive load may be confusing for participants, leading to attentional disengagement. The pupil dilation result helps us to understand these processes better. Pupils are more dilated in non-communicative actions under high load than in communicative actions. This intriguing finding may imply that socially meaningful communicative action might be processed more efficiently than the ambiguous, almost like mimicking actions like drinking or driving. Because in the communicative actions, the action is performed for the participants, whereas in the non-communicative actions, actions mimic (pantomime) the original actions.

## 4.1 Comparison of Hypotheses and Results

H1: High load leads to slower RT, lower accuracy, and more fixations on distractors. As predicted, Perceptual Load showed a strong and significant main effect on both behavioral measures (RT, Accuracy) and nearly all eye-tracking metrics (Total Dwell Time, Saccadic Velocity, First Fixation Latency, Pupil Dilation).

H2: Human distractors create a greater distractor effect than robots (slower RT, lower accuracy, more/longer fixations). This hypothesis was largely not supported as a main effect. A significant main effect of Agent was found only for Reaction Time and Saccadic Velocity. However, there was no main effect on the critical metric of Total Dwell Time, suggesting the agent's influence primarily emerged through interactions with other factors.

H3: Communicative actions attract more attention than non-communicative ones. The Communicativeness factor did not have a significant main effect on any metric. Its influence was only observed in interaction with other factors.

H4: High load reduces distractors processing, especially for robot agents. This interaction was significant for Reaction Time only. It was not observed in the eye-tracking metrics, indicating the hypothesis was only partially confirmed at the behavioral level.

H5: Under low load, communicative distractors have a greater effect. This interaction was significant for First Fixation Latency, Pupil Dilation, and Accuracy. However, the pattern was the opposite of what was hypothesized: the distinct effect of communicative cues emerged when the task was difficult (high load), not when it was easy. This is one of the most interesting findings of the study, with a contribution to the literature.

H6: The "Human-Communicative" condition attracts the most attention. This interaction was significant for Total Dwell Time and Reaction Time. Under high load, the hypothesis was supported for the human agent.

## 4.2 Limitations and Future Directions

These findings should be interpreted in light of some limitations. The actions on the screen are displayed for one second due to a framing issue when the actions are completed in two seconds. For most of the actions, the initial intention of the actions can be understood in one second, but we cannot control how participants perceive the actions as communicative or non-communicative.

The finding is specific to one type of humanoid robot (Pepper), and different robots may elicit a different range of anthropomorphism levels; the nature of the action is constrained by Pepper's ability to move. Future research may explore the effects of spectrum through the human and robot by the usage of mechanical industrial robots to more humanoid robots. Controlled laboratory settings may change the dynamics of attention, and distraction may occur differently in real-world settings since visual perception is not the only medium for perception. Future research could explore these factors in more naturalistic settings. Familiarity, emotional valence, or unpredictability of the distractor may modulate the attentional capture. Incorporating subjective measures can be used for the psychological mechanism of the social presence of robots. In addition, adaptation-based design may help to reveal whether attention to distractors diminished or increased with repeated exposure to a stimulus. Future research may use a different object, also adding different eye metrics, such as where the first fixation of the eye occurs and when, which can be important in order to understand if participants quickly regress into one area of interest and stay in that AOI or not.

## 4.3 Conclusion

These findings have important implications for theory, but it is also important for the practice. In theory, the multi-stage attention supports and demonstrates a dissociation between early and later mechanisms of attention. All results, excepting dwell time per fixation, aligned with the Perceptual Load Theory originally suggested by Nili Lavie [11]. Results are consistent and powerful in terms of the main effect of perceptual load across behavioral and eye metrics. Under cognitive strains, the processing of the communication of an action has a different nature than the human encounters. A speculation can be made by the anthropomorphic cues, because there is an interesting finding about the robot's communicative actions under a highly loaded context, which can be a source of conflict rather than a helpful social signal.

.

# Bibliography


[1] D. Broadbent, *Perception and Communication.* Pergamon Press, 1958.

[2] J. A. Deutsch and D. Deutsch, "Attention: Some theoretical considerations.," *Psychological review*, vol. 70, no. 1, p. 80, 1963.

[3] A. M. Treisman, "Strategies and models of selective attention.," *Psychological review*, vol. 76, no. 3, p. 282, 1969.

[4] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *Journal of the acoustical society of America*, vol. 25, pp. 975–979, 1953.

[5] N. Moray, *Attention: Selective processes in vision and hearing.* Routledge, 2017.

[6] N. Moray, "Attention in dichotic listening: Affective cues and the influence of instructions," *Quarterly journal of experimental psychology*, vol. 11, no. 1, pp. 56–60, 1959.

[7] H. Pashler, "The psychology of attention," *JOP*, vol. 14, no. 1, 2000.

[8] D. Kahneman, "Changing views of attention and automaticity," *Varieties of attention*, 1984.

[9] S. Yantis and J. C. Johnston, "On the locus of visual selection: evidence from focused attention tasks.," *Journal of experimental psychology: Human perception and performance*, vol. 16, no. 1, p. 135, 1990.

[10] G. W. Humphreys, "Flexibility of attention between stimulus dimensions," *Perception & Psychophysics*, vol. 30, pp. 291–302, 1981.

[11] N. Lavie and Y. Tsal, "Perceptual load as a major determinant of the locus of selection in visual attention," *Perception & psychophysics*, vol. 56, pp. 183–197, 1994.

[12] C. W. Eriksen and D. W. Schultz, "Information processing in visual search: A continuous flow conception and experimental results," *Perception & psychophysics*, vol. 25, no. 4, pp. 249–263, 1979.

[13] U. Cartwright-Finch and N. Lavie, "The role of perceptual load in inattentional blindness," *Cognition*, vol. 102, no. 3, pp. 321–340, 2007.

[14] N. Lavie and J. W. De Fockert, "Contrasting effects of sensory limits and capacity limits in visual selective attention," *Perception & Psychophysics*, vol. 65, no. 2, pp. 202–212, 2003.

[15] N. Lavie, "Perceptual load as a necessary condition for selective attention.," *Journal of Experimental Psychology: Human perception and performance*, vol. 21, no. 3, p. 451, 1995.

[16] S. Forster and N. Lavie, "High perceptual load makes everybody equal," *Psychological Science*, vol. 18, no. 5, pp. 377–381, 2007. PMID: 17576274.

[17] C. Folk, R. Remington, and J. Johnston, "Involuntary covert orienting is contingent on attentional control settings," *Journal of experimental psychology. Human perception and performance*, vol. 18, pp. 1030–44, 12 1992.

[18] J. Theeuwes, "Perceptual selectivity for color and form," *Perception psychophysics*, vol. 51, pp. 599–606, 07 1992.

[19] M. I. Posner, "Orienting of attention," *Quarterly journal of experimental psychology*, vol. 32, no. 1, pp. 3–25, 1980.

[20] A. Objects, "Steven yantis," *Visual Attention*, no. 8, p. 187, 1998.

[21] M. Cheal and D. R. Lyon, "Central and peripheral precuing of forced-choice discrimination," *The Quarterly Journal of Experimental Psychology*, vol. 43, no. 4, pp. 859–880, 1991.

[22] K. Nakayama and M. Mackeben, "Sustained and transient components of focal visual attention," *Vision research*, vol. 29, no. 11, pp. 1631–1647, 1989.

[23] J. Duncan and G. W. Humphreys, "Visual search and stimulus similarity.," *Psychological review*, vol. 96, no. 3, p. 433, 1989.

[24] S. Grossberg, E. Mingolla, and W. D. Ross, "A neural theory of attentive visual search: interactions of boundary, surface, spatial, and object representations.," *Psychological Review*, vol. 101, no. 3, p. 470, 1994.

[25] F. Katsuki and C. Constantinidis, "Bottom-up and top-down attention: different processes and overlapping neural systems," *The Neuroscientist*, vol. 20, no. 5, pp. 509–521, 2014.

[26] C. Bundesen, "Visual attention: Race models for selection from multielement displays," *Psychological Research*, vol. 49, no. 2, pp. 113–121, 1987.

[27] D. G. Lowe, *Perceptual organization and visual recognition*. 1985.

[28] S. P. Tipper, "The negative priming effect: Inhibitory effects of ignored primes," *Quarterly Journal of Experimental Psychology A*, vol. 37, pp. 571–590, 1985.

[29] D. Kahneman, *Attention and effort*, vol. 1063. Citeseer, 1973.

[30] Y. Yeshurun and H. Marciano, "Degraded stimulus visibility and the effects of perceptual load on distractor interference," *Frontiers in psychology*, vol. 4, p. 289, 2013.

[31] N. Lavie and E. Fox, "The role of perceptual load in negative priming.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 3, p. 1038, 2000.

[32] J. Driver, "A selective review of selective attention research from the past century," *British journal of psychology*, vol. 92, no. 1, pp. 53–78, 2001.

[33] P. Wei, G. Kang, and X. Zhou, "Attentional selection within and across hemispheres: Implications for the perceptual load theory," *Experimental brain research*, vol. 225, pp. 37–45, 2013.

[34] S. Fu, J. Fedota, P. M. Greenwood, and R. Parasuraman, "Early interaction between perceptual load and involuntary attention: An event-related potential study," *Neuroscience letters*, vol. 468, no. 1, pp. 68–71, 2010.

[35] N. A. Parks, D. M. Beck, and A. F. Kramer, "Enhancement and suppression in the visual field under perceptual load," *Frontiers in psychology*, vol. 4, p. 275, 2013.

[36] M. Sabri, C. Humphries, M. Verber, J. Mangalathu, A. Desai, J. R. Binder, and E. Liebenthal, "Perceptual demand modulates activation of human auditory cortex in response to task-irrelevant sounds," *Journal of cognitive neuroscience*, vol. 25, no. 9, pp. 1553–1562, 2013.

[37] D.-J. Yi, G. F. Woodman, D. Widders, R. Marois, and M. M. Chun, "Neural fate of ignored stimuli: dissociable effects of perceptual and working memory load," *Nature neuroscience*, vol. 7, no. 9, pp. 992–996, 2004.

[38] S. Forster and N. Lavie, "Failures to ignore entirely irrelevant distractors: the role of load.," *Journal of Experimental Psychology: Applied*, vol. 14, no. 1, p. 73, 2008.

[39] N. Kanwisher, J. McDermott, and M. M. Chun, "The fusiform face area: a module in human extrastriate cortex specialized for face perception," *Journal of neuroscience*, vol. 17, no. 11, pp. 4302–4311, 1997.

[40] P. Hains and J. Baillargeon, "La face animale et le visage humain sont-ils équivalents? une étude dans le cadre de la théorie de la charge perceptuelle," *L'Année psychologique*, vol. 111, no. 3, pp. 449–463, 2011.

[41] S. Sato and J. Kawahara, "Task-irrelevant faces capture attention regardless of perceptual load," *Journal of Vision*, vol. 14, no. 10, pp. 319–319, 2014.

[42] I. Gauthier, P. Skudlarski, J. C. Gore, and A. W. Anderson, "Expertise for cars and birds recruits brain areas involved in face recognition," *Nature neuroscience*, vol. 3, no. 2, pp. 191–197, 2000.

[43] T. Ro, A. Friggel, and N. Lavie, "Musical expertise modulates the effects of visual perceptual load," *Attention, Perception, & Psychophysics*, vol. 71, no. 4, pp. 671–674, 2009.

[44] E. J. Pereira, E. Birmingham, and J. Ristic, "Contextually-based social attention diverges across covert and overt measures," *Vision*, vol. 3, no. 2, 2019.

[45] J. K. Burgoon, V. Manusov, and L. K. Guerrero, *Nonverbal communication*. Routledge, 2021.

[46] M. Atkinson, A. Simpson, and G. Cole, "Visual attention and action: How cueing, direct mapping, and social interactions drive orienting," *Psychonomic Bulletin Review*, vol. 25, pp. 1–21, 08 2017.

[47] M. Tomasello, J. Call, and B. Hare, "Five primate species follow the visual gaze of conspecifics," *Animal Behaviour*, vol. 55, pp. 1063–1069, Apr. 1998.

[48] F. Capozzi and A. Kingstone, "The effects of visual attention on social behavior," *Social and Personality Psychology Compass*, vol. 18, no. 1, p. e12910, 2024.

[49] E. Birmingham, W. F. Bischof, and A. Kingstone, "Social attention and real-world scenes: The roles of action, competition and social content," *Quarterly Journal of Experimental Psychology*, vol. 61, no. 7, pp. 986–998, 2008. PMID: 18938281.

[50] R. Bedford, M. Elsabbagh, T. Gliga, A. Pickles, A. Senju, T. Charman, and M. Johnson, "Precursors to social and communication difficulties in infants at-risk for autism: Gaze following and attentional engagement," *Journal of autism and developmental disorders*, vol. 42, pp. 2208–2218, 01 2012.

[51] P. Lyyra, A. Myllyneva, and J. Hietanen, "Mentalizing eye contact with a face on a video : Gaze direction does not influence autonomic arousal," vol. 59, no. 4, pp. 360–367, 2018.

[52] N. Caruana, P. de Lissa, and G. McArthur, "Beliefs about human agency influence the neural processing of gaze during joint attention," *Social Neuroscience*, vol. 12, no. 2, pp. 194–206, 2017. PMID: 26942996.

[53] T. Reuscher, P. Toreini, and A. Maedche, "The state of the art of diagnostic multiparty eye tracking in synchronous computer-mediated collaboration," *Journal of Eye Movement Research*, vol. 16, 06 2023.

[54] H. Admoni, "Nonverbal communication in socially assistive human-robot interaction," *AI Matters*, vol. 2, pp. 9–10, 12 2016.

[55] P. E. Downing, D. Bray, J. Rogers, and C. Childs, "Bodies capture attention when nothing is expected," *Cognition*, vol. 93, no. 1, pp. B27–B38, 2004.

[56] A. Kingstone, C. Tipper, J. Ristic, and E. Ngan, "The eyes have it!: An fmri investigation," *Brain and Cognition*, vol. 55, no. 2, pp. 269–271, 2004.

[57] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, vol. 166, no. 1, pp. 140–164, 2005.

[58] R. A. Jones, "Relationalism through social robotics," *Journal for the Theory of Social Behaviour*, vol. 43, no. 4, pp. 405–424, 2013.

[59] D. Mcneill, "Hand and mind: What gestures reveal about thought," *Bibliovault OAI Repository, the University of Chicago Press*, vol. 27, 06 1994.

[60] M. Cook, "Gaze and mutual gaze in social encounters: How long—and when—we look others "in the eye" is one of the main signals in nonverbal communication," *American Scientist*, vol. 65, no. 3, pp. 328–333, 1977.

[61] S. Saunderson and G. Nejat, "How robots influence humans: A survey of nonverbal communication in social human–robot interaction," *International Journal of Social Robotics*, vol. 11, pp. 575 – 608, 2019.

[62] Y. Rivero-Moreno, S. Echevarria, C. E. Vidal Valderrama, L. Stefano-Pianetti, J. Cordova-Guilarte, J. Navarro Gonzalez, J. Acevedo-Rodriguez, G. Dorado-Avila, L. Osorio-Romero, C. Chavez-Campos, and K. Acero-Alvarracín, "Robotic surgery: A comprehensive review of the literature and current trends," *Cureus*, vol. 15, 07 2023.

[63] B. Hendriks, B. Meerbeek, S. Boess, S. Pauws, and M. Sonneveld, "Robot vacuum cleaner personality and behavior," *International Journal of Social Robotics*, vol. 3, pp. 187–195, Apr. 2011.

[64] I. Asimov, *I, robot*, vol. 1. Spectra, 2004.

[65] P. Wel and H. Steenbergen, "Pupil dilation as an index of effort in cognitive control tasks: A review," *Psychonomic Bulletin Review*, vol. 25, 02 2018.

[66] W. Chen, T. Sawaragi, and T. Hiraoka, "Comparing driver reaction and mental workload of visual and auditory take-over request from perspective of driver characteristics and eye-tracking metrics," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 97, pp. 396–410, 08 2023.

[67] M. Recarte and L. Nunes, "Effects of verbal and spatial–imagery tasks on eye fixations while driving," *Journal of Experimental Psychology: Applied*, vol. 6, 03 2000.

[68] K. Rayner, "Eye movements in reading and information processing: 20 years of research.," *Psychological bulletin*, vol. 124, no. 3, p. 372, 1998.

[69] K. F. Van Orden, T.-P. Jung, and S. Makeig, "Combined eye activity measures accurately estimate changes in sustained visual task performance," *Biological Psychology*, vol. 52, no. 3, pp. 221–240, 2000.

[70] J. William and R. Murugesh, "Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research," *Journal of scientific research*, vol. 64, pp. 168–175, 01 2020.

[71] J. Atkinson and O. Braddick, "Chapter 12 - visual development," in *Neurocognitive Development: Normative Development* (A. Gallagher, C. Bulteau, D. Cohen, and J. L. Michaud, eds.), vol. 173 of *Handbook of Clinical Neurology*, pp. 121–142, Elsevier, 2020.

[72] A. Bahill, M. R. Clark, and L. Stark, "The main sequence, a tool for studying human eye movements," *Mathematical Biosciences*, vol. 24, no. 3, pp. 191–204, 1975.

[73] A. Gegenfurtner, E. Lehtinen, and R. Säljö, "Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains," *Educational Psychology Review - EDUC PSYCHOL REV*, vol. 23, pp. 523–552, 12 2011.

[74] P. Binda and M. C. Morrone, "Vision during saccadic eye movements," *Annual review of vision science*, vol. 4, no. 1, pp. 193–213, 2018.

[75] D. Litchfield and L. J. Ball, "Rapid communication: Using another's gaze as an explicit aid to insight problem solving," *Quarterly Journal of Experimental Psychology*, vol. 64, no. 4, pp. 649–656, 2011. PMID: 21347990.

[76] M. Treviño, G. Birdsong, A. Carrigan, P. Choyke, T. Drew, M. Eckstein, A. Fernandez, B. Gallas, M. Giger, S. Hewitt, T. Horowitz, Y. Jiang, B. Kudrick, S. Martinez-Conde, S. Mitroff, L. Nebeling, J. Saltz, F. Samuelson, S. Seltzer, and J. Wolfe, "Advancing research on medical image perception by strengthening multidisciplinary collaboration," *JNCI Cancer Spectrum*, vol. 6, 12 2021.

[77] A. Andrychowicz-Trojanowska, "Experimental eyetracking glottodidactics," 01 2018.

[78] P. Viaene, P. Vansteenkiste, M. Lenoir, A. De Wulf, and P. De Maeyer, "Examining the validity of the total dwell time of eye fixations to identify landmarks in a building," *Journal of eye movement research*, vol. 9, no. 3, 2016.

[79] J. Duncan, R. Ward, and K. Shapiro, "Direct measurement of attentional dwell time in human vision," *Nature*, vol. 369, no. 6478, pp. 313–315, 1994.

[80] R. Mallick, D. Slayback, J. Touryan, A. J. Ries, and B. J. Lance, "The use of eye metrics to index cognitive workload in video games," in *2016 IEEE second workshop on eye tracking and visualization (etvis)*, pp. 60–64, IEEE, 2016.

[81] A. W. Joseph and R. Murugesh, "Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research," *J. Sci. Res*, vol. 64, no. 1, pp. 168–175, 2020.

[82] J. M. Henderson and T. J. S. and, "How are eye fixation durations controlled during scene viewing? further evidence from a scene onset delay paradigm," *Visual Cognition*, vol. 17, no. 6-7, pp. 1055–1082, 2009.

[83] C. K. Friesen and A. Kingstone, "The eyes have it! reflexive orienting is triggered by nonpredictive gaze," *Psychonomic bulletin & review*, vol. 5, no. 3, pp. 490–495, 1998.

[84] A. F. Ward, K. Duke, A. Gneezy, and M. W. Bos, "Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity," *Journal of the Association for Consumer Research*, vol. 2, no. 2, pp. 140–154, 2017.

[85] A. Parola and F. M. Bosco, "An eye-tracking investigation of the cognitive processes involved in the comprehension of simple and complex communicative acts," *Quarterly Journal of Experimental Psychology*, vol. 75, no. 10, pp. 1976–1995, 2022.

[86] 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.

[87] E. Wiese, H. J. Müller, and A. Wykowska, "Using a gaze-cueing paradigm to examine social cognitive mechanisms of individuals with autism observing robot and human faces," in *Social Robotics: 6th International Conference, ICSR 2014, Sydney, NSW, Australia, October 27-29, 2014. Proceedings 6*, pp. 370–379, Springer, 2014.

[88] T. D. Karaduman and A. B. Ozsu, "Robot distraction," May 2025.

[89] T. N. Pekçetin, G. Aşkın,  Evsen, T. Karaduman, B. Barinal, J. Tunç, C. Acarturk, and B. Urgen, "Hr-act (human-robot action) database: Communicative and noncommunicative action videos featuring a human and a humanoid robot," 02 2025.

[90] T. M. Inc., "Matlab version: 9.13.0 (r2024b)," 2024.

# Appendix A

# Data and Code

All data can be accessed through the link (https://osf.io/q7jwv/?view$_only = 9de897f91f2446179ef636920a9bb661)[AccessedJune23th, 2025]$