Characterizing Mental Workload in Physical Human-Robot Interaction Using Eye-Tracking Measures

Satyajit Upasani

Dissertation submitted to the faculty of
Virginia Polytechnic Institute and State University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy in
Industrial and Systems Engineering

Divya Srinivasan, Co-Chair

Joseph L. Gabbard, Co-Chair

Nathan Lau

Alexander Leonessa

Maury A. Nussbaum

May 2, 2023

Blacksburg, Virginia

Keywords: Visuomotor Coordination, Gaze Entropy, Collaborative Robotics, Exoskeletons, Virtual Reality, Motor Learning, Age Differences

Characterizing Mental Workload in Physical Human-Robot Interaction Using Eye-Tracking Measures

Satyajit Upasani

ABSTRACT

Recent technological developments have ushered in an exciting era for collaborative robots (cobots), which can operate in close proximity with humans, sharing and supporting task goals. While there is increasing research on the biomechanical and ergonomic consequences of using cobots, there is relatively little work on the potential motor-cognitive demand associated with these devices. These cognitive demands primarily stem from the need to form accurate internal (mental) models of robot behavior, while also dealing with the intrinsic motor-cognitive demands of physical co-manipulation tasks, and visually monitoring the environment to ensure safe operation. The primary aim of this work was to investigate the viability of eye-tracking measures for characterizing mental workload during the use of cobots, while accounting for the potential effects of learning, task-type, expertise, and age-differences. While eye-tracking is gaining traction in surgical/rehabilitation robotics domains, systematic investigations of eye tracking for studying interactions with industrial cobots are currently lacking. We conducted three studies in which participants of different ages and expertise levels learned to perform upper- and lower-limb tasks using a dual-armed cobot and a whole-body powered exoskeleton respectively, over multiple trials. Robot-control difficulty was manipulated by changing the joint impedance on one of the robot arms (for the dual-armed cobot).

The first study demonstrated that when individuals were learning to interact with a dual-armed cobot to perform an upper-limb co-manipulation task simulated in a virtual reality (VR) environment, pupil dilation (PD) and stationary gaze entropy (SGE) were the most sensitive and reliable measures of mental workload. A combination of eye-tracking measures predicted performance with greater accuracy than experimental task variables. Measures of visual attentional focus were more sensitive to task difficulty manipulations than typical eye-tracking workload measures, and PD was most sensitive to changes in workload over learning. The second study showed that compared to walking freely, walking while using a complex whole-body powered exoskeleton: a) increased PD of novices but not experts, b) led to reduced SGE in both groups and c) led to greater downward focused gaze (on the walking path) in experts compared to novices. In the third study using an upper-limb co-manipulation task similar to Study 1, we found that the PD of younger adults reduced at a faster rate over learning, compared to that of older adults, and older adults showed a significantly greater drop in gaze transition entropy with an increase in task difficulty, compared to younger adults. Also, PD was sensitive to learning and robot-difficulty but not environmental-complexity (collisions with objects in the task environment), and gaze-behavior measures were generally more sensitive to environmental-complexity.

This research is the first to conduct a comprehensive analysis of mental workload in physical human-robot interaction using eye-tracking measures. PD was consistently found to show larger effects over learning, compared to task difficulty. Gaze-behavior measures quantifying visual attention towards environmental areas of interest were found to show relatively large effects of task difficulty and should continue to be explored in future research. While walking in a powered exoskeleton, both novices and experts exhibited compensatory gaze strategies. This finding highlights potentially persistent effects of using cobots on visual attention, with potential implications to safety and situational awareness. Older adults were found to apply greater mental effort (indicated by sustained PD) and followed more constrained gaze patterns in order to maintain similar levels of performance to younger adults. Perceived workload measures could not capture these age-differences, thus highlighting the advantages of eye-tracking measures. Lastly, the differential sensitivity of pupillary- and gaze behavior metrics to different types of task demands highlights the need for future research to employ both kinds of measures for evaluating pHRI. Important questions for future research are the potential sensitivity of eye-tracking workload measures over long-term adaptations to cobots, and the potential generalizability of eye-tracking measures to real-world (non-VR) tasks.

Characterizing Mental Workload in Physical Human-Robot Interaction Using Eye-Tracking Measures

Satyajit Upasani

GENERAL AUDIENCE ABSTRACT

Collaborative robots (cobots) are an exciting and novel technology that may be used to assist human workers in manual industrial work, reduce physical demand, and potentially enable older adults to re-enter the workforce. However, relatively little is known about the potential cognitive demands that cobots may impose on the human user. Although intended to assist humans, some cobots have been found to be difficult to use, because of the time and effort that is needed to learn their control dynamics (i.e. to learn how to physically control them to perform a complex manual task). Thus, it is important to better understand the potential mental demand/workload that a human operator may experience, while using a cobot, and how this demand may vary over time and learning to use the cobot. Eye-tracking is a promising technique to measure a cobot-operators' mental workload, since it can provide various measures that correlate with the involuntary physiological response to mental workload (e.g. pupil dilation - PD), as well as voluntary gaze strategies (e.g. the durations and patterns of where people look) in order to perform a physical/motor task. Eye-tracking measures may be used to continuously and precisely evaluate whether a cobot imposes excessive workload on the human operator, and if high workload is observed, the cobot may be programmed to adapt its behavior to reduce workload. Although eye-tracking is gaining traction in surgical/rehabilitation robotics domains, systematic investigations of eye tracking for studying interactions with industrial cobots are currently lacking. We designed three studies in which we investigated 1) the ability of eye-tracking measures to measure changes in mental workload while participants learned to use a cobot under different difficulty-levels 2) the changes in pupil diameter and gaze behavior when participants walked while wearing a whole-body powered exoskeleton as opposed to walking freely, and potential differences between novice- and expert exoskeleton-users 3) the differences in mental workload and visual attention between younger and older adults while learning to use a cobot. The first and third studies used virtual reality (VR) to simulate the task environment, to allow for precise control over the presentation of stimuli.

In study 1, we found that in higher difficulty-levels, participants' pupils were significantly more dilated, i.e., participants experienced higher mental workload, than in lower-difficulty levels. Also, PD gradually reduced as participants learned to better perform the task. In difficult task-conditions, participants gazed more frequently at the robot, and showed higher randomness (entropy) in their gaze patterns. The proportion of gaze falling on certain objects was at least as sensitive an indicator of task-difficulty, as PD and gaze entropy. In study 2, we found that walking in a whole-body exoskeleton was cognitively demanding, but only for novice participants. However, both novice and expert participants showed changes in their gaze patterns while walking in the exoskeleton – both groups lowered their gaze and focused on the walking path to a greater extent, compared to walking freely. Lastly, in study 3, we also found that older adults applied greater mental effort for maintaining similar levels of performance as younger adults. Older adults also exhibited more repetitive scanning patterns compared to younger adults, when task difficulty increased. This may have been due to potential reduction in the capacity to control attention with age. Our work demonstrates that eye-tracking measures are sensitive and reliable metrics of workload, and that different metrics are sensitive to different sources of workload. Specifically, PD was sensitive to robot-difficulty, and measures of visual attention were generally more sensitive to the complexity of the task environment. Important questions for future research are the potential changes in eye-tracking workload measures over longer time periods of learning to use cobots, and how these results generalize to real-world tasks that are not performed in virtual reality.

Acknowledgements

It takes a village. I have so many people to thank for their support, encouragement, and patience, as I embarked on this exciting and anxiety-ridden, but ultimately the most rewarding, quest. I would like to begin by thanking my advisor, Dr. Divya Srinivasan, for being my mentor and hero, for helping me mature as a student, researcher, and communicator, and for calling me an "expert" long before I could even hope to be one. She showed me that it is possible to achieve great success in life while also building a healthy and loving family and engaging in multiple creative pursuits outside of work. She is the primary reason for why this PhD journey has been such an enriching experience for me.

I am also thankful to my co-advisor, Dr. Joe Gabbard, for inspiring me to pursue UX research, and for giving me my first real-world UX project experience. He did this at a difficult time in my life, when my self-worth was in tatters, and I was drifting along as if a kite with a cut string. My life may have taken a very different path, had it not been for his support.

I am thankful to Dr. Nathan Lau, Dr. Alexander Leonessa, and Dr. Maury Nussbaum, for challenging me and providing their honest, expert feedback and continued support. I continue to be awed by their presence on my PhD committee.

I am thankful to my friends from Virginia, South Carolina and Mississippi for being nothing short of awesome. Harsh, Aishwarya, Sreyoshi, Tamoghna, Sarang, Taha, Disha, Ayaan, Almasji, Shreyas, Maithri, the Arabs, Eric, Rahul, Vishwajeet, Alec, Mithun and Eshan - thank you for all our shared experiences, and for the most fun and laughs I've ever had in my life. Every single one of you has inspired and changed me in unique ways, and I am truly honored to have met you all.

I am thankful to my mother- and father-in-law for their unshakeable trust and kindness – you have truly been my second set of parents. Thank you, Sonia and Alok, for welcoming me into your home, for keeping me well-fed, and for showing me a kind of generosity that was beyond anything I could have hoped for. It would have been very difficult for Sneha and me to get through this journey, had it not been for your support.

I am eternally thankful to Aai, Baba, and Sanat, for their unending love, and for making me who I am today. Aai instilled in me a sense of wonder, a love for reading and writing, and the courage to be able to laugh at myself. Baba showed me what it means to truly, truly care for people, to charm with words, and to dare to be ambitious. Sanat showed me the true meaning of persistence and resilience - he will always be my best friend and best man.

My dear Sneha – I dedicate my PhD to you. Thank you for being my rock, and for waiting for me all these years. A PhD takes a lot of patience and grit, but it takes a whole other kind of resilience to support a loved one while they go through it. You had no control over the process, and all you could do was to stay hopeful that I would figure it out one day. You are amazing, I love you, and I can't wait to start our new life together!

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Industrial Collaborative Robots

Recent developments in robotic control algorithms and material design have ushered in an exciting era for collaborative robots (cobots), which have grown much more usable, versatile, and safer to operate in close proximity with humans compared to older industrial robots (Haddadin & Croft, 2016). In fact, robotic exoskeletons, prosthetics, and some types of cobots remain in complete physical contact with the human operator, mimicking the operator's movements, often cooperating in the performance of a task, or augmenting strength, endurance and precision (Bergamasco & Herr, 2016; Chen & Kemp, 2010; Zhu et al., 2020). In this context, it is important to understand the cognitive challenges involved in controlling these complex devices, and the learning/training needs for diverse operators to effectively utilize them. Recent research suggests that although cobots have achieved higher standards of safety and compliance, they can still impose a significant workload on the user's attentional and cognitive-motor resources (Chadwell et al., 2016; Stirling et al., 2020; Wu et al., 2019) and that cobot operation may require time and effort to learn (Aronson et al., 2018; Cornwall, 2015; J. V. V. Parr et al., 2019). Thus, it is critical to develop metrics to quantify a user's mental workload for the safe and effective implementation of cobots, especially in environments that are inherently hazardous and safety-critical. Additionally, cobots will likely be used by a diverse set of individuals with widely varying demographics, physical and cognitive capabilities, and skill-levels, wherein the experience of the collaboration may be quite different, also resulting in different cognitive effort and learning rates. Importantly, the rise in the average age of industrial workers (Calzavara et al., 2020; Hedge et al., 2006) necessitates that future research takes aging-related changes in capabilities and task performance into account, in order to be inclusive in presenting a complete picture of the workload dynamics in physical human-robot interaction (pHRI).

There are numerous ways in which a human interacts physically with a robot. Multiple classifications and taxonomies of pHRI have been proposed (Aaltonen et al., 2018; Haddadin & Croft, 2016; Hentout et al., 2019), but they share the common theme of being defined along a continuum that shows an increasing degree of proximity and/or contact between the robot and the human operator. As defined by (Aaltonen et al., 2018), human-robot collaboration spans across four levels – 1) No co-existence (and no interaction), in which there is complete physical separation between the human and the robot such as in traditional fenced robot cells, 2) Coexistence, in which the human and the robot may work (partially or completely) in a shared space, but with different goals, 3) Cooperation, in which the human and the robot work towards a shared goal in (partially or completely) shared space, either simultaneously or sequentially, and 4) Collaboration, in which the human and robot work simultaneously on a shared object, or co-manipulate the object in shared space. A schematic of these four levels of collaborative pHRI is illustrated in figure 1.

Among these levels of human-robot collaboration, we are particularly interested in co-manipulation tasks involving continuous force interactions between the human and the robot, as this is a relatively new area presenting several important research challenges in the development of fluid human-robot interactions. Co-manipulation is defined by having human(s), robot(s) and the environment come to contact with each other (or through the manipulated object) to form a tightly coupled dynamical system to accomplish a task (Ajoudani et al., 2018; Bauer et al., 2008; Krüger et al., 2009). Ideally, each active component of such a system must be capable of observing and estimating the counterparts' contributions to the overall system's response through the fusion and processing of sensory information (Argall & Billard, 2010; Ebert & Henrich, 2002; Lallee et al., 2012). Many sensing strategies continue to be proposed for the robot to detect the human's behavior and respond in a manner that improves the fluidity and intuitiveness of the interaction (Haddadin & Croft, 2016; Hentout et al., 2019). However, the design of intuitive co-manipulation is still a nascent field of research (Haddadin & Croft, 2016), and there are no well-established metrics for evaluating "intuitiveness". Current approaches mostly infer intuitiveness or ease of interaction based on a reduction in metabolic cost (Haddadin & Croft, 2016) or subjective and performance metrics (Hoffman, 2019); however, in such complex and dynamic interactions, it is also important to consider the potential mental workload and motor-control demands that the human operator may experience, so that the robot can anticipate and/or respond to human state to form a “fluent” interaction (Hoffman, 2019).

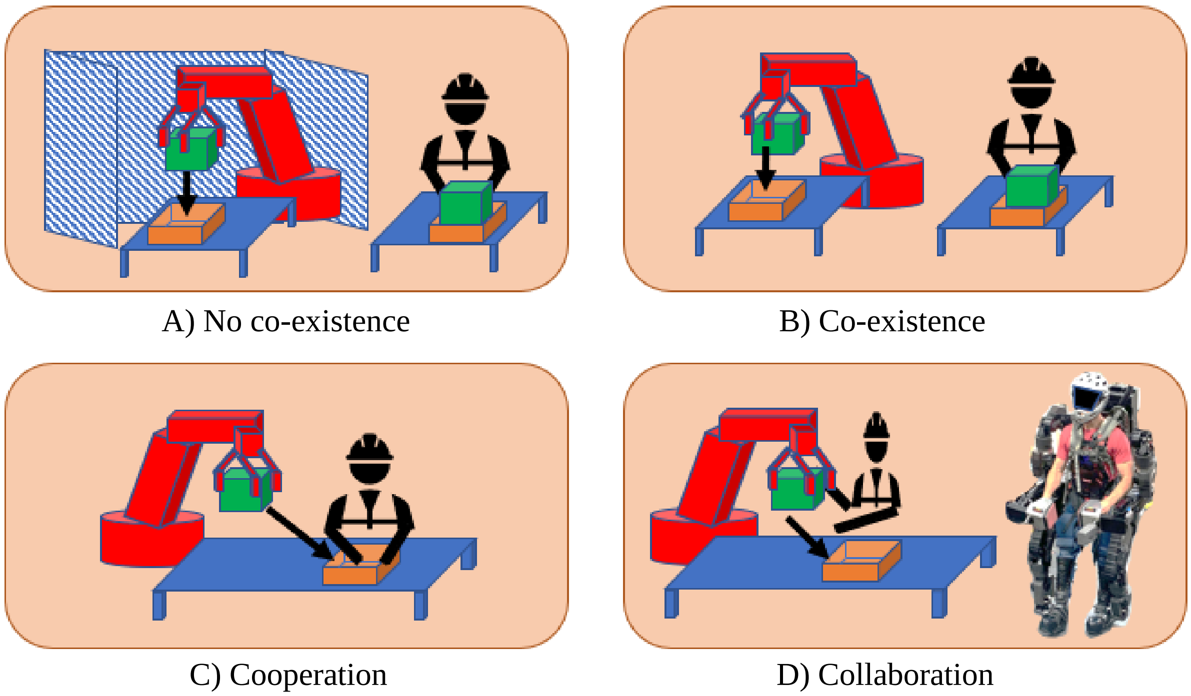


Figure 1. Schematic of levels of physical human-robot collaboration

1.2 Mental Workload in Physical Human-Robot Interaction

Mental workload is defined as the level of attentional resources that is demanded by a task (Tsang & Vidulich, 2006; Wickens, 2008). There are two main determinants of mental workload – exogenous task demands, and endogenous supply of attentional resources. Exogenous task demands are “out in the world”: some examples include the intrinsic difficulty of the task being performed, task priorities, and situational constraints; whereas endogenous factors pertain to the individual, and relate to the attentional supply that is needed for perceiving, planning, decision-making and response processing (Tsang & Vidulich, 2006). As a task becomes more difficult, the attentional resources required to maintain performance, and hence mental workload will also increase, although this relationship between task difficulty and workload is mediated by choice of strategy and individual differences in age, ability and expertise. For example, older adults may exhibit lower working memory capacity (Hale et al., 2011) and hence require more cognitive effort to meet task demands, leading to increased workload. Alternatively, age and expertise are also correlated with enhanced crystallized intelligence (Beier & Ackerman, 2005) and better choice of strategies to meet task demands, potentially leading to more frequent resolution of task goals and consequently lower workload (Tsang & Vidulich, 2006). Thus, the accurate measurement of mental workload should account for individual differences in age, ability and expertise in addition to the effects of task difficulty (M. S. Young et al., 2015).

In this context, it is worth noting that performing motor skills can be cognitively demanding, especially in the initial stages of learning (Seidler et al., 2012). Cognitive phenomena such as attention, working memory, and decision-making are relevant not only to knowledge-based cognitive work, but also to complex motor skills such as those required in competitive sports (Buszard & Masters, 2018; Draheim et al., 2021; Lee et al., 1994). Even simpler tasks such as maintaining a posture or walking on level ground have been shown to require cognitive involvement under specific conditions (Ellmers et al., 2016; Kahya et al., 2018; Walsh, 2021). There is growing neurophysiological support from studies using electroencephalography and event-related potentials that complex motor performance places demands attentional and working memory resources, and that this demand attenuates with learning and expertise development in the task (Jaquess et al., 2017; Rietschel et al., 2014; Shaw et al., 2018). As mentioned previously, mental workload can arise out of various task-related factors, and these factors can be specific to a given domain. The following section will discuss the potential task-related workload factors that are specific to co-manipulation tasks in pHRI.

1.3 Factors affecting workload in co-manipulation tasks

There are three main task-related factors that can lead to mental workload in co-manipulation tasks:

1) The attentional demands associated with forming an internal model of the cobot’s dynamics (Aronson et al., 2018; Bequette et al., 2020; Stirling et al., 2020; Zhang et al., 2016) for being able to predict the behavior of the robotic system.

2) The cognitive and motor-control demands associated with completing complex task goals (Buszard et al., 2017; Van Acker et al., 2020; Wu et al., 2019).

3) The need to continuously monitor the environment to avoid collisions and other safety incidents (Chen & Kemp, 2010; Steinfeld et al., 2006).

1.3.1. Internal Model Formation:

Learning a new skill, especially a 'motor' skill, requires a process known as "Internal Model Formation" (Wolpert et al., 2011), in which the learner, with practice, gradually develops the ability to predict the sensory consequences or results of the physical actions associated with a skill. Internal models are control functions in the central nervous system (see figure 2 for a schematic) which predict the sensory consequences (results) of a motor command (movement). If the actual sensory consequences of an action do not match the predicted ones, the central nervous system generates an error signal representing this difference, and this error signal "tunes" or "updates" the internal model parameters towards a more accurate prediction. An enhanced ability to predict the results of one's own actions is a critical aspect of what formulates expertise in the task. Learning to use novel or complex tools is also characterized by the formation of new internal models for tool behaviors, as well as updating of the internal models for the limb controlling the tool (Wolpert et al., 2011).

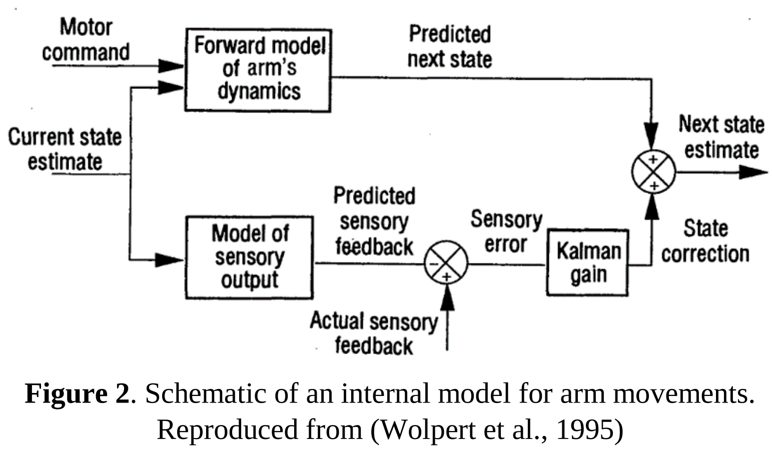


Figure 2. Schematic of an internal model for arm movements.
Reproduced from (Wolpert et al., 1995)

From a cognitive perspective, internal model formation involves the engagement of attentional and memory processes to drive decision-making (Krakauer et al., 2019; Lee et al., 1994). In this process, attention is actively directed for **acquiring the perceptual information** that is relevant to the task (Ellmers et al., 2016; Foerster et al., 2011; Koren et al., 2022), and multiple movement **strategies are retrieved from memory** in order to **select the strategy** that is best suited for a particular situation (Lee et al., 1994). Although internal models have largely been discussed in the motor learning and neuroscience literature, they were also referenced in early human factors research on higher-order manual control (Wickens & Derrick, 1981). This research suggested that greater central-executive resources are demanded while forming an internal model of higher-order control behavior such as that observed in aircraft-control and other similar systems with a high inertial response or "sluggishness". Recent research has shown that wearable robots such as myoelectric prostheses and powered exoskeletons tend to be initially difficult to use because the user cannot easily predict the devices' control dynamics (Chadwell et al., 2016; Cornwall, 2015; Kao, 2009; Kim et al., 2021; Stirling et al., 2020). The uncertainty and delays in prosthetic control systems and/or the absence of proprioceptive feedback can make internal model formation difficult (Chadwell et al., 2016; Sobuh et al., 2014). This difficulty has been observed through an increased reliance on vision to monitor the movement of the robot. For example, it has been found that upper-limb prosthetic users devote much of their attention towards monitoring the prosthetic itself rather than other task-relevant objects (J. V. V. Parr et al., 2019). Similar difficulty in predicting robot behavior has also been observed with other types of cobots, e.g., joystick-operated robotic arms (Aronson et al., 2018).

Early lower-limb active exoskeletons such as the Human Universal Load Carrier (HULC) were observed to cause unnatural gait kinematics (Cornwall, 2015), as there was a mismatch between operator expectations and system behavior, thereby making the device difficult to learn to use. A more recent study found that walking with a lower-limb active exoskeleton increased perceived mental workload and reduced performance on secondary cognitive tasks, although these effects varied across individuals (Bequette et al., 2020). The authors cited motor coordination demands as the potential source of the mental demand. Another recent study found that compared to walking without an exoskeleton, users who walked with a whole-body powered exoskeleton activated their lower-limb muscles more frequently and at a higher level, and took longer to initiate steps during each gait-cycle (i.e. they increased the time spent in dual stance) (Kim et al., 2021). The authors stated that these behaviors may reflect increased mental and physical effort for maintaining walking stability, possibly due to an imperfect mental model of the exoskeleton's dynamics.

Finally, learning also impacts mental workload. For instance, although workload may be high in the early phase of learning, gradual skill-acquisition and formation of internal models can reduce workload. Studies have found that the eye-movement patterns of prosthetic-users can approach those of anatomically-intact individuals over the course of training, suggesting that the initial perceived difficulty may reduce as users develop appropriate internal models (Sobuh et al., 2014). A reduction in workload over the course of learning has also been demonstrated in studies on robotic surgical skills training (Wu et al., 2021). Thus, we can expect a high mental workload to be present during initial stages of learning to use a cobot, as the user attempts to build an internal model of the device and movement (Sailer, 2005), and over the course of practice, mental workload is expected to attenuate due to refinement of neural processes and increasing automaticity in the task (Rietschel et al., 2014; Sailer, 2005; O. White & French, 2017).

While internal model formation and associated mental workload can change over the course of learning, the demands associated with visual monitoring of the environment and the cognitive demands due to task goals may persist, especially if the intrinsic difficulty of the task is high, e.g., in robotic surgery or complex automotive assembly processes. The nature of these persistent task-relevant demands is discussed in the following sections.

1.3.2 Cognitive Demands due to Task Goals:

Even if an operator has a well-developed internal model of a tool or device, co-manipulation tasks can be intrinsically difficult, due to the need to remember multiple task steps with conditional relationships and perform complex spatial transformations mentally (Van Acker et al., 2020; Wu et al., 2021), deal with time pressures (Bommer & Fendley, 2018), and ensure high-precision (Jiang et al., 2015; Wu et al., 2021).

It is well-known that it can be challenging to operate remote-controlled robots under reduced visual and haptic feedback (Lanfranco et al., 2004; Zhu et al., 2022). Thus, task goals are an important determinant of mental workload in physical human-robot interaction.

The nature of modern industrial work, especially in the Industry 5.0 era, has become increasingly cognitive and multi-task oriented (Maddikunta et al., 2022). In addition to the physical and motor-control demands that have always been present in manufacturing jobs, workers now regularly perform cognitive work using advanced digital technologies and artificial intelligence to support their tasks (Neumann et al., 2021). Although these technologies are intended to help worker productivity, it is possible that they may unintentionally create additional multitasking loads when coupled with the existing motor-control demands.

1.3.3 Visual Monitoring of the Environment:

There may be multiple instances in which the human operator may have to intervene, or engage in on-line monitoring and control while using a cobot (Steinfeld et al., 2006). For example, during co-manipulation, the human may have to monitor the object being manipulated as well as the surrounding environment in order to avoid collisions and to ensure that the object's orientation and placement is aligned with task goals. These monitoring demands may be especially relevant in a high-precision task, for example, controlling a robot arm to skewer food (Aronson et al., 2018) (Figure 3), or in cluttered and unstructured environments where the potential for collisions may be higher. High monitoring demands on an already limited human attentional capacity may increase mental fatigue and the risk of safety incidents (Li et al., 2019), thus making the accurate measurement of these demands an important and relevant research problem.

Although there has been continued success in the effort to design intelligent collision-avoidance (Haddadin & Croft, 2016) and stabilization (Haddadin & Croft, 2016; Lanfranco et al., 2004) capabilities for cobots, it is impossible to predict and account for all types of instabilities and anomalies that may occur in the human-robot interaction (Aronson & Admoni, 2018; Prewett et al., 2010). Additionally, the task environment may change rapidly in dynamic work scenarios, elevating the role of the human as a decision-maker (Honig & Oron-Gilad, 2018). Thus, it is very likely that the human operator will need to continue to support task goals through monitoring and intervention from time to time.

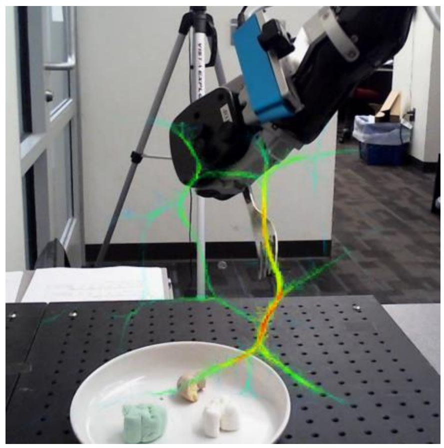


Figure 3. Aggregated gaze scanning pattern while using a robot arm to skewer food. The user fixated on the food items to help align the robot arm correctly. Reproduced from (Aronson et al., 2018)

1.4 The role of individual differences on mental workload

In addition to the task-related influences mentioned above, individual differences can also influence motor performance, mental workload, and learning rates. For example, complex upper- and lower-limb motor control tasks may be especially mentally demanding for older adults due to their tendency to control their movements more consciously and engage a greater proportion of their attentional and working memory resources compared to younger adults (Seidler et al., 2010). A study by Coats et al. compared differences in eye-movement patterns between older and younger adults while they performed consecutive pick-and-place movements under varying precision demands (Coats et al., 2016). In the baseline (easy) condition, participants had to pick and place two balls onto a stationary tray located on a table, and in the stabilizing (hard) condition, participants held the tray in their hands, thus imposing an additional task demand associated with stabilizing the plate and the balls. The study found that in the baseline condition, older adults performed a greater number of look-ahead eye movements, i.e. their eyes moved to the second ball before placing the first ball. This finding suggested that older adults tend to require more advance visual information from the environment for planning their movements. Additionally, in the harder, stabilizing condition, older adults gazed at the tray for a longer duration compared to younger adults due to the need to maximize visual information needed to keep the tray stable. Thus, older adults may spend greater effort to maximize visual information from the environment in motor-precision tasks. Another study (W. R. Young et al., 2012) found that high-risk older adults may experience higher levels of anxiety during obstacle-navigation and target-stepping tasks, leading to higher attentional demands, a reduced ability to form a "spatial map" of the upcoming path, and consequently lower stepping accuracy. That study observed that high-risk older adults spent a disproportionate amount of time focusing their gaze on an upcoming obstacle or target, as opposed to low-risk older adults who were able to monitor the future travel path to acquire a better spatial map.

Individual personality differences may also influence mental workload while performing motor skills. Based on the theoretical framework of Fitts and Posner (Fitts & Posner, 1967), motor learning progresses from a declarative knowledge stage requiring conscious control and considerable attention to a procedural knowledge stage in which performance is automatic and requires little attention. However, in a phenomenon termed as "reinvestment", certain individuals may regress to a declarative stage of processing even during tasks that are well-learned and automated, potentially because of personality influences, aging, or anxiety/pressure about performance (Masters & Maxwell, 2008). Formally, reinvestment is defined as the "manipulation of conscious, explicit, rule-based knowledge, by working memory, to control the mechanics of one's movement during motor output" (Masters & Maxwell, 2008). Thus, it is conceivable that a greater tendency to reinvest can increase the load on working-memory and attention, and consequently increase mental workload. It has been found that under high motor-cognitive demands, individuals who self-reported a greater tendency to reinvest in their movements exhibited lower accuracy in a target-stepping task, as well as greater decrements in dual-task performance (Ellmers et al., 2016). As stated previously, movement reinvestment is cognitively demanding (Masters & Maxwell, 2008), and the effects of reinvestment may be exacerbated in individuals such as high-risk older adults who may already experience cognitive impairment or lower working memory resources (Ellmers et al., 2016; W. R. Young et al., 2012).

A number of studies have found that performance on working memory tests predicted rates of learning in tasks involving sensorimotor adaptation and motor sequence learning; (see (Buszard & Masters, 2018; Seidler et al., 2012) for reviews). The function of working memory in motor control tasks may be to store movement error information, which can then be used to update and correct future movements (Seidler et al., 2012). The engagement of working memory resources to meet motor task demands may increase mental workload, and this effect may be exacerbated in older adults due to potential age-related reduction in working memory capacity (Hale et al., 2011; Seidler et al., 2012). A study involving a joystick-based targeting task under visuomotor rotation found that that older adults were not be able to engage their spatial working memory resources as effectively as younger adults (Anguera et al., 2011; Seidler et al., 2012), which negatively affected older adults' motor visuomotor learning rates. A failure to engage working memory may lead to compensatory strategies in situations with high task demands.

1.5 Measures of Mental Workload

Mental workload measures are classified into three types (Cain, 2007; M. S. Young et al., 2015):

1. **Self-report measures of workload using standardized questionnaires -** e.g., NASA-TLX: While self-report measures are technically simple and cost-efficient to administer, they usually require the participant to take breaks from the experimental task, so they cannot be acquired continuously. Thus, self-report measures may also show retrospective bias if administered post-test, or at less frequent intervals (M. S. Young et al., 2015).

2. **Performance measures:** Changes in mental workload can typically be inferred based on decrements in primary or secondary task performance. Performance measures can help relate mental workload to actual task performance and can be acquired in a more continuous manner compared to self-report measures. However, task performance has been found to dissociate with mental workload, (Brookhuis & Waard, 2002; Cain, 2007), thus making it necessary to combine performance measures with other indicators. Also, secondary tasks need to be carefully designed such that they do not interfere with the primary task (M. S. Young et al., 2015).

3. **Physiological measures, e.g., electroencephalography (EEG)-based measures, heart rate, and eye-tracking measures**: Physiological measures provide many advantages over self-report and performance measures since they are relatively unintrusive, and more importantly, they can provide continuous data on the user's mental workload (M. S. Young et al., 2015). Physiological sensors such as eye trackers can potentially provide rich data on both mental workload and motor performance dynamics.

Recent reviews have provided evidence that physiological measures can be sensitive indices of not only task difficulty and mental workload, but also learning (Tinga, 2019; Tinga et al., 2020). Several studies have employed the use of brain-related measures such as electroencephalography (EEG) and event-related potentials (ERP) (Deeny et al., 2014; Memar & Esfahani, 2020; Novak et al., 2015; J. V. V. Parr et al., 2019) and measures of the autonomic nervous system (ANS) (Darzi et al., 2017; Guerrero et al., 2013; Novak et al., 2011, 2015) to calculate mental workload while using physically-coupled robots. Brain-activity measures, in particular, have been increasingly popular, likely because of their ability to 'directly' measure mental workload in the brain (Tsang & Vidulich, 2006), and the ongoing hardware and software advances that make reliable and accurate workload measurement possible in naturalistic environments (Aricò et al., 2016; Dehais et al., 2019; Di Flumeri et al., 2018; Jebelli et al., 2018; So et al., 2017) . However, brain-related measures are prone to motion artifacts caused by body movements, which increases the amount of post-processing and data cleaning required in applications with high physical activity (Brouwer et al., 2015; Novak et al., 2014; Tinga et al., 2020; M. M. White et al., 2017) . ANS measures are more sensitive to physical activity and effort than they are to mental effort (C. E. Adams & Leverland, 1985; Charles & Nixon, 2019; Grassmann et al., 2016; Novak et al., 2011), making them less suitable for motor control tasks. Also, brain measurement devices such EEG, require the user to wear some form of securely fitted cap, or an array of electrodes on the head, which may be perceived as more intrusive than other measures (Bigliassi et al., 2019; Tinga et al., 2020) . Lastly, and most importantly, neither EEG nor ANS measures offer a direct measure of visual attention, which is closely coupled with motor performance and motor learning (Ariff et al., 2002; Binsted et al., 2001; de Vries et al., 2018; J. V. V. Parr et al., 2019; Sailer, 2005; Wilmut et al., 2006).

Eye-tracking, on the other hand, can overcome many of the practical disadvantages of brain-related and ANS measures due to its reported ease-of-use and wearability (Kovesdi et al., 2018; Pasqualotto et al., 2015; M. M. White et al., 2017), and its ability to provide a direct measure of visual attention which is critical for goal-directed motor performance. Additionally, since eye tracking is relatively unaffected by body movement, it is being used increasingly in motor-control and ambulatory tasks (Jiang et al., 2015; Matthis et al., 2018; Novak et al., 2015; Wu et al., 2019; Zhang et al., 2016). For these reasons, eye-tracking measures are becoming ubiquitous in HRI, especially for measuring visual attention and workload in motor-control tasks. **Yet, the sensitivity and reliability of different eye-tracking measures in collaborative physical human-robot interaction tasks, in terms of differences in task difficulty, individual characteristics of the humans, and learning stages, are relatively unexplored.**

1.5.1 Eye-tracking measures of mental workload:

The uniqueness of eye-tracking lies in its ability to provide physiological measures (e.g. pupil dilation, saccade velocity, workload indices) which correlate with the involuntary neural response to mental workload (Just et al., 2003) as well as gaze behavior measures that quantify the spatio-temporal characteristics of eye-movements and/or their relationship with objects in the real world (e.g. fixation rate, eye-hand span, and gaze entropy). Gaze-behavior measures reflect voluntary visuomotor strategies that are influenced by task demands (Lavoie et al., 2018; Leeuwen et al., 2017; Luke et al., 2006; Sailer, 2005;

Sarter et al., 2007; Srinivasan & Martin, 2010). Since mental workload is a multidimensional construct which is intricately related with attention, task performance, and strategies (Tsang & Vidulich, 2006), the ability of eye tracking to provide different types of information related to mental workload is advantageous. Lastly, eye-tracking measures have been shown to change over the course of learning both cognitive and motor skills (Foroughi et al., 2017; Sailer, 2005; Tinga et al., 2020; O. White & French, 2017).

Figure 4. **a)** Tobii Glasses 2 head-worn eye-tracker (*https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/*) **b)** Smart Eye Pro remote eye-tracking cameras; Adapted from (Lotz et al., 2020)

Several studies have used eye-tracking metrics as indicators of mental workload and motor performance in physical human-robot interaction (pHRI). Aronson et al. measured gaze patterns and pupil diameter while using a joystick-controlled Kinova MICO robot to spear food under different levels of automation (Aronson et al., 2018). They found that gaze patterns could effectively discriminate between different automation levels, and also that pupil diameter was elevated in the manual condition compared to the automated condition. Other studies found that participants' pupil diameter was elevated while using a prosthetic arm with a direct control algorithm compared to a pattern recognition algorithm, owing to the higher number of mental computations and mode switches present in direct control (M. M. White et al., 2017; Zhang et al., 2016). A series of studies have employed eye tracking metrics to quantify the difficulty of using prostheses during goal-directed reaching tasks (Bouwsema et al., 2014; Chadwell et al., 2016; J. V. V. Parr et al., 2018; Raveh et al., 2018; Sobuh et al., 2014). These studies found that prosthesis-users largely devoted their attention to the prosthesis itself, rather than the targets of their movements, which is contrary to gaze behavior observed in natural daily tasks such as tea-making and sandwich-making (Land, 2009). Additionally, prosthesis users exhibited fewer gaze transitions between different areas of interest (AOIs) and fewer look-ahead fixations than those who used their anatomical hand. These results suggest that prosthesis users may be experiencing high mental workload and hence lower spare attentional capacity available for tracking objects in their surroundings. Studies on the Da Vinci surgical robot have found that stationary gaze entropy (SGE) and pupil diameter were sensitive to perceived task difficulty (as measured through NASA-TLX ratings)(Wu et al., 2019, 2021).

There is a growing body of evidence that pupil dilation and gaze behavior can index cognitive effort and visual strategies in goal-directed gait (Ellmers et al., 2016), unstable walking conditions (Koren et al., 2020), postural demands (Kahya et al., 2018) and standing balance (Walsh, 2021). One study found that increased postural demands due to visual occlusion resulted in an increase in mental effort measured using the (pupil-based) Index of Cognitive Activity (ICA) (Kahya et al., 2018). A study by Ellmers et al. found that while performing a target-stepping task, the addition of a secondary cognitive task resulted in healthy young adults disengaging their gaze from the walkway and focusing instead on irrelevant information outside the walkway, thus reducing the intake of relevant visual information and negatively affecting their stepping accuracy (Ellmers et al., 2016). Another study found that high-risk older adults exhibited less-variable gaze in difficult obstacle-avoidance and target-stepping tasks compared to low-risk older adults (W. R. Young et al., 2012). In a quiet standing task, older adults exhibited lower gaze complexity (measured using sample entropy) than younger adults, and gaze complexity increased with the addition of a distracting cognitive task (Walsh, 2021). Lastly, a greater proportion of downward-oriented gaze was observed in healthy young adults who experienced unstable walking conditions (Koren et al., 2020). Combined, these results suggest that difficult coordination, movement planning, and postural control tasks may influence gaze strategies, particularly the variability of gaze, which may reduce with greater task difficulty. Measures such as gaze entropy may thus prove to be effective indices of task difficulty in ambulatory motor tasks.

Eye-tracking metrics can also be sensitive to the reductions in mental workload over the course of learning. A few studies on traditional cognitive or motor tasks have shown learning effects on eye-tracking metrics of mental workload, such as a decrease in pupil diameter (Foroughi et al., 2017), increased eye-hand span (Sailer, 2005) and an increase in target-fixations (Law et al., 2004). One of the studies on prosthetic use found that over multiple practice sessions, prosthesis users began to direct their attention towards the targets of their movement rather than the prosthesis itself, thus showing similarities to gaze behavior of participants using their anatomical hand (Sobuh et al., 2014). The study by Wu et al. found that SGE was negatively correlated with improvements in performance (Wu et al., 2021). These findings suggest that eye-tracking may be able to quantify the gradual increase in available attentional capacity (and decrease in mental workload) that is typically associated with expertise.

Past studies on eye-tracking and mental workload during cobot-use have largely focused on assistive or surgical robots. To our knowledge, there has been no research on the sensitivity of eye-tracking measures to quantify mental workload while learning to use a cobot for an industrial task. There have also been no studies on the cognitive effects of using a whole-body active exoskeleton. Additionally, while most of the studies in pHRI have compared eye-tracking metrics across multiple task conditions, the gradual changes in eye-gaze metrics as individuals adapt and learn novel interaction dynamics, as well as exhibit differences in the rates of learning due to underlying differences in task difficulty, have not been explored. Finally, given the large inter-individual variance in eye-tracking metrics and physiological measures in general (Hogervorst et al., 2014; Matthews et al., 2015; Sibley et al., 2020), the relative contributions of individual variance vs. external factors such as task difficulty or learning state in eye tracking measures applied to pHRI is also unknown. A better understanding of these factors will help establish the usefulness or potential utility of eye-tracking as a continuous measure of human cognitive state in pHRI, thereby guiding online control/re-adjustments of robotic systems to be responsive to human state.

1.5.2 Description of Eye-tracking Measures

Pupil diameter (PD)

Pupil diameter (PD) has been used as a measure of mental workload in multiple domains and a wide range of cognitive and psychomotor tasks, as described in the reviews (Charles & Nixon, 2019; Iqbal et al., 2005; Just et al., 2003; Marquart et al., 2015; Tao et al., 2019). Typically, PD increases with mental demands caused by an increasing task difficulty (figure 5), and this effect is called the task-evoked pupillary response (TEPR) (Beatty, 1982). Importantly, the TEPR reflects the level of mental effort that is invested in the task (van der Wel & van Steenbergen, 2018).

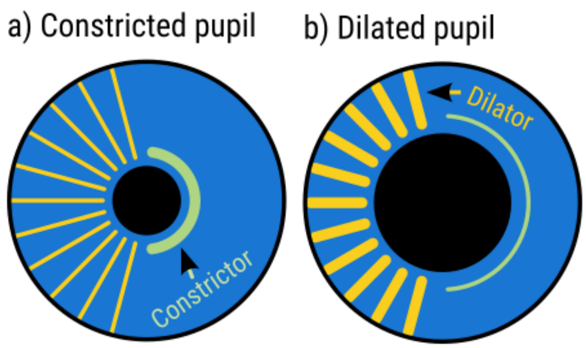


Figure 5. Task-evoked pupillary response (TEPR). Reproduced from (Mathot, 2018)

While the large majority of research on PD has been in the context of cognitive tasks, it is only more recently that studies have begun to examine the sensitivity of PD to motor-control workload (Fletcher et al., 2017; Jiang et al., 2015; Kahya et al., 2018; Novak et al., 2014; O. White & French, 2017; Wu et al., 2019; Zhang et al., 2016). PD has been shown to exhibit a positive linear relationship with the Index of Difficulty (ID) in a task based on Fitts' paradigm (Jiang et al., 2014, 2015). In addition to the task-evoked response, Jiang and colleagues also found differential pupil dynamics (magnitude and rate of change) with respect to different movement phases. PD has been used to differentiate between the mental workload of two different control-algorithms for myoelectric prostheses (M. M. White et al., 2017; Zhang et al., 2016). PD has also found to be sensitive to task difficulty in robotic and traditional laparoscopic surgery (Di Stasi et al., 2016; Jiang et al., 2015; Wu et al., 2019) and motor control tasks performed using a physically-coupled robotic arm (Novak et al., 2015). A recent study (O. White & French, 2017) showed that PD reduced over multiple trials of practicing a tracking task, which suggested a reduction in mental workload as a result of learning and automaticity. Other studies (Foroughi et al., 2017; Sibley et al., 2011) have also found pupil diameter to decrease over the course of learning, albeit in cognitive tasks. Lastly, there is emerging evidence that baseline PD reflects not only cognitive effort, but also individual differences in working memory capacity and fluid intelligence, i.e. individuals with greater working memory capacities exhibit higher baseline PD (Coyne et al., 2017; Tsukahara et al., 2016). The combined results of the above studies suggest that PD is a potentially sensitive measure of motor-control workload and its variation over learning.

Fixation Duration and Count:

A fixation is defined as a period of time (typically around 200 ms but extending upto several seconds) during which the eye is relatively still (Holmqvist et al., 2011), thus allowing for the acquisition of visual information from the environment. Fixation durations are affected by mental workload, however, the direction of the effect is highly task-dependent (De Rivecourt et al., 2008). For example, increased visual tracking demands can lead to a faster monitoring strategy, resulting in more frequent and shorter fixations (De Rivecourt et al., 2008; Van Orden et al., 2001), whereas the addition of a secondary task or an increase in mental effort spent in interpreting visual information can increase the duration of fixations (Holmqvist et al., 2011). The number of fixations has been found to increase with perceived mental workload in a ‘whack-a-mole’ task using a physically-coupled robotic arm (Novak et al., 2015). Precision demands in a bimanual task led older adults to exhibit longer fixations than younger adults in order to maximize visual input (Coats et al., 2016). The same study also found that older adults performed more look-ahead fixations when precision demands were not present, probably because they required more advance visual information for motor planning. Cognitive interference during standing-balance tasks can cause individuals to exhibit longer and fewer fixations in an effort to maximize visual input and improve balance, and this effect is amplified in older adults (Walsh, 2021).

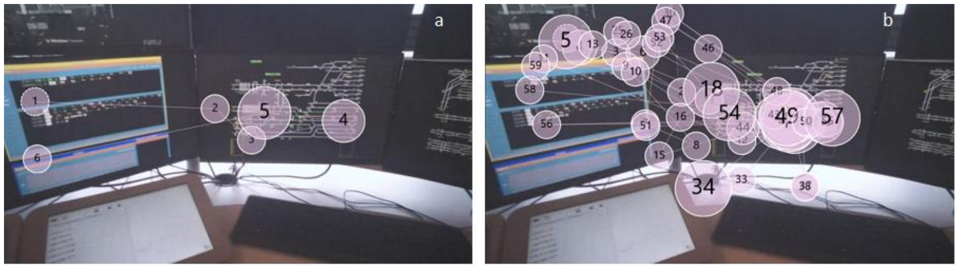


Figure 6. Eye-fixations recorded across a computer workstation. **a)** A scenario with six fixations represented as circles, and numbered in the order of occurrence. The size of each circle/fixation corresponds to its duration **b)** A scenario with a greater number of fixations

Gaze Entropy

Gaze Entropy is a measure of the randomness or complexity of gaze distribution across the visual field (Shiferaw et al., 2019). Entropy provides a way to quantify how an observer samples visual information from the environment, and whether this sampling is widely dispersed and/or erratic (high entropy) or confined and/or ordered (low entropy) (figure 7). These different gaze patterns can be driven by salience of visual stimuli (bottom-up processing), as well as by task difficulty, past experience and physiological state (top-down processing) (R. A. Adams et al., 2013; T. Parr & Friston, 2017). Visual sampling, and consequently, entropy, is consciously optimized in order to maximize perception and performance to meet task demands (Shiferaw et al., 2019). This optimization occurs due to activation of top-down processes (e.g. a particular gaze strategy, or prediction of movement) which modulate the effect of bottom-up influences from the environment (e.g. visual clutter). Thus, entropy can represent the level of effort spent towards efficiently sampling the visual environment for a given set of task demands. Specifically, gaze transition entropy (GTE) is thought to be correlated with the degree of top-down processing, and stationary gaze entropy (SGE) is correlated with bottom-up influence.

Gaze entropy during standing-balance has been found to increase under the influence of a distracting cognitive task, potentially due to a shift from controlled to automatic processing (Walsh, 2021). In contrast, another study found that the addition of a secondary cognitive task during simulated driving reduced GTE (Schieber & Gilland, 2008), indicating reduced exploratory behavior. Thus, different measures of entropy (SGE and GTE) can be differentially affected based on task-specific demands. A study that investigated visuomotor behavior during reaching and grasping tasks using a prosthesis found that the number of transitions between different task-related areas initially increased, but appeared to reduce over multiple sessions (Sobuh et al., 2014). This suggests that goal-directed reaching movements using a cobot may initially produce greater dispersion of gaze fixations (higher SGE), if participants attempt to maximize visual sampling. This dispersion may gradually reduce, or become less random (lower SGE) as users learn to better predict the motion of the robot. Studies have found that SGE was positively correlated with the level of perceived mental workload while operating a surgical robot (Wu et al., 2019, 2021), and also that SGE reduced as performance improved across multiple practice sessions, indicating a learning effect (Wu et al., 2021).

(a) $H_t = 0.89,\ H_s = 1.00$ (b) $H_t = 0.69,\ H_s = 0.98$

Figure 7. Gaze transition entropy (H_t) and stationary gaze entropy (H_s) (in bits), of two different individuals. **a)** An individual exhibiting higher entropy due to frequent gaze-switching between different areas of the painting is shown on the left panel **b)** and an individual exhibiting lower entropy due to more careful viewing of the painting, which manifested as longer fixations in the same areas, is shown on the right panel. Reproduced from (Krejtz et al., 2015)

1.6 Conclusion

Thus, mental workload is an important construct relevant to evaluating the effectiveness of physical human-robot interaction, and its study should account for the influences of learning and individual differences such as age and expertise. Eye-tracking offers great promise as a technique that can provide continuous measures of workload and visuomotor behaviors. Studies in domains other than pHRI have shown that eye-tracking metrics may be sensitive, not only to task-induced workload, but also to the variations in workload that occur over the course of learning and to individual differences. However, there has been relatively little research exploring these relationships in pHRI, and especially little when concerning industrial applications of collaborative robots.

While one of the leading arguments in favor of developing and advancing collaborative robots is that they may diversify the workforce, and help older adults re-enter the workforce later in their lives (Calzavara et al., 2020), we are not aware of any research on investigating differences in performance, mental workload and learning between younger and older populations while using industrial cobots, or including consideration of constructs such as working memory capacity and expertise.

1.7 Study Aims

The **primary goal** of this work is to determine the viability of eye-tracking for measuring variations in mental workload while learning to operate physically-coupled robots. We have chosen eye-tracking as our candidate measure, since it can provide near-instantaneous physiological responses to mental workload, as well as information about visuomotor behaviors relevant to motor performance and learning. While eye-tracking is gaining traction in surgical/rehabilitation robotics domains, systematic investigations of eye tracking for studying interactions with industrial cobots is currently lacking. We propose a comprehensive approach to study the sensitivity of eye-tracking measures to changing mental workload by considering the effects of a variety of task demands (upper- and lower-limb motor tasks), individual differences in age, and expertise, and the effects of learning. Additionally, we also investigate the ability of eye-tracking measures to predict performance, since information regarding when a failure might occur can inform robot-adaptation strategies. This work supports our **long-term goal** of facilitating fluent human-robot interactions through accurate and reliable estimation of operator state.

<u>Aim 1</u>: To quantify the sensitivity and reliability of eye-tracking metrics to variations in mental workload while learning to use an industrial cobot to perform object-manipulation tasks.

<u>Hypothesis</u>: Pupil diameter (PD), Fixation count (FC), Stationary Gaze Entropy (SGE), Gaze Transition Entropy (GTE), and fixations on the manipulated cobot will increase with task difficulty, and reduce over the course of learning to use the cobot, although the response sensitivity and rate of reduction of each measure may be different.

Absolute and relative reliability metrics will be computed and interpreted in the context of existing reliability criteria and findings from previous reliability studies in other domains.

<u>Aim 2</u>: Quantify the extent to which eye-tracking metrics during can predict pHRI task performance (success/failure).

<u>Hypothesis</u>: Eye-tracking measures will be able to predict task-performance more accurately than experimental independent variables.

<u>Aim 3</u>: Quantify the changes in gaze behavior due to the cognitive and attentional demands of using a whole-body powered exoskeleton for performing a routine task (level gait), and compare gaze behaviors between expert and novice users.

<u>Hypothesis</u>: Pupil dilation (PD) and fixations on the walking path (PF) will be higher, and stationary gaze entropy (SGE) will be lower in the EXO condition compared to the no-EXO condition for novice operators, and the size of these effects will be attenuated for expert operators. We expect these measures to trend towards the no-EXO responses over the course of adaptation, as mental workload attenuates over practice.

<u>Aim 4</u>: Investigate age-related differences in mental workload, visuomotor strategies, and learning rates as younger and older adults learn to operate a physically-coupled industrial cobot to perform object manipulations.

<u>*Hypotheses*</u>:

1. Task performance and learning rates are expected to vary according to task difficulty, with no significant age differences (based on prior literature, we expected older adults to be able to perform just as well as younger adults, but with a greater cost, as reflected by mental workload).

2. Older adults are expected to experience elevated mental workload in higher task difficulties (task x group interaction), as evidenced by higher ratings on NASA-TLX, higher pupil dilation (PD), fixation rate (FR), stationary gaze entropy (SGE) and gaze transition entropy (GTE). Measures were expected to change in the direction of reduced workload over the course of learning.

In addition, we also aimed to explore whether robot control or task environmental difficulty had selective effects on the different age groups, and any potential age-related differences in visual attention on different areas of interest, to infer whether younger and older adults utilized different strategies to accomplish the tasks.

References

1. Aaltonen, I., Salmi, T., & Marstio, I. (2018). Refining levels of collaboration to support the design and evaluation of human-robot interaction in the manufacturing industry. *Procedia CIRP*, *72*, 93–98. https://doi.org/10.1016/j.procir.2018.03.214
2. Adams, C. E., & Leverland, M. B. (1985). Environmental and Behavioral Factors That Can Affect Blood Pressure. *The Nurse Practitioner*, *10*(11), 39.
3. Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5
4. Ajoudani, A., Zanchettin, A. M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., & Khatib, O. (2018). Progress and prospects of the human–robot collaboration. *Autonomous Robots*, *42*(5), 957–975. https://doi.org/10.1007/s10514-017-9677-2
5. Anguera, J. A., Reuter-Lorenz, P. A., Willingham, D. T., & Seidler, R. D. (2011). Failure to Engage Spatial Working Memory Contributes to Age-related Declines in Visuomotor Learning. *Journal of Cognitive Neuroscience*, *23*(1), 11–25. https://doi.org/10.1162/jocn.2010.21451
6. Argall, B. D., & Billard, A. G. (2010). A survey of Tactile Human–Robot Interactions. *Robotics and Autonomous Systems*, *58*(10), 1159–1176. https://doi.org/10.1016/j.robot.2010.07.002

7. Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., & Babiloni, F. (2016). Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00539
8. Ariff, G., Donchin, O., Nanayakkara, T., & Shadmehr, R. (2002). A Real-Time State Predictor in Motor Control: Study of Saccadic Eye Movements during Unseen Reaching Movements. *The Journal of Neuroscience*, *22*(17), 7721–7729. https://doi.org/10.1523/JNEUROSCI.22-17-07721.2002
9. Aronson, R. M., & Admoni, H. (2018). *Gaze for Error Detection During Human-Robot Shared Manipulation*. Proceedings of RSS '18 Towards a Framework for Joint Action Workshop.
10. Aronson, R. M., Santini, T., Kübler, T. C., Kasneci, E., Srinivasa, S., & Admoni, H. (2018). *Eye-Hand Behavior in Human-Robot Shared Manipulation*. 10.
11. Bauer, A., Wollherr, D., & Buss, M. (2008). Human–robot collaboration: A survey. *International Journal of Humanoid Robotics*, *05*(01), 47–66. https://doi.org/10.1142/S0219843608001303
12. Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292.
13. Beier, M. E., & Ackerman, P. L. (2005). Age, Ability, and the Role of Prior Knowledge on the Acquisition of New Domain Knowledge: Promising Results in a Real-World Learning Environment. *Psychology and Aging*, *20*(2), 341–355. https://doi.org/10.1037/0882-7974.20.2.341
14. Bequette, B., Norton, A., Jones, E., & Stirling, L. (2020). Physical and Cognitive Load Effects Due to a Powered Lower-Body Exoskeleton. *Human Factors*, *62*(3), 411–423. https://doi.org/10.1177/0018720820907450
15. Bergamasco, M., & Herr, H. (2016). Human–Robot Augmentation. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1875–1906). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_70
16. Bigliassi, M., Karageorghis, C. I., Hoy, G. K., & Layne, G. S. (2019). The Way You Make Me Feel: Psychological and cerebral responses to music during real-life physical activity. *Psychology of Sport and Exercise*, *41*, 211–217. https://doi.org/10.1016/j.psychsport.2018.01.010
17. Binsted, G., Chua, R., Helsen, W., & Elliott, D. (2001). Eye–hand coordination in goal-directed aiming. *Human Movement Science*, *20*(4–5), 563–585. https://doi.org/10.1016/S0167-9457(01)00068-9
18. Bommer, S. C., & Fendley, M. (2018). A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations. *International Journal of Industrial Ergonomics*, *63*, 7–17. https://doi.org/10.1016/j.ergon.2016.10.007
19. Bouwsema, H., van der Sluis, C. K., & Bongers, R. M. (2014). Changes in performance over time while learning to use a myoelectric prosthesis. *Journal of NeuroEngineering and Rehabilitation*, *11*(1), 16. https://doi.org/10.1186/1743-0003-11-16
20. Brookhuis, K. A., & Waard, D. de. (2002). On the assessment of (mental) workload and other subjective qualifications. *Ergonomics*, *45*(14), 1026–1030. https://doi.org/10.1080/00140130210166799
21. Brouwer, A.-M., Zander, T. O., van Erp, J. B. F., Korteling, J. E., & Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: Six recommendations to avoid common pitfalls. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00136
22. Buszard, T., Farrow, D., Verswijveren, S. J. J. M., Reid, M., Williams, J., Polman, R., Ling, F. C. M., & Masters, R. S. W. (2017). Working Memory Capacity Limits Motor Learning When Implementing Multiple Instructions. *Frontiers in Psychology*, *8*, 1350. https://doi.org/10.3389/fpsyg.2017.01350
23. Buszard, T., & Masters, R. S. W. (2018). Adapting, correcting and sequencing movements: Does working-memory capacity play a role? *International Review of Sport and Exercise Psychology*, *11*(1), 258–278. https://doi.org/10.1080/1750984X.2017.1323940

24. Cain, B. (2007). *A Review of the Mental Workload Literature*. 35.
25. Calzavara, M., Battini, D., Bogataj, D., Sgarbossa, F., & Zennaro, I. (2020). Ageing workforce management in manufacturing systems: State of the art and future research agenda. *International Journal of Production Research*, *58*(3), 729–747. https://doi.org/10.1080/00207543.2019.1600759
26. Chadwell, A., Kenney, L., Thies, S., Galpin, A., & Head, J. (2016). The Reality of Myoelectric Prostheses: Understanding What Makes These Devices Difficult for Some Users to Control. *Frontiers in Neurorobotics*, *10*. https://doi.org/10.3389/fnbot.2016.00007
27. Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, *74*, 221–232. https://doi.org/10.1016/j.apergo.2018.08.028
28. Chen, T. L., & Kemp, C. C. (2010). Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 367–374. https://doi.org/10.1109/HRI.2010.5453162
29. Coats, R. O., Fath, A. J., Astill, S. L., & Wann, J. P. (2016). Eye and hand movement strategies in older adults during a complex reaching task. *Experimental Brain Research*, *234*(2), 533–547. https://doi.org/10.1007/s00221-015-4474-7
30. Cornwall, W. (2015). In pursuit of the perfect power suit. *Science*, *350*(6258), 270–273. https://doi.org/10.1126/science.350.6258.270
31. Coyne, J. T., Foroughi, C., & Sibley, C. (2017). Pupil Diameter and Performance in a Supervisory Control Task: A Measure of Effort or Individual Differences? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 865–869. https://doi.org/10.1177/1541931213601689
32. Darzi, A., Gorsic, M., & Novak, D. (2017). Difficulty adaptation in a competitive arm rehabilitation game using real-time control of arm electromyogram and respiration. *2017 International Conference on Rehabilitation Robotics (ICORR)*, 857–862. https://doi.org/10.1109/ICORR.2017.8009356
33. De Rivecourt, M., Kuperus, M. N., Post, W. J., & Mulder, L. J. M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, *51*(9), 1295–1319. https://doi.org/10.1080/00140130802120267
34. de Vries, S., Huys, R., & Zanone, P. G. (2018). Keeping your eye on the target: Eye–hand coordination in a repetitive Fitts' task. *Experimental Brain Research*, *236*(12), 3181–3190. https://doi.org/10.1007/s00221-018-5369-1
35. Deeny, S., Chicoine, C., Hargrove, L., Parrish, T., & Jayaraman, A. (2014). A Simple ERP Method for Quantitative Analysis of Cognitive Workload in Myoelectric Prosthesis Control and Human-Machine Interaction. *PLOS ONE*, *9*(11), e112091. https://doi.org/10.1371/journal.pone.0112091
36. Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., & Lotte, F. (2019). Monitoring Pilot's Mental Workload Using ERPs and Spectral Power with a Six-Dry-Electrode EEG System in Real Flight Conditions. *Sensors*, *19*(6), Article 6. https://doi.org/10.3390/s19061324
37. Di Flumeri, G., Borghini, G., Aricò, P., Sciaraffa, N., Lanzi, P., Pozzi, S., Vignali, V., Lantieri, C., Bichicchi, A., Simone, A., & Babiloni, F. (2018). EEG-Based Mental Workload Neurometric to Evaluate the Impact of Different Traffic and Road Conditions in Real Driving Settings. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00509
38. Di Stasi, L. L., Diaz-Piedra, C., Rieiro, H., Sánchez Carrión, J. M., Martin Berrido, M., Olivares, G., & Catena, A. (2016). Gaze entropy reflects surgical task load. *Surgical Endoscopy*, *30*(11), 5034–5043. https://doi.org/10.1007/s00464-016-4851-8
39. Draheim, C., Pak, R., Draheim, A., & Engle, R. W. (2021). *The role of attention control in complex real-world tasks* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/9ekpu
40. Ebert, D. M., & Henrich, D. D. (2002). Safe human-robot-cooperation: Image-based collision detection for industrial robots. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, *2*, 1826–1831 vol.2. https://doi.org/10.1109/IRDS.2002.1044021
41. Ellmers, T. J., Cocks, A. J., Doumas, M., Williams, A. M., & Young, W. R. (2016). Gazing into Thin Air: The Dual-Task Costs of Movement Planning and Execution during Adaptive Gait. *PLOS ONE*, *11*(11), e0166063. https://doi.org/10.1371/journal.pone.0166063

42. Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Brooks/Cole.
43. Fletcher, K., Neal, A., & Yeo, G. (2017). The effect of motor task precision on pupil diameter. *Applied Ergonomics*, *65*, 309–315. https://doi.org/10.1016/j.apergo.2017.07.010
44. Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision*, *11*(7), 9. https://doi.org/10.1167/11.7.9
45. Foroughi, C. K., Sibley, C., & Coyne, J. T. (2017). Pupil size as a measure of within-task learning. *Psychophysiology*, *54*(10), 1436–1443. https://doi.org/10.1111/psyp.12896
46. Grassmann, M., Vlemincx, E., von Leupoldt, A., Mittelstädt, J. M., & Van den Bergh, O. (2016, June 14). *Respiratory Changes in Response to Cognitive Load: A Systematic Review* [Review Article]. Neural Plasticity; Hindawi. https://doi.org/10.1155/2016/8146809
47. Guerrero, C. R., Fraile Marinero, J. C., Turiel, J. P., & Muñoz, V. (2013). Using “human state aware” robots to enhance physical human–robot interaction in a cooperative scenario. *Computer Methods and Programs in Biomedicine*, *112*(2), 250–259. https://doi.org/10.1016/j.cmpb.2013.02.003
48. Haddadin, S., & Croft, E. (2016). Physical Human–Robot Interaction. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1835–1874). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_69
49. Hale, S., Rose, N. S., Myerson, J., Strube, M. J., Sommers, M., Tye-Murray, N., & Spehar, B. (2011). The Structure of Working Memory Abilities across the Adult Life Span. *Psychology and Aging*, *26*(1), 92–110. https://doi.org/10.1037/a0021483
50. Hedge, J. W., Borman, W. C., & Lammlein, S. E. (2006). *The aging workforce: Realities, myths, and implications for organizations* (pp. v, 203). American Psychological Association. https://doi.org/10.1037/11325-000
51. Hentout, A., Aouache, M., Maoudj, A., & Akli, I. (2019). Human–robot interaction in industrial collaborative robotics: A literature review of the decade 2008–2017. *Advanced Robotics*, *33*(15–16), 764–799. https://doi.org/10.1080/01691864.2019.1636714
52. Hoffman, G. (2019). Evaluating Fluency in Human–Robot Collaboration. *IEEE Transactions on Human-Machine Systems*, *49*(3), 209–218. https://doi.org/10.1109/THMS.2019.2904558
53. Hogervorst, M. A., Brouwer, A.-M., & van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00322
54. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press. http://lup.lub.lu.se/record/1852359
55. Honig, S., & Oron-Gilad, T. (2018). Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.00861
56. Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 311–320. https://doi.org/10.1145/1054972.1055016
57. Jaquess, K. J., Gentili, R. J., Lo, L.-C., Oh, H., Zhang, J., Rietschel, J. C., Miller, M. W., Tan, Y. Y., & Hatfield, B. D. (2017). Empirical evidence for the relationship between cognitive workload and attentional reserve. *International Journal of Psychophysiology*, *121*, 46–55. https://doi.org/10.1016/j.ijpsycho.2017.09.007
58. Jebelli, H., Hwang, S., & Lee, S. (2018). EEG Signal-Processing Framework to Obtain High-Quality Brain Waves from an Off-the-Shelf Wearable EEG Device. *Journal of Computing in Civil Engineering*, *32*(1), 04017070. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000719

59. Jiang, X., Atkins, M. S., Tien, G., Zheng, B., & Bednarik, R. (2014). Pupil dilations during target-pointing respect Fitts' law. *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '14*, 175–182. https://doi.org/10.1145/2578153.2578178
60. Jiang, X., Zheng, B., Bednarik, R., & Atkins, M. S. (2015). Pupil responses to continuous aiming movements. *International Journal of Human-Computer Studies*, *83*, 1–11. https://doi.org/10.1016/j.ijhcs.2015.05.006
61. Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 56–88. https://doi.org/10.1080/14639220210159735
62. Kahya, M., Wood, T. A., Sosnoff, J. J., & Devos, H. (2018). Increased Postural Demand Is Associated With Greater Cognitive Workload in Healthy Young Adults: A Pupillometry Study. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00288
63. Kao, P.-C. (2009). *Principles of Motor Adaptation when Walking with a Powered Exoskeleton*. 120.
64. Kim, S., Srinivasan, D., Nussbaum, M. A., & Leonessa, A. (2021). Human Gait During Level Walking With an Occupational Whole-Body Powered Exoskeleton: Not Yet a Walk in the Park. *IEEE Access*, *9*, 47901–47911. https://doi.org/10.1109/ACCESS.2021.3068836
65. Koren, Y., Mairon, R., Sofer, I., Parmet, Y., Ben-Shahar, O., & Bar-Haim, S. (2020). *Downward Gazing for Steadiness* (p. 2020.02.28.969162). bioRxiv. https://doi.org/10.1101/2020.02.28.969162
66. Koren, Y., Mairon, R., Sofer, I., Parmet, Y., Ben-Shahar, O., & Bar-Haim, S. (2022). Vision, cognition, and walking stability in young adults. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-021-04540-w
67. Kovesdi, C., Spielman, Z., LeBlanc, K., & Rice, B. (2018). Application of Eye Tracking for Measurement and Evaluation in Human Factors Studies in Control Room Modernization. *Nuclear Technology*, *202*(2–3), 220–229. https://doi.org/10.1080/00295450.2018.1455461
68. Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., & Haith, A. M. (2019). Motor Learning. In R. Terjung (Ed.), *Comprehensive Physiology* (1st ed., pp. 613–663). Wiley. https://doi.org/10.1002/cphy.c170043
69. Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., Vilaro, A., & Villalobos, N. (2015). Gaze Transition Entropy. *ACM Transactions on Applied Perception*, *13*(1), 1–20. https://doi.org/10.1145/2834121
70. Krüger, J., Lien, T. K., & Verl, A. (2009). Cooperation of human and machines in assembly lines. *CIRP Annals*, *58*(2), 628–646. https://doi.org/10.1016/j.cirp.2009.09.009
71. Lallee, S., Pattacini, U., Lemaignan, S., Lenz, A., Melhuish, C., Natale, L., Skachek, S., Hamann, K., Steinwender, J., Sisbot, E. A., Metta, G., Guitton, J., Alami, R., Warnier, M., Pipe, T., Warneken, F., & Dominey, P. F. (2012). Towards a Platform-Independent Cooperative Human Robot Interaction System: III An Architecture for Learning and Executing Actions and Shared Plans. *IEEE Transactions on Autonomous Mental Development*, *4*(3), 239–253. https://doi.org/10.1109/TAMD.2012.2199754
72. Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, *26*(1), 51–62. https://doi.org/10.1017/S0952523808080899
73. Lanfranco, A. R., Castellanos, A. E., Desai, J. P., & Meyers, W. C. (2004). Robotic Surgery. *Annals of Surgery*, *239*(1), 14–21. https://doi.org/10.1097/01.sla.0000103020.19595.7d
74. Lavoie, E. B., Valevicius, A. M., Boser, Q. A., Kovic, O., Vette, A. H., Pilarski, P. M., Hebert, J. S., & Chapman, C. S. (2018). Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. *Journal of Vision*, *18*(6), 18–18. https://doi.org/10.1167/18.6.18
75. Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004). Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. *Proceedings of the Eye Tracking Research & Applications Symposium on Eye Tracking Research & Applications - ETRA'2004*, 41–48. https://doi.org/10.1145/968363.968370

76. Lee, T. D., Swinnen, S. P., & Serrien, D. J. (1994). Cognitive Effort and Motor Learning. *Quest*, *46*(3), 328–344. https://doi.org/10.1080/00336297.1994.10484130
77. Leeuwen, P. M. van, Groot, S. de, Happee, R., & Winter, J. C. F. de. (2017). Differences between racing and non-racing drivers: A simulator study using eye-tracking. *PLOS ONE*, *12*(11), e0186871. https://doi.org/10.1371/journal.pone.0186871
78. Li, J., Li, H., Wang, H., Umer, W., Fu, H., & Xing, X. (2019). Evaluating the impact of mental fatigue on construction equipment operators' ability to detect hazards using wearable eye-tracking technology. *Automation in Construction*, *105*, 102835. https://doi.org/10.1016/j.autcon.2019.102835
79. Lotz, A., Russwinkel, N., & Wohlfarth, E. (2020). Take-over expectation and criticality in Level 3 automated driving: A test track study on take-over behavior in semi-trucks. *Cognition, Technology & Work*, *22*(4), 733–744. https://doi.org/10.1007/s10111-020-00626-z
80. Luke, T., Brook-Carter, N., Parkes, A. M., Grimes, E., & Mills, A. (2006). An investigation of train driver visual strategies. *Cognition, Technology & Work*, *8*(1), 15–29. https://doi.org/10.1007/s10111-005-0015-7
81. Maddikunta, P. K. R., Pham, Q.-V., B, P., Deepa, N., Dev, K., Gadekallu, T. R., Ruby, R., & Liyanage, M. (2022). Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration*, *26*, 100257. https://doi.org/10.1016/j.jii.2021.100257
82. Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufacturing*, *3*, 2854–2861. https://doi.org/10.1016/j.promfg.2015.07.783
83. Masters, R., & Maxwell, J. (2008). The theory of reinvestment. *International Review of Sport and Exercise Psychology*, *1*(2), 160–183. https://doi.org/10.1080/17509840802287218
84. Mathot, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), Article 1. https://doi.org/10.5334/joc.18
85. Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The Psychometrics of Mental Workload: Multiple Measures Are Sensitive but Divergent. *Human Factors*, *57*(1), 125–143. https://doi.org/10.1177/0018720814539505
86. Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the Control of Foot Placement When Walking in Natural Terrain. *Current Biology*, *28*(8), 1224-1233.e5. https://doi.org/10.1016/j.cub.2018.03.008
87. Memar, A. H., & Esfahani, E. T. (2020). Objective Assessment of Human Workload in Physical Human-robot Cooperation Using Brain Monitoring. *ACM Transactions on Human-Robot Interaction*, *9*(2), 1–21. https://doi.org/10.1145/3368854
88. Neumann, W. P., Winkelhaus, S., Grosse, E. H., & Glock, C. H. (2021). Industry 4.0 and the human factor – A systems framework and analysis methodology for successful development. *International Journal of Production Economics*, *233*, 107992. https://doi.org/10.1016/j.ijpe.2020.107992
89. Novak, D., Beyeler, B., Omlin, X., & Riener, R. (2014). Passive Brain-Computer Interfaces for Robot-Assisted Rehabilitation. In C. Guger, T. Vaughan, & B. Allison (Eds.), *Brain-Computer Interface Research* (pp. 73–95). Springer International Publishing. https://doi.org/10.1007/978-3-319-09979-8_7
90. Novak, D., Beyeler, B., Omlin, X., & Riener, R. (2015). Workload Estimation in Physical Human–Robot Interaction Using Physiological Measurements. *Interacting with Computers*, *27*(6), 616–629. https://doi.org/10.1093/iwc/iwu021
91. Novak, D., Mihelj, M., Ziherl, J., Olensek, A., & Munih, M. (2011). Psychophysiological Measurements in a Biocooperative Feedback Loop for Upper Extremity Rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *19*(4), 400–410. https://doi.org/10.1109/TNSRE.2011.2160357
92. Parr, J. V. V., Vine, S. J., Harrison, N. R., & Wood, G. (2018). Examining the Spatiotemporal Disruption to Gaze When Using a Myoelectric Prosthetic Hand. *Journal of Motor Behavior*, *50*(4), 416–425. https://doi.org/10.1080/00222895.2017.1363703

93. Parr, J. V. V., Vine, S. J., Wilson, M. R., Harrison, N. R., & Wood, G. (2019). Visual attention, EEG alpha power and T7-Fz connectivity are implicated in prosthetic hand control and can be optimized through gaze training. *Journal of NeuroEngineering and Rehabilitation*, *16*(1), 52. https://doi.org/10.1186/s12984-019-0524-x
94. Parr, T., & Friston, K. J. (2017). Working memory, attention, and salience in active inference. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-15249-0
95. Pasqualotto, E., Matuz, T., Federici, S., Ruf, C. A., Bartl, M., Olivetti Belardinelli, M., Birbaumer, N., & Halder, S. (2015). Usability and Workload of Access Technology for People With Severe Motor Impairment: A Comparison of Brain-Computer Interfacing and Eye Tracking. *Neurorehabilitation and Neural Repair*, *29*(10), 950–957. https://doi.org/10.1177/1545968315575611
96. Prewett, M. S., Johnson, R. C., Saboe, K. N., Elliott, L. R., & Coovert, M. D. (2010). Managing workload in human–robot interaction: A review of empirical studies. *Computers in Human Behavior*, *26*(5), 840–856. https://doi.org/10.1016/j.chb.2010.03.010
97. Raveh, E., Friedman, J., & Portnoy, S. (2018). Visuomotor behaviors and performance in a dual-task paradigm with and without vibrotactile feedback when using a myoelectric controlled hand. *Assistive Technology*, *30*(5), 274–280. https://doi.org/10.1080/10400435.2017.1323809
98. Rietschel, J. C., McDonald, C. G., Goodman, R. N., Miller, M. W., Jones-Lush, L. M., Wittenberg, G. F., & Hatfield, B. D. (2014). Psychophysiological support of increasing attentional reserve during the development of a motor skill. *Biological Psychology*, *103*, 349–356. https://doi.org/10.1016/j.biopsycho.2014.10.008
99. Sailer, U. (2005). Eye-Hand Coordination during Learning of a Novel Visuomotor Task. *Journal of Neuroscience*, *25*(39), 8833–8842. https://doi.org/10.1523/JNEUROSCI.2658-05.2005
100. Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data. *Human Factors*, *49*(3), 347–357. https://doi.org/10.1518/001872007X196685
101. Schieber, F., & Gilland, J. (2008). Visual Entropy Metric Reveals Differences in Drivers' Eye Gaze Complexity across Variations in Age and Subsidiary Task Load. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(23), 1883–1887. https://doi.org/10.1177/154193120805202311
102. Seidler, R. D., Bernard, J. A., Burutolu, T. B., Fling, B. W., Gordon, M. T., Gwin, J. T., Kwak, Y., & Lipps, D. B. (2010). Motor control and aging: Links to age-related brain structural, functional, and biochemical effects. *Neuroscience & Biobehavioral Reviews*, *34*(5), 721–733. https://doi.org/10.1016/j.neubiorev.2009.10.005
103. Seidler, R. D., Bo, J., & Anguera, J. A. (2012). Neurocognitive Contributions to Motor Skill Learning: The Role of Working Memory. *Journal of Motor Behavior*, *44*(6), 445–453. https://doi.org/10.1080/00222895.2012.672348
104. Shaw, E. P., Rietschel, J. C., Hendershot, B. D., Pruziner, A. L., Miller, M. W., Hatfield, B. D., & Gentili, R. J. (2018). Measurement of attentional reserve and mental effort for cognitive workload assessment under various task demands during dual-task walking. *Biological Psychology*, *134*, 39–51. https://doi.org/10.1016/j.biopsycho.2018.01.009
105. Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews*, *96*, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007
106. Sibley, C., Coyne, J., & Baldwin, C. (2011). Pupil Dilation as an Index of Learning. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 237–241. https://doi.org/10.1177/1071181311551049
107. Sibley, C., Foroughi, C. K., Brown, N. L., Phillips, H., Drollinger, S., Eagle, M., & Coyne, J. T. (2020). More than Means: Characterizing Individual Differences in Pupillary Dilations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 57–61. https://doi.org/10.1177/1071181320641017

108. So, W. K. Y., Wong, S. W. H., Mak, J. N., & Chan, R. H. M. (2017). An evaluation of mental workload with frontal EEG. *PLOS ONE*, *12*(4), e0174949. https://doi.org/10.1371/journal.pone.0174949
109. Sobuh, M. M. D., Kenney, L. P. J., Galpin, A. J., Thies, S. B., McLaughlin, J., Kulkarni, J., & Kyberd, P. (2014). Visuomotor behaviours when using a myoelectric prosthesis. *Journal of NeuroEngineering and Rehabilitation*, *11*(1), 72. https://doi.org/10.1186/1743-0003-11-72
110. Srinivasan, D., & Martin, B. J. (2010). Eye–hand coordination of symmetric bimanual reaching tasks: Temporal aspects. *Experimental Brain Research*, *203*(2), 391–405. https://doi.org/10.1007/s00221-010-2241-3
111. Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*, 33. https://doi.org/10.1145/1121241.1121249
112. Stirling, L., Kelty-Stephen, D., Fineman, R., Jones, M. L. H., Daniel Park, B.-K., Reed, M. P., Parham, J., & Choi, H. J. (2020). Static, Dynamic, and Cognitive Fit of Exosystems for the Human Operator. *Human Factors*, *62*(3), 424–440. https://doi.org/10.1177/0018720819896898
113. Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A Systematic Review of Physiological Measures of Mental Workload. *International Journal of Environmental Research and Public Health*, *16*(15), 2716. https://doi.org/10.3390/ijerph16152716
114. Tinga, A. M. (2019). Non-invasive neurophysiological measures of learning_ A meta-analysis. *Neuroscience and Biobehavioral Reviews*, 31.
115. Tinga, A. M., de Back, T. T., & Louwerse, M. M. (2020). Non-invasive Neurophysiology in Learning and Training: Mechanisms and a SWOT Analysis. *Frontiers in Neuroscience*, *14*, 589. https://doi.org/10.3389/fnins.2020.00589
116. Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In *Handbook of human factors and ergonomics, 3rd ed* (pp. 243–268). John Wiley & Sons, Inc. https://doi.org/10.1002/0470048204.ch9
117. Tsukahara, J. S., Harrison, T. L., & Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology*, *91*, 109–123. https://doi.org/10.1016/j.cogpsych.2016.10.001
118. Van Acker, B. B., Bombeke, K., Durnez, W., Parmentier, D. D., Mateus, J. C., Biondi, A., Saldien, J., & Vlerick, P. (2020). Mobile pupillometry in manual assembly: A pilot study exploring the wearability and external validity of a renowned mental workload lab measure. *International Journal of Industrial Ergonomics*, *75*, 102891. https://doi.org/10.1016/j.ergon.2019.102891
119. van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y
120. Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T.-P. (2001). Eye Activity Correlates of Workload during a Visuospatial Memory Task. *Human Factors*, *43*(1), 111–121. https://doi.org/10.1518/001872001775992570
121. Walsh, G. S. (2021). Visuomotor control dynamics of quiet standing under single and dual task conditions in younger and older adults. *Neuroscience Letters*, *761*, 136122. https://doi.org/10.1016/j.neulet.2021.136122
122. White, M. M., Zhang, W., Winslow, A. T., Zahabi, M., Zhang, F., Huang, H., & Kaber, D. B. (2017). Usability Comparison of Conventional Direct Control Versus Pattern Recognition Control of Transradial Prostheses. *IEEE Transactions on Human-Machine Systems*, *47*(6), 1146–1157. https://doi.org/10.1109/THMS.2017.2759762
123. White, O., & French, R. M. (2017). Pupil Diameter May Reflect Motor Control and Learning. *Journal of Motor Behavior*, *49*(2), 141–149. https://doi.org/10.1080/00222895.2016.1161593
124. Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

125. Wickens, C. D., & Derrick, W. (1981). *The Processing Demands of Higher Order Manual Control: Application of Additive Factors Methodology*. ILLINOIS UNIV AT URBANA DEPT OF PSYCHOLOGY. https://apps.dtic.mil/sti/citations/ADA098077
126. Wilmut, K., Wann, J. P., & Brown, J. H. (2006). How active gaze informs the hand in sequential pointing movements. *Experimental Brain Research*, *175*(4), 654–666. https://doi.org/10.1007/s00221-006-0580-x
127. Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*(12), Article 12. https://doi.org/10.1038/nrn3112
128. Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931
129. Wu, C., Cha, J., Sulek, J., Sundaram, C. P., Wachs, J., Proctor, R. W., & Yu, D. (2021). Sensor-based indicators of performance changes between sessions during robotic surgery training. *Applied Ergonomics*, *90*, 103251. https://doi.org/10.1016/j.apergo.2020.103251
130. Wu, C., Cha, J., Sulek, J., Zhou, T., Sundaram, C. P., Wachs, J., & Yu, D. (2019). Eye-Tracking Metrics Predict Perceived Workload in Robotic Surgical Skills Training. *Human Factors*, 0018720819874544. https://doi.org/10.1177/0018720819874544
131. Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17. https://doi.org/10.1080/00140139.2014.956151
132. Young, W. R., Wing, A. M., & Hollands, M. A. (2012). Influences of State Anxiety on Gaze Behavior and Stepping Accuracy in Older Adults During Adaptive Locomotion. *The Journals of Gerontology: Series B*, *67B*(1), 43–51. https://doi.org/10.1093/geronb/gbr074
133. Zhang, W., White, M., Zahabi, M., Winslow, A. T., Zhang, F., Huang, H., & Kaber, D. (2016). Cognitive workload in conventional direct control vs. Pattern recognition control of an upper-limb prosthesis. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 002335–002340. https://doi.org/10.1109/SMC.2016.7844587
134. Zhu, Q., Wei, P., Shi, Y., & Du, J. (2020). Cognitive Benefits of Human-Robot Collaboration in Complex Industrial Operations: A Virtual Reality Experiment. *Construction Research Congress 2020*, 129–138. https://doi.org/10.1061/9780784482858.015
135. Zhu, Q., Zhou, T., & Du, J. (2022). Upper-body haptic system for snake robot teleoperation in pipelines. *Advanced Engineering Informatics*, *51*, 101532. https://doi.org/10.1016/j.aei.2022.101532

Chapter 2. Eye-Tracking in Physical Human-Robot Interaction: Mental Workload and Performance Prediction

Abstract

In Physical Human-Robot Interaction (pHRI), the need to learn the robot's motor-control dynamics is associated with increased cognitive load. Eye-tracking metrics can help understand the dynamics of fluctuating mental workload over the course of learning. The aim of this study was to test eye-tracking measures' sensitivity and reliability to variations in task difficulty, as well as their performance-prediction capability, in physical human-robot collaboration tasks involving an industrial robot for object co-manipulation. Participants (9M,9F) learned to co-perform a virtual pick-and-place task with a bimanual robot over multiple trials. Joint stiffness of the robot was manipulated to increase motor-coordination demands. The psychometric properties of eye-tracking measures and their ability to predict performance was investigated. We found that stationary gaze entropy and pupil diameter were the most reliable and sensitive measures of workload associated with changes in task difficulty and learning. Increased task difficulty was more likely to result in a robot-monitoring strategy. Eye-tracking measures were able to predict the occurrence of success or failure in each trial with 70% sensitivity and 71% accuracy. Overall, the sensitivity and reliability of eye-tracking measures was acceptable, although values were lower than those observed in cognitive domains. Measures of gaze behaviors indicative of visual monitoring strategies were most sensitive to task difficulty manipulations, and should be explored further for the pHRI domain where motor-control and internal-model formation will likely be strong contributors to workload. Future collaborative robots can adapt to human cognitive state and skill-level measured using eye-tracking measures of workload and visual attention.

Keywords: Collaborative Robot, Virtual Reality, Gaze Entropy, Areas of Interest, Psychometrics, Performance Prediction

2.1 Introduction

Collaborative robots (cobots) are becoming more usable, versatile, and increasingly safe to operate in close proximity with humans (Haddadin & Croft, 2016). In Physical Human-Robot Interaction (pHRI), although cobots have achieved higher standards of safety and compliance in recent years, they can still impose a significant workload on the user's attentional and cognitive-motor resources (Marchand et al., 2021; Stirling et al., 2020), and may require time and effort to learn (Aronson et al., 2018; Cornwall, 2015). Hence, it is important to understand the cognitive challenges involved in controlling these complex devices and be able to predict the consequences of such cognitive challenges on performance.

Past research has shown that performing joint actions with a robot, e.g., using wearable robots such as myoelectric prostheses and powered exoskeletons or using joystick-operated robotic arms, tends to pose cognitive challenges because the user cannot easily predict the device's control dynamics (Aronson et al., 2018; Chadwell et al., 2016; Cornwall, 2015; Kao, 2009). In other words, users may find it difficult to develop an internal mental model of the cobot and hence not be able to anticipate the consequences of a joint action with the robot, thereby exhibiting an increased reliance on vision to monitor robot behavior. Human motor control literature has theorized that learning to use novel or complex tools such as cobots requires the formation of new internal models for tool behaviors, as well as updating of the internal mental models of the limb controlling the tool (Wolpert et al., 2011). Additionally, using cobots for co-manipulation tasks (i.e., tasks where the human and robot cooperate to manipulate shared objects) may also be challenging because of the need to monitor the manipulated object as well as the surrounding environment for hazards or potential collisions (Steinfeld et al., 2006). Lastly, co-manipulation tasks can be intrinsically difficult, due to the need to remember multiple task steps with conditional relationships, and needing to perform complex spatial transformations mentally (Van Acker et al., 2020) or deal with time pressures (Bommer & Fendley, 2018).

Thus, it is expected that there is a high mental workload present during initial stages of learning to perform joint actions with a cobot, as the user attempts to build an internal model of the device and movement (Sailer, 2005). Over the course of practice, mental workload is expected to attenuate due to refinement of neural processes and increasing automaticity in the task (Sailer, 2005; O. White & French, 2017). Understanding the dynamics of mental workload over the course of motor learning and continuous measurement of these constructs can help in designing learning/training protocols and help minimize workload for users of cobots. Eye-tracking is a promising technique for measuring mental workload and predicting performance, since it can provide both physiological measures, (e.g., pupil dilation) that correlate with the involuntary neural response to mental workload (Just et al., 2003), as well as eye-movement measures (e.g. fixation rate), which reflect voluntary gaze behavior and strategies for maximizing performance (Land, 2009; Srinivasan & Martin, 2010). Eye-tracking measures have also been shown to change over the course of learning and to be able to classify distinct stages of motor-skill-acquisition (Sailer, 2005). In surgical tasks, the ability to focus on the most informative anatomical regions is an important determinant of expertise (Law et al., 2004; Zheng et al., 2021), which suggests that gaze patterns may correlate with, and predict task performance. The versatility of eye-tracking, coupled with its increasing wearability and ubiquity, make eye-tracking a viable technology to implement in dynamic, real-world environments (Cognolato et al., 2018).

Previous research on the psychometric properties of eye-tracking measures has largely focused on cognitive tasks involving change-detection or working memory (Matthews et al., 2015; Zargari Marandi et al., 2018). In the pHRI domain involving tasks requiring motor skills, although some studies have been conducted on changes in specific eye-tracking measures, the psychometric properties of eye tracking metrics are yet to be established as there have been no systematic investigations of the sensitivity and reliability of eye tracking in predicting changes in mental workload and/or performance. Studies in the domains of assistive and surgical robotics have found that higher mental workload increased pupil diameter (Aronson et al., 2018; M. M. White et al., 2017) and stationary gaze entropy (SGE) (Wu et al., 2019), and reduced the fixation rate (Novak et al., 2015). Additionally, SGE and the gaze transitions between areas of interest have been found to reduce over the course of learning (Sobuh et al., 2014; Wu et al., 2021). Recent work has also begun to explore the use of eye-tracking measures to predict performance in the pHRI domain (Aronson et al., 2018; Wu et al., 2021). Using machine-learning techniques, Aronson and colleagues were able to associate distinct scanning behaviors with different control modes of an assistive robot (Aronson et al., 2018). Other work by the same authors discussed the potential ability of gaze behavior to predict unexpected performance conditions or errors, based on which a cobot could take corrective actions (Aronson & Admoni, 2018). In robotic surgery, which requires efficient visual scanning to identify key anatomical features, stationary gaze entropy was found to predict task performance improvements (Wu et al., 2021). This finding supports research that found that gaze behavior could discriminate between expert and novice surgical performance (Law et al., 2004; Wilson et al., 2010; Zheng et al., 2021).

A better and more systematic understanding of the reliability of these different eye tracking metrics, and their sensitivity to variations in task difficulty and learning, can help establish the potential utility of eye-tracking metrics as a continuous measure of human cognitive state in human-cobot interaction, thereby guiding online/adaptive control of robotic systems to be responsive to human state. A combination of pupillometric and gaze-behavior measures may also be able to effectively predict task performance in pHRI. If imminent failure can be predicted by eye-tracking metrics, the adaptability of robotic systems can be improved by enabling the design of anticipatory safety mechanisms.

Thus, this study aimed to quantify the sensitivity and reliability of eye-tracking measures of workload to variations in task difficulty and learning in a pHRI task. The ability of gaze-behavior measures to distinguish different visual strategies across task difficulty and learning, and predict task performance, were also studied. We designed a task that required participants to control a bimanual cobot to pick and place virtual objects at different target locations, while avoiding collisions with virtual objects. The task was timed, and participants performed it under two different levels of task difficulty. Several eye-tracking metrics such as pupil dilation, fixation count, fixations in different areas of interest (AOI), and gaze entropies were computed. Based on the literature reviewed, we expected that these metrics would provide different types of information related to mental workload. Specifically, pupil dilation and fixation count indicate overall workload and visual monitoring demand respectively. Stationary gaze entropy (SGE), accounting for the fixation distribution across different regions of the environment, indicates the need for visual monitoring of the task environment (to detect/avoid collisions in our task). Gaze transition entropy (GTE) can account for the transitions between different areas of the environment, and hence indicate the potential emergence of repetitive scanning patterns over the course of learning. Fixations in the different AOI extracted from eye tracking can be used to infer internal model formation. Since fixations are typically directed towards the targets in goal-directed movements, increased fixations on the manipulated cobot (vs. targets) would indicate an increased reliance on vision for controlling the cobot.

Considering these aspects of eye-tracking measures, we proposed the following aims and hypotheses. First, we quantified the sensitivity of eye-tracking metrics in a human-robot co-manipulation task, by studying how human participants learned to use a cobot under two different levels of task difficulty. We hypothesized that SGE, GTE, pupil dilation and fixation count would be higher in the more difficult task condition, and these metrics would decrease over time (with learning) in both tasks, although the rate of change would depend on task difficulty. Next, we estimated the reliability of the different eye-tracking metrics, to provide a more detailed characterization of their effectiveness as workload measures in pHRI. As a third aim, we explored the changes in visual strategies (in terms of the relative visual focus on different AOI in the environment) over the course of learning to use the robot. Fixations on the manipulated cobot were expected to be higher in the high-difficulty task condition and reduce over the course of learning in both conditions. Finally, we explored the extent to which eye-tracking metrics during could predict pHRI task performance on a trial-to-trial basis.

2.2 Methods

2.2.1 Experimental setup

The inspiration for our task environment is a potential industrial scenario in which a human operator uses a physically-coupled robot to lift heavy objects (perhaps oddly shaped ones) and place them in assigned target locations. The task was simulated in virtual reality (VR). Participants performed the task using the

Rethink Robotics Baxter Robot, which has two 7-DOF arms that can be manipulated freely by grasping their wrists. Custom handles were 3-D printed and attached to the robot wrists to enable a secure and comfortable grasp. We visualized a virtual model of Baxter inside a Unity VR environment (Unity version 2020.1.9f1) by sending Baxter's real-time joint positions to Unity at 140 Hz using the ROS# package (Zhou et al., 2020). The schematic for this process is shown in Figure 8a. The HTC Vive Pro Eye VR headset was used for rendering the virtual task and environment in this study. Participants stood close to, and facing, the Baxter robot while wearing the VR headset, to co-perform virtual object manipulations without needing any large head movements to scan their environment (Figure 8b). They experienced a first-person view of the task (VR visualization shown in Figure 8c. Participants manipulated a virtual plate with a ball on it, which simulated physical properties of an actual ball such as rolling around inside the plate and falling out if tilted. When cued by the experimenter, participants had to pick and transfer the object (plate with ball) on to a selected target location (any one of three target towers chosen randomly), without colliding with any virtual objects, within 10 seconds. In the case of any errors (collision, ball drop, or excessive time), the trial would be marked incomplete, and the object would be reset at the starting location.

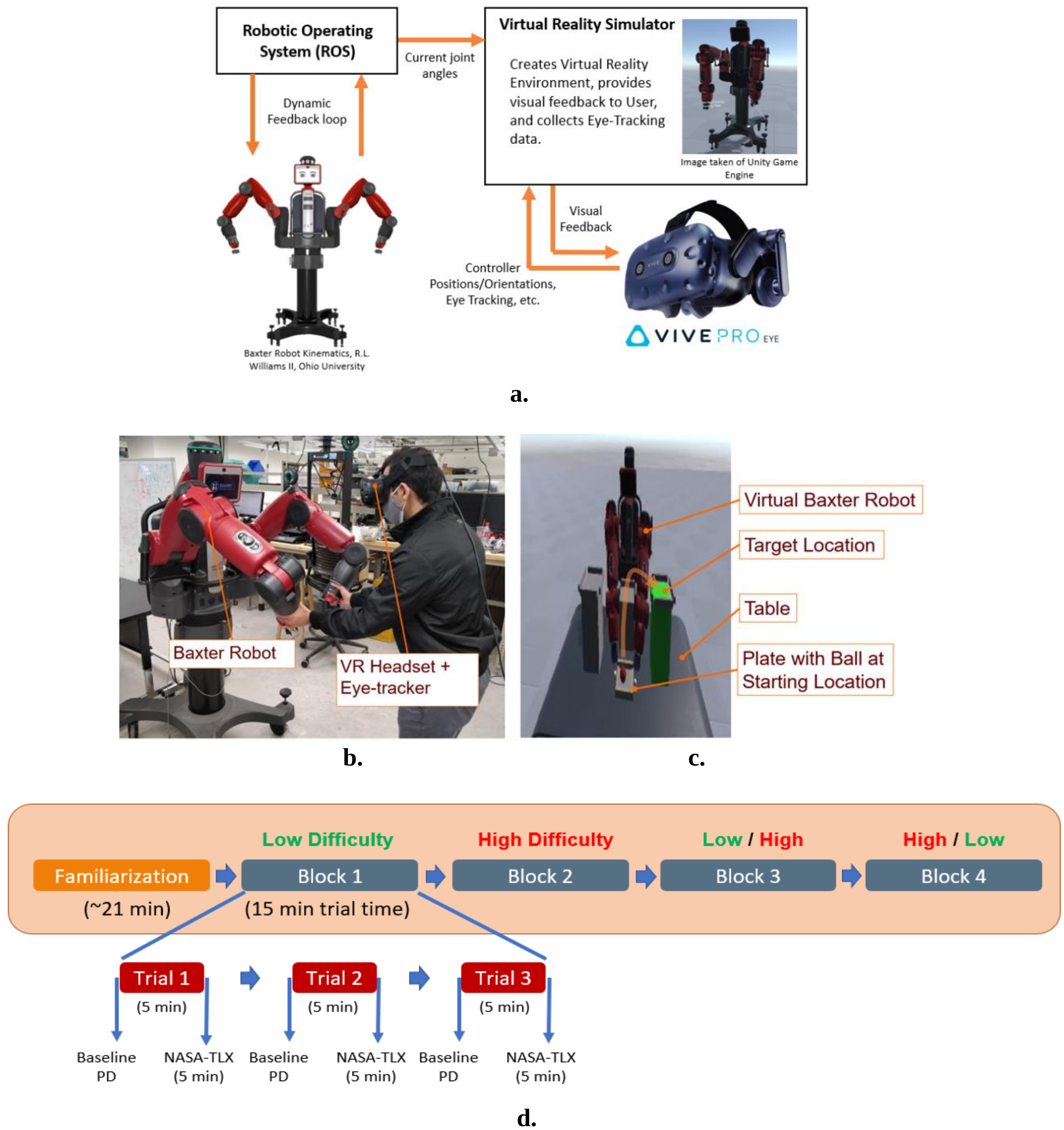


Figure 8. **(a)** Schematic of custom-built communication module between Baxter and VR; **(b)** Participant wearing the VR headset and operating the robot; **(c)** Participant's view in VR, with robot, plate, and target locations; **(d)** Experimental protocol showing familiarization followed by 4 blocks of 12 experimental trials. Baseline PD was obtained before, and NASA-TLX after, each trial.

2.2.2 Participants

A convenience sample of 24 participants was recruited from the local community. Data from 6 participants were not included in the analysis due to data quality issues, resulting in a total of 18 participants (9M, 9F) for data analysis. Their mean age was 26.8 years (SD 4 years). Participants were included if they could read at arm's length without the use of corrective lenses, and were free of any recent history of musculoskeletal disorders (past 12 months). Individuals with a history of migraine, vertigo and epilepsy were excluded, since these conditions can increase the susceptibility to VR sickness. Participants signed a written informed consent, and the research was approved by the Virginia Tech Institutional Review Board (#21-203).

2.2.3 Experiment Design and Protocol

The two independent variables in the experiment were task difficulty and trial number. Task difficulty was manipulated by changing the degree of "match" between the joint impedances of the two arms of the robot, with two levels, low difficulty (LD; matched impedances) and high difficulty (HD; mis-matched impedances). The stiffness and damping parameters of the Baxter robot were used to vary joint impedances. The mis-matched HD condition was created by stiffening some degrees of freedom on one Baxter arm, thereby limiting the arm's range of motion (ROM). This made the control of the robot less intuitive and increased motor coordination demands needed to bimanually balance the plate and ball. Joint names and spring and damping coefficient values are shown in Figure 9 and table 1. In a first familiarization session (figure 8d), participants learnt to control the Baxter robot, put on the VR headset and practice the individual components of the main task. Familiarization was followed by the experimental session that included four blocks of trials with three trials of 5 minutes each. The first and second blocks were always performed in the order of LD followed by HD, to avoid transfer of learning effects. The order of difficulty (LD vs. HD) was randomized for the third and the fourth blocks (counterbalanced across participants).

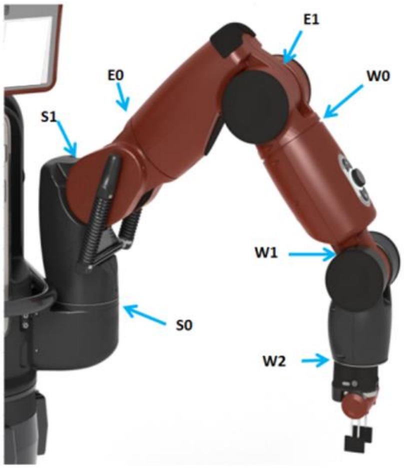


Figure 9. Joint names of the Baxter robot-arms. (https://sdk.rethinkrobotics.com/wiki/Joint_Position_Example)

Table 1. Spring stiffness (K) and Damping coefficient (C) of each joint for both arms and difficulty conditions.

	Low Difficulty (Matched)				High Difficulty (Mis-matched)			
	Left Arm		Right Arm		Left Arm		Right Arm	
Joint Name	**K**	**C**	**K**	**C**	**K**	**C**	**K**	**C**
S0	3	1	3	1	30	10	14.7	5
S1	2	0.75	2	0.75	3	0.75	3	0.75
E0	1.5	0.75	1.5	0.75	15	7.5	7.5	3.75
E1	1.5	0.5	1.5	0.5	0.45	0.15	1.5	0.5
W0	0	0	0	0	0.6	0.1	0.09	0.015
W1	0.4	0.15	0.4	0.15	0.2	0.1	0.4	0.15
W2	4	1	4	1	4	1	4	1

2.2.4 Data Collection and Processing

Although the sampling frequency of the eye-tracker was set at 90 Hz, the actual sampling frequency depended on the graphics rendering speed of the computer and was hence variable over time (Llanes-Jurado et al., 2020). Hence, data were uniformly resampled at 90 Hz using a custom MATLAB script. Next, all samples which corresponded to a pupil diameter value of -1 (which meant that the eyes were not tracked for those times) were removed. Any remaining gaps in the data, except those larger than 100 ms were linearly interpolated (Hessels et al., 2017).

Eye-movement Classification

We classified the gaze samples into fixations using the I-VDT velocity-dispersion algorithm (Komogortsev & Karpov, 2013). Prior to applying the algorithm, we computed the visual angle between successive gaze points using the scalar product of the gaze point vectors. Then, we obtained angular velocity by applying the Savitzky-Golay filter to the visual angle data (Holmqvist et al., 2011). Finally, as specified in the I-VDT algorithm, gaze points with an angular velocity greater than 75 deg/s were classified as saccades, and the remaining data were classified into fixations using a dispersion threshold of 1 deg and a minimum fixation duration of 80 ms (Aronson et al., 2018). When the participant looked at any VR object (also referred to as an area of interest or AOI), the gaze point intersected with the surface of the object, thus registering as a 'hit' on the AOI. For each fixation, we recorded the most frequently-hit AOI as the "fixated-AOI", which resulted in a time series of successive fixated-AOIs for each trial (an example is provided in Appendix A, in the section on entropy computation).

Metrics Computation

The script also recorded the 3-D coordinates of the virtual plate, ball, and the virtual Baxter's end-effectors, as well as the instantaneous timer value corresponding to transfer completion. To quantify performance, each transfer was first categorized as a ***success*** if the plate was transferred to the target within 10-s or a ***failure*** if either a drop or collision occurred, or if the 10-s timer ran out before the plate was transferred. Numbers of successes and failures were computed per 30-s interval and expressed as a proportion of the total number of transfers in each 30-s interval. <u>***Proportions of successes and failures***</u> were used as measures of task performance. <u>***Fixation count***</u> was calculated as the number of fixations per 30-s interval during the trial. We computed the <u>***median pupil dilation (PD)***</u> of the left eye from each 30 s interval, by subtracting the pupil diameter at baseline from the raw pupil diameter during the 30s interval. Baseline pupil diameter was computed as the average pupil diameter across a 10s period during which the participant fixated on the top surface of the center target-tower. <u>***SGE***</u> was computed using the following equation adapted from (B. A. Shiferaw, 2018):

$$H_s(x) = -\sum_{i=1}^{N} p(i) \times \log_2 p(i) \qquad (1)$$

where $H_s(x)$ is the SGE value for a particular time-bin 'x' (equal to 30s), 'i' represents the successive AOI-fixations in the interval 'x' and 'p' is the proportion of fixations on the i^{th} AOI in the interval 'x'. For computing ***GTE***, we first generated a transition matrix for each 30-second interval, in which the element at row 'i' and column 'j' represented the probability of transitions from object 'i' to object 'j' within the 30-second interval. GTE was computed using the following equation adapted from (B. A. Shiferaw, 2018):

$$H_g(x) = -\sum_{i=1}^{N} p(i) \left[\sum_{j=1}^{N} p(i|j) \times \log_2 p(i|j)\right] \qquad (2)$$

where $H_g(x)$ is the value of GTE, 'p(i)' is the stationary distribution of fixation locations, and 'p(i | j)' is the probability of transitioning to AOI 'j' given current position of 'i'. SGE and GTE were both normalized to the maximum theoretical entropy for each 30-s interval, equal to $\log_2 N$ (B. Shiferaw et al., 2019). A more detailed description of the entropy computation is provided in Appendix A. Lastly, we also computed the ***ratio of fixations*** on the following AOI - the plate (which included the ball), the robot-arms, and the top surface of the target pillars, per 30-second interval, in order to understand the relative attentional focus directed towards these AOI. These AOI were selected because they were directly related to participants' goals and actions, and hence were critical components of the task. Fixations on the table and the floor were coded as "miscellaneous" fixations that were not task-relevant.

2.2.5 Statistical Analysis

Effect of task difficulty and learning on Performance and Workload: A mixed factor ANOVA was used to test for the effects of difficulty Condition (LD, HD), Gender (male, female), Trial (1 to 6 for each condition) and their interactions on performance, eye-tracking metrics, and NASA-TLX mental workload ratings. Significant effects were followed by post hoc pairwise comparisons using Tukey's HSD test. Sensitivity of eye-tracking metrics was estimated using effect size (partial eta-squared η_p^2) for main and interaction effects.

<u>Reliability analysis</u>: The standard error of measurement (SEM), as an index of absolute reliability, and the intra-class correlation coefficient (ICC), as an index of relative reliability, were computed using the following procedures:

The total variability of the data set was partitioned according to the mixed effects model (Searle et al., 2009):

$$E_{cstr} = \mu + \beta_c + \gamma_{tr} + \alpha_s + e_{cstr} \qquad (3)$$

where E_{cstr} is the measured value of an eye-tracking measure for a specific condition *c* in trial *tr* of subject *s*; μ is the grand mean; β_c is the fixed effect due to condition (task difficulty); γ_{tr} is the fixed effect of the learning trial, α_s is the random effect of subject; and e_{cstr} is the residual. Both random effects (α_s, e_{cstr}) were assumed to be independently and identically distributed, to have zero covariance between any pair of values and to have a mean of zero. The mixed-effects model was resolved using a REML procedure in JMP® Pro (Version 14, SAS Institute Inc., Cary, NC) to estimate the variance between subjects (S^2_{BS}, i.e., the variance of α_s) and within subject (S^2_{WS}, i.e., the variance of e_{cstr}) with 95% confidence intervals. These variance components were then used to calculate reliability metrics.

The standard error of measurement (SEM) was calculated as the square-root of the within-subjects variance. ICC was computed using the following equation (Shrout & Fleiss, 1979):

$$ICC = \frac{S^2_{BS}}{S^2_{BS} + S^2_{WS}} \qquad (4)$$

<u>Effect of task difficulty and learning on AOI-measures</u>: To understand differences in visual behavior and strategies, the same mixed factor ANOVA described above used to test for the effects of our independent variables on AOI-measures.

<u>Performance prediction using eye-tracking metrics</u>: Two logistic regression models were generated to predict the occurrence of a success (0) or failure (1) – the first model included only the independent

variables of the experiment (condition, Trial, and their interaction) as predictor variables (for comparison purposes), whereas the second included each participant's eye-tracking and AOI-based metrics as predictor variables. Confusion matrices were generated for both models, along with measures of classifier performance (accuracy, precision, recall, specificity, negative predictive value (NPV) and F1-score) (Webb, 2010).

For all analyses, the significance level was set at α = 0.05, and all statistical analyses were performed in JMP Pro (version 16.0.0, SAS Institute Inc., USA).

2.3 Results

2.3.1 Performance and perceived workload

The proportion of successes was significantly lower in the HD condition ($F(1, 2121) = 34.6$; $p < 0.0001$, as seen in Figure 10(a). Additionally, successes increased significantly with time (trials) ($F(5, 2121) = 19.5$; $p < 0.0001$), indicating that participants learnt to better perform the task. There was no significant interaction effect of condition and trial on performance, suggesting that participants improved at a similar rate in both conditions. Females exhibited significantly lower performance than males: ($F(1, 16) = 14.88$; $p = 0.0014$). NASA-TLX mental workload ratings were significantly higher in the HD condition compared to the LD condition, ($F(1, 2121) = 152.9$; $p<0.0001$), and ratings for both conditions reduced over the course of learning ($F(5, 2121) = 67.5$; $p < 0.0001$) accompanied by a significant Condition x Trial interaction effect ($F(5, 2121) = 10.08$; $p<0.0001$). Ratings were not significantly different across genders.

2.3.2 Sensitivity of eye-tracking metrics to changes in task difficulty and learning

All eye tracking metrics showed significant differences across condition (Figure 10). PD, FC, and SGE increased in the HD condition compared to the LD condition, and GTE was reduced in the HD condition compared to the LD condition. The effect of Trial was also significant on PD, SGE and FC. PD

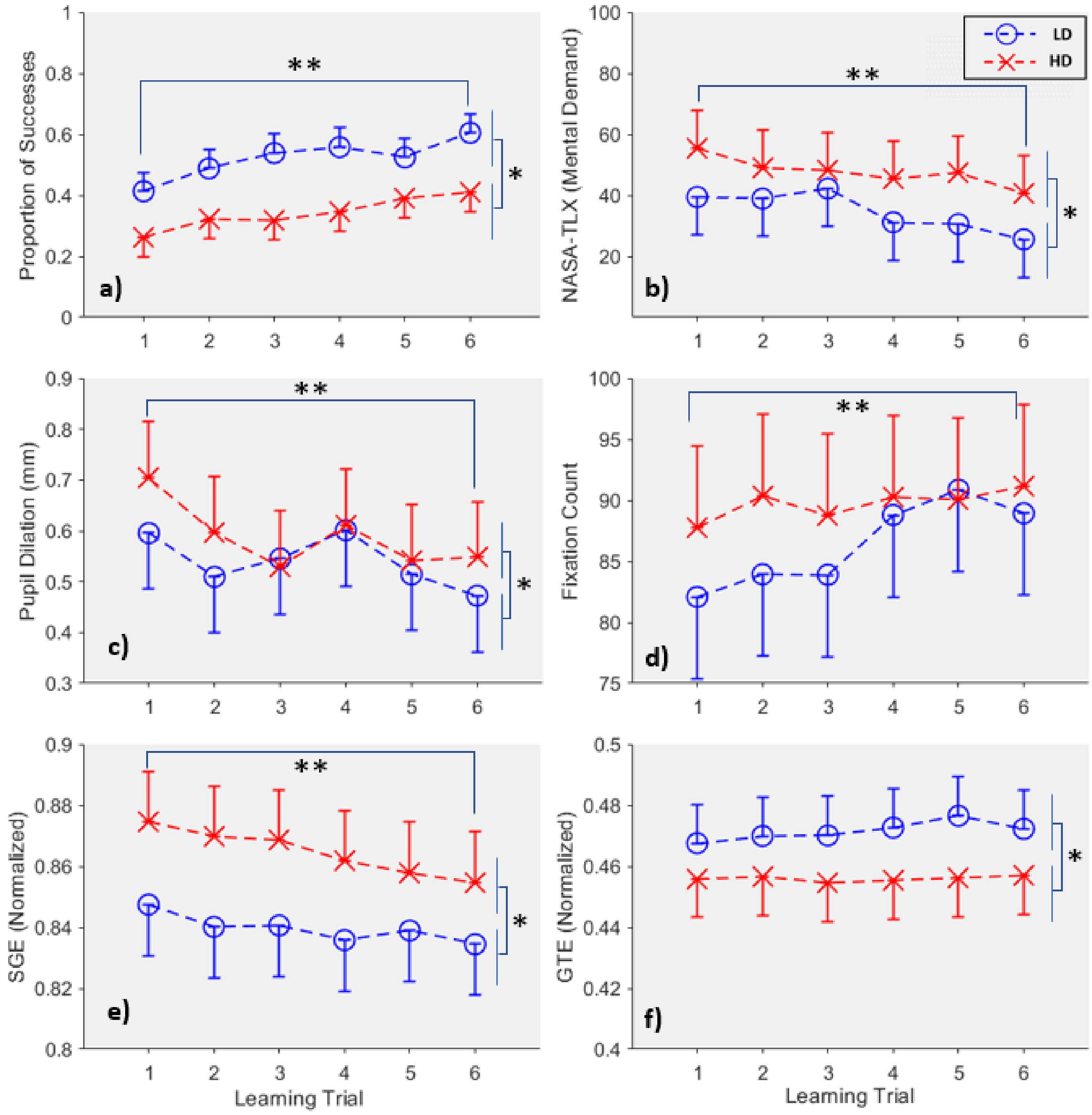


Figure 10. (a) Performance **(b)** NASA-TLX ratings on mental demand, and **(c-f)** eye-tracking workload metrics over the course of six learning trials for each task condition (low difficulty LD, high difficulty HD). Individual data points represent least squares means, and the error bars represent 1 standard error.

<u>**Note**</u>: * denotes a significant main effect of Condition, and ** denotes a significant main effect of Trial. Significant main effects of Condition and Trial were seen in Performance **(a)** and NASA-TLX ratings **(b)**. Among the eye-tracking metrics, pupil dilation, SGE and GTE **(c,e,f)** showed significant main effects of Condition, while pupil dilation, fixation count and SGE **(c-e)** showed significant main effects of Trial.

and SGE reduced, and FC increased across successive trials, although the change in PD was non-monotonic (Figure 10c). The Condition x Trial interaction effect was significant on PD and FC. There was no effect of Gender on eye-tracking workload metrics. The F-statistics, p-values and effect sizes for these statistical comparisons are shown in table 2.

2.3.3 Reliability of eye-tracking metrics

Table 2 also shows the participant variance components and ICC values for each metric. Based on the criteria stated in (Becser et al., 1998; Ettinger et al., 2003; Zargari Marandi et al., 2018), all eye-tracking metrics showed good reliability (>0.4) except GTE, which showed poor reliability. Pupil dilation showed the highest reliability (0.68), followed by SGE (0.44), fixation count (0.41), and GTE (0.17).

Table 2. F-statistics, sensitivity, variance components, and reliability (ICC and SEM) of eye tracking metrics. Sensitivity is quantified using partial eta-squared (η_p^2) as a measure of effect size. Bold fonts indicate significant effects ($p < 0.05$). The 95% lower and 95% upper confidence limits are shown in square brackets.

Workload Measures	F-statistic p-value (Effect Size - η_p^2)			Participant Variance Components		Reliability Metrics	
	Condition	Trial	Condition x Trial	Between (S_{Bs}^2) [LCI-UCI]	Within (S_{Ws}^2) [LCI-UCI]	ICC	SEM
Pupil Dilation (mm)	**F (1, 2121) = 41.26 <0.0001 (0.019)**	**F (5, 2121) = 39.44 <0.0001 (0.085)**	**F (5, 2121) = 8.3 <0.0001 (0.019)**	0.054 [0.0128, 0.0661]	0.026 [0.024, 0.027]	0.68	0.16
Fixation Count	**F (1, 2121) = 10.98 0.0009 (0.005)**	**F (5, 2121) = 7.01 <0.0001 (0.016)**	**F (5, 2121) = 2.65 0.0216 (0.006)**	183.98 [58.81, 309.14]	263.69 [248.51, 280.31]	0.41	16.24
SGE (normalized)	**F(1, 2119) = 45.3 <0.0001 (0.021)**	**F (5,2119) = 8.6 <0.0001 (0.02)**	F (5, 2119) = 1.15 0.3295 (0.0027)	0.00112 [0.0004, 0.0019]	0.0015 [0.0014, 0.0015]	0.44	0.04
GTE (normalized)	**F (1, 2119) = 4.82 0.0283 (0.002)**	F (5, 2119) = 0.39 0.8516 (0.0009)	F (5, 2119) = 0.34 0.8853 (0.0008)	0.0005 [0.0002, 0.0009]	0.0025 [0.0023, 0.0026]	0.17	0.05

2.3.4 Effect of Task Difficulty and Learning on AOI-measures

Considering AOI-based measures, plate-fixations significantly reduced, and robot-arm fixations significantly increased, in the HD condition compared to the LD condition (Figure 11a and Figure 11b). Target-fixations were not significantly different across conditions. The main effect of Trial on plate-fixations was significant, corresponding to an increase over time, whereas robot-arm fixations decreased

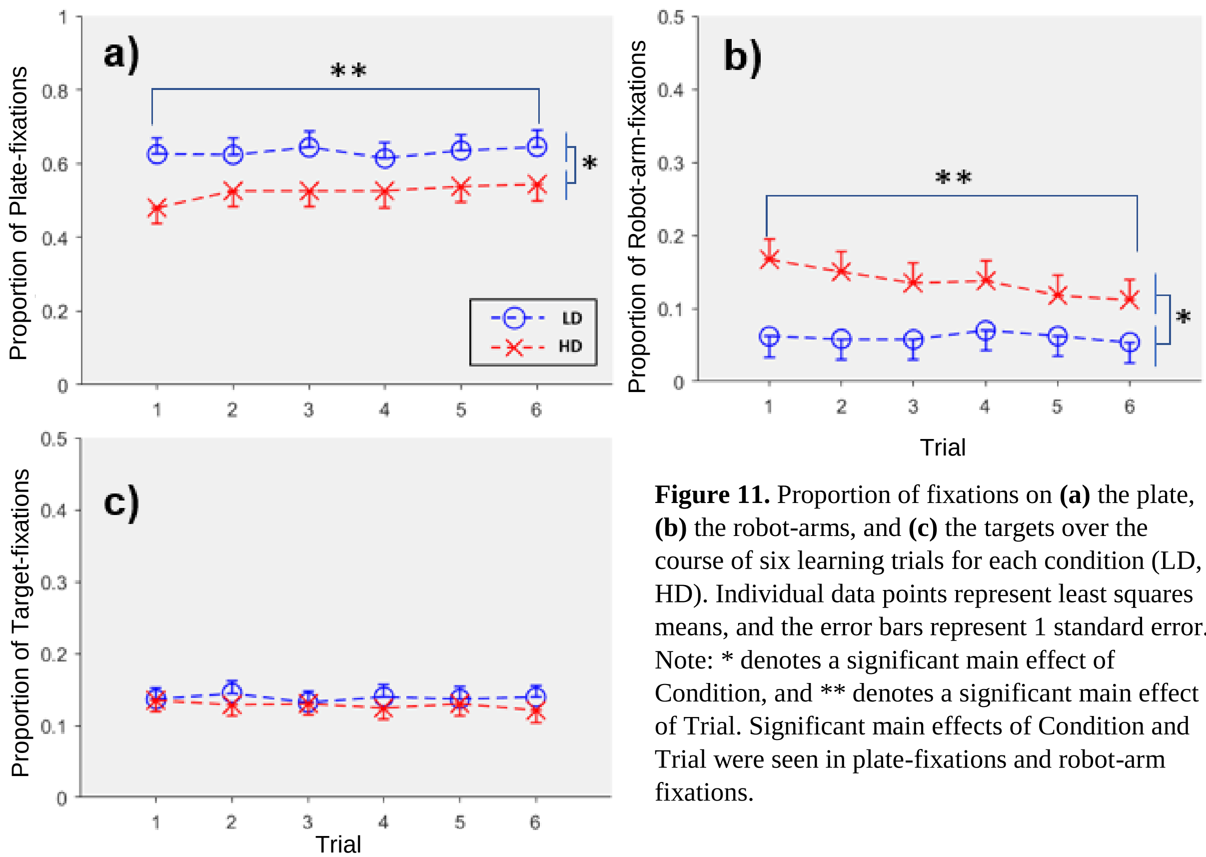


Figure 11. Proportion of fixations on **(a)** the plate, **(b)** the robot-arms, and **(c)** the targets over the course of six learning trials for each condition (LD, HD). Individual data points represent least squares means, and the error bars represent 1 standard error. Note: * denotes a significant main effect of Condition, and ** denotes a significant main effect of Trial. Significant main effects of Condition and Trial were seen in plate-fixations and robot-arm fixations.

over time. Significant Condition x Trial interaction effects (F-statistics, effect sizes and p-values in table 3) and post-hoc tests indicated that these changes over time occurred only in the HD condition. Target-fixations did not change significantly over trials. Males exhibited significantly more plate-fixations ($F(1, 16) = 7.8$; $p = 0.0129$), compared to females, who exhibited significantly more robot-fixations ($F(1, 16) = 10.84$; $p = 0.0046$).

Table 3. F-statistics, **e**ffect sizes and p-values for AOI-measures. Bold fonts indicate significant effects ($p < 0.05$)

AOI-Measures	F-statistic p-value (Effect Size - η_p^2)		
	Condition	Trial	Condition x Trial
Plate-Fixations	**F (1, 2119) = 183.4 <0.0001 (0.08)**	**F (5, 2119) = 7.87 <0.0001 (0.018)**	**F (5, 2119) = 3.51 0.0036 (0.008)**
Robot-Fixations	**F (1, 2119) = 182.5 <0.0001 (0.08)**	**F (5, 2119) = 8.43 <0.0001 (0.019)**	**F (5, 2119) = 6.26 <0.0001 (0.014)**
Target-Fixations	F (1, 2119) = 0.02 <0.0001 (0.00001)	F (5,2119) = 1.03 <0.0001 (0.002)	F (5, 2119) = 1.9 <0.0001 (0.004)

2.3.5 Performance prediction using eye-tracking metrics

The confusion matrices for the two logistic regression models are represented together in table 4, along with the model performance metrics. Overall, the model with eye-tracking measures as predictors (Model 2) was more sensitive (recall of 70%), specific (71%), and accurate (71%) than the model with performance measures only (Model 1 – recall 52%, specificity 67% and accuracy 60%).

Table 4: Confusion Matrices for logistic regression models comparing the predictive ability of manipulated task variables (a; Model 1) vs eye-tracking measures (b; Model 2).

	Model 1: Condition, Trial and Condition x Trial (R-squared = 0.062)				
		Predicted Success/Failure			
		1	**0**		
Actual Success/ Failure	**1**	0.516	0.484	Recall = 0.52	
	0	0.328	0.672	Specificity = 0.67	
		Precision = 0.52	NPV = 0.67	Accuracy = 0.60	F1 Score = 0.54

	Model 2: Eye-Tracking Measures **(R-squared = 0.304)**				
		Predicted Success/Failure			
		1	**0**		
Actual Success/ Failure	**1**	0.704	0.296	Recall = 0.70	
	0	0.294	0.706	Specificity = 0.71	
		Precision = 0.68	NPV = 0.73	Accuracy = 0.71	F1 Score = 0.69

2.4 Discussion

This study explored the potential of eye-tracking as a continuous and non-intrusive technique for measuring the mental workload associated with learning to use a bi-manual robot, as elaborated below. Our results indicated that performance and perceived mental workload were affected by task difficulty, but improved over the course of multiple trials (as intended). These changes were similar across males and females, however, females exhibited lower overall performance and reported higher workload compared to males. This performance advantage for males in manipulating the robot may be explained by their average physical strength and dexterity being higher than females (Thomas & French, 1985).

2.4.1 Sensitivity of eye-tracking metrics to changes in task difficulty and learning

Among the eye-tracking workload measures, stationary gaze entropy (SGE) and pupil diameter (PD) increased with task difficulty and gradually reduced over multiple trials, although PD did not reduce monotonically. A number of past studies have found SGE to increase while monitoring highly variable or complex visual environments (B. Shiferaw et al., 2019) or more difficult surgical procedures (Wu et al., 2019). In this study, higher SGE in the HD condition may have been due to the need to monitor and track the robot's (more unpredictable) behavior, or to monitor the environment for potential collisions. Higher fixation count (FC) in the HD condition may have been the result of more frequent collision-monitoring. FC did not change significantly over time in the HD condition, suggesting that the visual monitoring demands in the HD condition may have been persistently high throughout the experiment. Interestingly, FC also increased in the last three trials of the LD condition. While this may suggest an increased visual monitoring workload, collisions remained infrequent in these trials and the NASA-TLX ratings did not rise. A possible explanation is that after learning to minimize errors during the first half of the LD condition, participants may have shifted to more frequent visual sampling of the plate's orientation to ensure more precise placement, or tracking the ball, or improving other (secondary) aspects of performance. Fixations have been found to be not only sensitive to workload but also be task-specific and increase due to visual monitoring strategies in prior work (Rivecourt et al., 2008; Y. Zhang et al., 2022), which may be a reason for the dissociation of FC with collisions and NASA-TLX ratings in our study. It should be noted that although all our workload metrics showed significant differences across task conditions and most measures showed an effect of learning trial, all effect sizes were small, with GTE and FC effects likely being too small to be practically important. Pupil dilation had a medium-to-large effect of learning trial ($\eta_p^2 = 0.085$).

On the other hand, AOI-based measures of gaze behavior had medium-to-large effect sizes for task difficulty, specifically the plate-fixations ($\eta_p^2 = 0.0797$) and robot-fixations ($\eta_p^2 = 0.0793$). Further insight into the nature of task demands and the source of workload is provided by these AOI-based measures. It was found that the plate accrued the most fixations of all the task-relevant AOI. This was likely because participants relied on visual feedback of the plate in an effort to maintain their virtual 'grip', as well as to prevent the ball from rolling off. In the HD condition, participants seemed to significantly reduce plate-fixations and increase robot-fixations (Figure 11a and Figure 11c). Further analysis revealed a significantly higher likelihood of collisions in the HD condition, which might have led participants to look away from the plate to monitor the robot-arms and avoid collisions with targets. Additionally, participants may also have directed their visual attention towards the robot in order to better evaluate, and to develop an internal model of, the robot's control dynamics. This is supported by the observation that some participants tended to move the robot-arms in a random, exploratory manner in the HD condition for some time initially, presumably to better understand their dynamics. Robot-arm fixations reduced significantly over the course of the trials, suggesting that participants reduced their dependence on vision for monitoring the arms, likely as a result of forming an improved internal model. The lack of a significant difference in target-fixations between conditions may be due to a fixed target-monitoring strategy in both conditions, and not devoting additional visuomotor effort towards accurate placement in the HD condition. Regarding gender differences, females tended to fixate more on the robot compared to males in both the LD and HD conditions. Previous research has found females to be more risk-averse (Byrnes et al., 1999), which may have contributed to greater "checking" behavior towards the movement of the robot.

In summary, to detect differences due to task difficulty and learning, measures of gaze behaviors may need to be included in addition to traditional eye-tracking measures of mental workload in future pHRI studies.

2.4.2 Reliability of eye-tracking metrics

The relative reliability of PD (ICC = 0.68) was highest among all metrics, followed by SGE (0.44) and FC (0.41), and considered to be good based on criteria stated in previous research on eye-tracking measures (Zargari Marandi et al., 2018). However, it should be noted that PD is generally more reliable in studies involving cognitive (as opposed to motor-control) tasks (Krejtz et al., 2018; Matthews et al., 2015). The SEM value for PD was 0.16 mm, which is considered acceptable, since mental workload effects on PD typically range from 0.1 mm to 0.5 mm from baseline (Pfleging et al., 2016). However, in our study, the highest difference in the least squares means of pupil dilation between each trial in the LD and HD conditions was ~0.1 mm (Figure 10c). Thus, the SEM of PD may have been too large (indicating a large within-subject variance) for PD to be a reliable measure of workload. The small effect size of condition on PD further supports this conclusion. However, the effect of Trial on PD was significantly larger, indicating that in this specific study, PD may be a better indicator of changes in mental workload that occur over learning, as opposed to those due to task difficulty.

It is worth discussing the reliability and sensitivity of eye-tracking metrics in reference to standard measures of workload. A review of mental workload in the context of assistive wearable devices found that the NASA-TLX is still among the most widely used methods of measuring mental workload, owing to its convenience and ease of implementation (Marchand et al., 2021). In our study, NASA-TLX was found to exhibit higher sensitivity ($\eta_p^2 = 0.067$) and reliability (ICC = 0.78) than PD, fixation count and entropy measures. However, the AOI-measures in our study showed greater sensitivity to task difficulty than NASA-TLX, suggesting that task-relevant and context-dependent information on visual strategies may be more sensitive to task difficulty and learning manipulations than subjective measures such as the NASA-TLX. Previous work has also reported that the relative reliability of NASA-TLX was similar to that in our study, and higher than that of a pupil-based index of cognitive activity (Devos et al., 2020). However, sensitivity of NASA-TLX may be lower (Zargari et al., 2018) or higher (Matthews et al., 2015) than that of eye-tracking measures, depending on the specific metric considered, as well as the study context. For example, the sensitivity of pupil dilation has been found to be much higher ($\eta_p^2 = 0.195$) in a study that involved higher precision-demands (controlling a laparoscopic surgical tool) than those in our work (Jiang et al., 2015). Finally, although the NASA-TLX may match the sensitivity and reliability of objective eye-tracking metrics in some contexts, its major limitation will always be that it cannot be used to continuously measure instantaneous changes in workload. With these considerations, our results provide encouraging evidence regarding the potential use of eye-tracking measures as continuous indices of workload in pHRI tasks.

2.4.3 Performance prediction using eye-tracking metrics

The model with eye-tracking measures was more sensitive, specific, and accurate, clearly outperforming the model with task variables (Table 4), for classifying the outcome of each trial to be a success or failure. This is likely because eye-tracking measures continuously capture individual variation in workload and visual behavior, which our task-based independent variables cannot. Much existing work in adaptive robotics, largely in the rehabilitation domain (Brown et al., 2016), tends to adapt robot behavior based solely on task performance. However, in industrial use-cases, adapting robot behavior based on performance may prove to be too late due to the high safety costs of failures. Thus, using eye-tracking workload measures may provide proactive, anticipatory information about mental and visuomotor effort preceding a success or a failure, potentially leading to more timely and accurate adaptation of robot behavior.

2.5 Limitations and Future Work

A limitation of our study is the potential effect of physical workload on our results, specifically PD, which has been shown to increase with the perception of physical effort (Zénon et al., 2014). Although we attempted to control for physical workload by offering rest periods after each trial, multiple participants reported wrist strain during the study, which may have affected PD in addition to mental workload. A second limitation is that this study only recruited college students as participants, thus, limiting the generalizability of our results to other populations. With the rising age of industrial workers, it is expected that older adults will frequently use, and benefit from, physically-coupled robots in the future workplace. Thus, future research needs to examine mental workload and strategies associated with pHRI across diverse users. Finally, since our task environment was virtual, there may be some differences in terms of eliciting appropriate gaze behaviors and cognitive responses between the virtual and physical world. Specifically, the lack of haptic feedback usually utilized for physically manipulating an object may have increased reliance on visual feedback in the VR tasks studied here. Future studies could be conducted in real-world environments or in higher-fidelity VR environments to determine the extent to which these findings can be generalized to real-world robot manipulation scenarios.

2.6 Conclusion

Overall, this work found that SGE and PD may be usable indicators of mental workload in pHRI, specifically under precision- and monitoring demands. Pupil dilation was a more sensitive indicator of workload changes due to learning than those due to task difficulty in our task setup. Visual-attention metrics, specifically those that quantify the number of fixations in different AOIs, may be highly useful and informative in a pHRI context. Interestingly, the sensitivity of these metrics to both task difficulty and learning was higher than our workload metrics from both eye-tracking and NASA-TLX. Although AOI-based metrics are not direct measures of workload, they provide important information regarding visual strategies and performance, which are important mediators of workload (Loft et al., 2007; Tsang & Vidulich, 2006). Thus, their validity should be explored further, especially in the pHRI domain where motor-control and internal-model formation will likely be strong contributors to workload.

References

1. Aronson, R. M., & Admoni, H. (2018). *Gaze for Error Detection During Human-Robot Shared Manipulation*. Proceedings of RSS '18 Towards a Framework for Joint Action Workshop.
2. Aronson, R. M., Santini, T., Kübler, T. C., Kasneci, E., Srinivasa, S., & Admoni, H. (2018). *Eye-Hand Behavior in Human-Robot Shared Manipulation*. 10.
3. Becser, N., Sand, T., & Zwart, J.-A. (1998). Reliability of Cephalic Thermal Thresholds in Healthy Subjects. *Cephalalgia*, *18*(8), 574–582. https://doi.org/10.1046/j.1468-2982.1998.1808574.x
4. Bommer, S. C., & Fendley, M. (2018). A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations. *International Journal of Industrial Ergonomics*, *63*, 7–17. https://doi.org/10.1016/j.ergon.2016.10.007
5. Brown, D. A., Lee, T. D., Reinkensmeyer, D. J., & Duarte, J. E. (2016). Designing Robots That Challenge to Optimize Motor Learning. In D. J. Reinkensmeyer & V. Dietz (Eds.), *Neurorehabilitation Technology* (pp. 39–58). Springer International Publishing. https://doi.org/10.1007/978-3-319-28603-7_3
6. Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*, 367–383. https://doi.org/10.1037/0033-2909.125.3.367
7. Chadwell, A., Kenney, L., Thies, S., Galpin, A., & Head, J. (2016). The Reality of Myoelectric Prostheses: Understanding What Makes These Devices Difficult for Some Users to Control. *Frontiers in Neurorobotics*, *10*. https://doi.org/10.3389/fnbot.2016.00007
8. Cognolato, M., Atzori, M., & Müller, H. (2018). Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances. *Journal of Rehabilitation and Assistive Technologies Engineering*, *5*, 2055668318773991. https://doi.org/10.1177/2055668318773991
9. Cornwall, W. (2015). In pursuit of the perfect power suit. *Science*, *350*(6258), 270–273. https://doi.org/10.1126/science.350.6258.270
10. Ettinger, U., Kumari, V., Crawford, T. J., Davis, R. E., Sharma, T., & Corr, P. J. (2003). Reliability of smooth pursuit, fixation, and saccadic eye movements. *Psychophysiology*, *40*(4), 620–628. https://doi.org/10.1111/1469-8986.00063
11. Haddadin, S., & Croft, E. (2016). Physical Human–Robot Interaction. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1835–1874). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_69
12. Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. C. (2017). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, *49*(5), 1802–1823. https://doi.org/10.3758/s13428-016-0822-1
13. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press. http://lup.lub.lu.se/record/1852359
14. Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 56–88. https://doi.org/10.1080/14639220210159735
15. Kao, P.-C. (2009). *Principles of Motor Adaptation when Walking with a Powered Exoskeleton*. 120.

16. Komogortsev, O. V., & Karpov, A. (2013). Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior Research Methods*, *45*(1), 203–215. https://doi.org/10.3758/s13428-012-0234-9
17. Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE*, *13*(9), e0203629. https://doi.org/10.1371/journal.pone.0203629
18. Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, *26*(1), 51–62. https://doi.org/10.1017/S0952523808080899
19. Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004). Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. *Proceedings of the Eye Tracking Research & Applications Symposium on Eye Tracking Research & Applications - ETRA'2004*, 41–48. https://doi.org/10.1145/968363.968370
20. Llanes-Jurado, J., Marín-Morales, J., Guixeres, J., & Alcañiz, M. (2020). Development and Calibration of an Eye-Tracking Fixation Identification Algorithm for Immersive Virtual Reality. *Sensors*, *20*(17), Article 17. https://doi.org/10.3390/s20174956
21. Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and Predicting Mental Workload in En Route Air Traffic Control: Critical Review and Broader Implications. *Human Factors*, *49*(3), 376–399. https://doi.org/10.1518/001872007X197017
22. Marchand, C., De Graaf, J. B., & Jarrassé, N. (2021). Measuring mental workload in assistive wearable devices: A review. *Journal of NeuroEngineering and Rehabilitation*, *18*(1), 160. https://doi.org/10.1186/s12984-021-00953-w
23. Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The Psychometrics of Mental Workload: Multiple Measures Are Sensitive but Divergent. *Human Factors*, *57*(1), 125–143. https://doi.org/10.1177/0018720814539505
24. Novak, D., Beyeler, B., Omlin, X., & Riener, R. (2015). Workload Estimation in Physical Human–Robot Interaction Using Physiological Measurements. *Interacting with Computers*, *27*(6), 616–629. https://doi.org/10.1093/iwc/iwu021
25. Pfleging, B., Fekety, D. K., Schmidt, A., & Kun, A. L. (2016). A Model Relating Pupil Diameter to Mental Workload and Lighting Conditions. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5776–5788. https://doi.org/10.1145/2858036.2858117
26. Rivecourt, M. D., Kuperus, M. N., Post, W. J., & Mulder, L. J. M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, *51*(9), 1295–1319. https://doi.org/10.1080/00140130802120267
27. Sailer, U. (2005). Eye-Hand Coordination during Learning of a Novel Visuomotor Task. *Journal of Neuroscience*, *25*(39), 8833–8842. https://doi.org/10.1523/JNEUROSCI.2658-05.2005
28. Searle, S. R., Casella, G., & McCulloch, C. E. (2009). *Variance Components*. John Wiley & Sons.
29. Shiferaw, B. A. (2018). Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *SCIENTIfIC REPOrTS*, 10.
30. Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews*, *96*, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007
31. Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. https://doi.org/10.1037/0033-2909.86.2.420
32. Sobuh, M. M. D., Kenney, L. P. J., Galpin, A. J., Thies, S. B., McLaughlin, J., Kulkarni, J., & Kyberd, P. (2014). Visuomotor behaviours when using a myoelectric prosthesis. *Journal of NeuroEngineering and Rehabilitation*, *11*(1), 72. https://doi.org/10.1186/1743-0003-11-72
33. Srinivasan, D., & Martin, B. J. (2010). Eye–hand coordination of symmetric bimanual reaching tasks: Temporal aspects. *Experimental Brain Research*, *203*(2), 391–405. https://doi.org/10.1007/s00221-010-2241-3

34. Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*, 33. https://doi.org/10.1145/1121241.1121249
35. Stirling, L., Kelty-Stephen, D., Fineman, R., Jones, M. L. H., Daniel Park, B.-K., Reed, M. P., Parham, J., & Choi, H. J. (2020). Static, Dynamic, and Cognitive Fit of Exosystems for the Human Operator. *Human Factors*, *62*(3), 424–440. https://doi.org/10.1177/0018720819896898
36. Thomas, J. R., & French, K. E. (1985). Gender differences across age in motor performance: A meta-analysis. *Psychological Bulletin*, *98*, 260–282. https://doi.org/10.1037/0033-2909.98.2.260
37. Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In *Handbook of human factors and ergonomics, 3rd ed* (pp. 243–268). John Wiley & Sons, Inc. https://doi.org/10.1002/0470048204.ch9
38. Van Acker, B. B., Bombeke, K., Durnez, W., Parmentier, D. D., Mateus, J. C., Biondi, A., Saldien, J., & Vlerick, P. (2020). Mobile pupillometry in manual assembly: A pilot study exploring the wearability and external validity of a renowned mental workload lab measure. *International Journal of Industrial Ergonomics*, *75*, 102891. https://doi.org/10.1016/j.ergon.2019.102891
39. Webb, G. I. (2010). Model Evaluation. In *Encyclopedia of Machine Learning* (pp. 683–683). Springer.
40. White, M. M., Zhang, W., Winslow, A. T., Zahabi, M., Zhang, F., Huang, H., & Kaber, D. B. (2017). Usability Comparison of Conventional Direct Control Versus Pattern Recognition Control of Transradial Prostheses. *IEEE Transactions on Human-Machine Systems*, *47*(6), 1146–1157. https://doi.org/10.1109/THMS.2017.2759762
41. White, O., & French, R. M. (2017). Pupil Diameter May Reflect Motor Control and Learning. *Journal of Motor Behavior*, *49*(2), 141–149. https://doi.org/10.1080/00222895.2016.1161593
42. Wilson, M., McGrath, J., Vine, S., Brewer, J., Defriend, D., & Masters, R. (2010). Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts. *Surgical Endoscopy*, *24*(10), 2458–2464. https://doi.org/10.1007/s00464-010-0986-1
43. Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*(12), Article 12. https://doi.org/10.1038/nrn3112
44. Wu, C., Cha, J., Sulek, J., Sundaram, C. P., Wachs, J., Proctor, R. W., & Yu, D. (2021). Sensor-based indicators of performance changes between sessions during robotic surgery training. *Applied Ergonomics*, *90*, 103251. https://doi.org/10.1016/j.apergo.2020.103251
45. Wu, C., Cha, J., Sulek, J., Zhou, T., Sundaram, C. P., Wachs, J., & Yu, D. (2019). Eye-Tracking Metrics Predict Perceived Workload in Robotic Surgical Skills Training. *Human Factors*, 0018720819874544. https://doi.org/10.1177/0018720819874544
46. Zargari Marandi, R., Madeleine, P., Omland, Ø., Vuillerme, N., & Samani, A. (2018). Reliability of Oculometrics During a Mentally Demanding Task in Young and Old Adults. *IEEE Access*, *6*, 17500–17517. https://doi.org/10.1109/ACCESS.2018.2819211
47. Zénon, A., Sidibé, M., & Olivier, E. (2014). Pupil size variations correlate with physical effort perception. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00286
48. Zhang, Y., Hopko, S., Y, A., & Mehta, R. K. (2022). Capturing Dynamic Trust Metrics during Shared Space Human Robot Collaboration: An eye-tracking approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 536–536. https://doi.org/10.1177/1071181322661296
49. Zheng, B., Jiang, X., Bednarik, R., & Atkins, M. S. (2021). Action-related eye measures to assess surgical expertise. *BJS Open*, *5*(5), zrab068. https://doi.org/10.1093/bjsopen/zrab068
50. Zhou, T., Zhu, Q., & Du, J. (2020). Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction. *Advanced Engineering Informatics*, *46*, 101170. https://doi.org/10.1016/j.aei.2020.101170

Appendix A

Computation of Entropy

The computation of *stationary gaze entropy (SGE)* and *gaze transition entropy (GTE)* required knowledge of which VR objects (AOIs) were being successively fixated on. The raw data output from our study included the VR object on which the gaze point was located at every sample, although this sample could be part of a saccade, fixation, or smooth pursuit. Thus, for each fixation classified by the I-VDT algorithm, we computed the mode of the list of VR objects that were associated all the gaze samples within that fixation. The resulting mode was considered as the "fixated object" or "fixated AOI". Figure 12 illustrates our approach -

Sample ID	Eye Movement Type	VR Object	VR Object Fixated
1	Fixation	2	2
2	Fixation	2	2
3	Fixation	2	2
4	Fixation	2	2
5	Fixation	1	2
6	Fixation	3	2
7	Fixation	2	2
8	Fixation	2	2
9	Fixation	2	2
10	Fixation	3	2
11	Saccade	3	---
12	Saccade	4	---
13	Saccade	5	---
14	Saccade	6	---
15	Saccade	7	---
16	Fixation	5	6
17	Fixation	6	6
18	Fixation	6	6
19	Fixation	6	6
20	Fixation	6	6
21	Fixation	6	6
22	Fixation	7	6
23	Fixation	6	6

Figure 12. Samples 1->10 and 16->23 represent two different fixations. The "VR Object" column represents the object on which the gaze point falls at every sample. We computed the mode of the VR Object column for the duration of the fixation, i.e. from sample 1->10 and 16->23, resulting in objects 2 and 6 being determined as the fixated-AOI. Each unique value of "VR Object Fixated" represented a single state space, and a set of consecutive state spaces are used in the entropy calculation.

Two types of entropy were computed – Stationary Gaze Entropy (SGE) and Gaze Transition Entropy (GTE). SGE was computed using the following equation adapted from (B. A. Shiferaw, 2018):

$$H_s(x) = -\sum_{i=1}^{N} p(i) \times \log_2 p(i) \qquad (1)$$

where $H_s(x)$ is the SGE value for a particular time-bin 'x' (equal to 30s), 'i' represents the successive AOI-fixations in the interval 'x' and 'p' is the proportion of fixations on the i^{th} AOI in the interval 'x'. This calculation is illustrated in Figure 13. The figure shows N = 7 objects fixated on during a specified time interval. First, the probabilities of each object "i" are computed, so p(object 2) = 3/7, p(object 6) = 1/7 and p(object 1) = 3/7. These probability values are then entered into equation 1 to compute SGE.

Fixation Number	VR Object Fixated
1	2
2	6
3	2
4	1
5	1
6	1
7	2

Figure 13. Array of successive state spaces (VR objects fixated) where N = 7.

For GTE, we first generated a transition matrix for each 30-second interval, in which the element at row 'i' and column 'j' represented the probability of transitions from object 'i' to object 'j' within the 30-second interval. An example of such a transition matrix from our data is given in figure 14:

	1	2	3	4	5	6	7	8	9	10	11	12
1	8	0	0	1	1	1	0	2	1	0	1	0
2	1	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	1
4	1	0	0	0	0	0	0	0	0	0	0	0
5	1	0	0	0	0	0	1	1	0	0	0	0
6	0	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0	0	1
8	3	0	0	0	0	0	0	2	0	0	0	0
9	0	0	0	0	2	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	1
11	0	0	0	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	1	1	0	1

Figure 14. Transition matrix with frequencies for a 30-second interval. In this particular case, 12 unique VR objects were fixated on (out of a total of 39 possible objects). Each element (i,j) represents the number of transitions from object "i" to object "j" in the 30-second interval. For example, there were two transitions from object 1 to object 8 (cell with red border). Note that there were no transitions from object 11 to any other object, resulting in all zero values for row 11.

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.5333	0	0	0.0667	0.0667	0.0667	0	0.1333	0.0667	0	0.0667	0
2	1	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	1
4	1	0	0	0	0	0	0	0	0	0	0	0
5	0.3333	0	0	0	0	0	0.3333	0.3333	0	0	0	0
6	0	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0.5000	0	0	0	0	0.5000
8	0.6000	0	0	0	0	0	0	0.4000	0	0	0	0
9	0	0	0	0	1	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	1
11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12	0.2500	0	0	0	0	0	0	0	0.2500	0.2500	0	0.2500

Figure 15. Transition matrix with probabilities for a single 30-seond interval. Each element (i,j) in the matrix shown in figure 7 was divided by the sum of the elements in the corresponding i^{th} row. As a result, each element (i,j) in this matrix represents the probability of transitioning to object "j", given that the previous object was "i". Row 11 contains NaN values since there were no transitions from object 11 to any other object. This row is ignored in the GTE computation.

Based on the transition matrix, GTE was computed using the following equation adapted from (B. A. Shiferaw, 2018):

$$H_g(x) = -\sum_{i=1}^{N} p(i) \left[\sum_{j=1}^{N} p(i|j) \times \log_2 p(i|j)\right] \qquad (2)$$

where $H_g(x)$ is the value of GTE, 'p(i)' is the stationary distribution of fixation locations, and 'p(i | j)' is the probability of transitioning to AOI 'j' given current position of 'i'. SGE and GTE were both normalized to the maximum theoretical entropy for each 30-s interval, equal to $\log_2 N$ (B. Shiferaw et al., 2019).

Lastly, we also computed the ratio of fixations on the following AOI - the plate (which included the ball), the robot-arms, and the top surface of the target pillars, per 30-second interval, in order to understand the relative attentional focus directed towards these AOI. These AOI were selected because they were directly related to participants' goals and actions, and hence were critical components of the task. Fixations on the table and the floor were coded as "miscellaneous" fixations that were not task-relevant.

Chapter 3. Gaze Behavior and Mental Workload while using a Whole-body Powered Exoskeleton

Abstract

Wearable assistive devices such as exoskeletons and prosthetic devices aim to improve the physical function of their users, by either enhancing motor performance (or restoring motor abilities). However, as the capabilities of these devices continue to improve, the mental workload associated with controlling them has been identified as a significant concern affecting device acceptance, usability, and safe operation. The need for better methods and metrics to understand the mental workload associated with operating assistive devices has been highlighted recently by a systematic review of relevant literature. In this context, the goal of this study is to characterize the mental demands associated with operating complex whole-body powered exoskeletons. This study aimed to quantify the overall mental workload associated with using a powered whole-body exoskeleton among expert and novice users, as well as the changes in workload resulting from novices adapting to exoskeleton-use over time. We used eye-tracking measures to quantify the differences in workload of six novices and five experts while they performed a level-walking task, with and without wearing a whole-body powered exoskeleton, over three repeated sessions. Pupil dilation (PD), gaze fixation rate (FR), stationary gaze entropy (SGE), and the ratio of gaze fixations directed downward on to the walking path (vs. looking elsewhere; PF) were quantified. We found that PD of only the novices increased with exoskeleton use, and continued to be high even after three exoskeleton sessions, as compared to the control (no-exoskeleton) condition. SGE was significantly reduced among both experts and novices in the exoskeleton condition, compared to the no-exoskeleton condition. We also found that while both experts and novices exhibited higher PF while using the exoskeleton, compared to the no-exoskeleton condition, there was a significant interaction effect, in that experts showed higher PF than novices. These results indicate that novices' mental demands were higher, and that both experts and novices exhibited more focused gaze patterns (less visual exploration as indicated by lower SGE) while using the exoskeleton. In addition, downward direction of gaze, a visuomotor strategy to potentially increase walking stability and help with anticipatory movement planning, was found to be more developed among expert users than novice

users of the exoskeleton. Furthermore, eye-tracking measures may potentially be useful for detecting differences in workload and skill-level associated with operating complex assistive devices such as exoskeletons, and be considered as inputs for estimating a user's cognitive load in developing future adaptive exoskeleton control algorithms.

Keywords: Mental Workload, Coordination, Balance, Eye Tracking, Downward Gaze, Expertise

3.1 Introduction

Advancements in exoskeleton technologies have allowed whole-body powered exoskeletons to gain increasing attention as potential interventions to reduce physical demands in occupational work tasks (Kim et al., 2021; Park et al., 2022) However, while more studies on physical performance assessments of exoskeleton use are appearing in the literature, the cognitive and attentional demands associated with operating complex exoskeletons, and how these change with adaptation to continued exoskeleton use, are poorly understood.

While walking is normally considered an activity requiring little or no cognitive effort, the use of an assistive device such as a lower limb prosthesis has been shown to require a high level of concentration (Morgan et al., 2016), preventing the user from walking on uneven ground at the risk of falling (Miller et al., 2001), or performing a second parallel task such as holding a conversation (Heller et al., 2000). The use of an exoskeleton by an able-bodied user can also potentially produce similar effects. For example, early lower-limb active exoskeletons such as the Human Universal Load Carrier (HULC) were notoriously difficult to use and caused unnatural gait kinematics due to a mismatch between operator expectations and exoskeleton behavior (Cornwall, 2015). A more recent study found that the motor-coordination demands of walking with an active lower-limb exoskeleton increased subjective mental workload and reduced performance on secondary cognitive tasks, although these effects varied across individuals (Bequette et al., 2020). Another recent study found that compared to walking without an exoskeleton, users who walked with a whole-body powered exoskeleton activated their lower-limb muscles more frequently and at a higher level, and took longer to initiate steps during each gait-cycle (i.e. they increased the time spent in dual stance) (Kim et al., 2021). The authors stated that these behaviors may reflect increased mental and physical effort for maintaining walking stability, possibly due to an imperfect mental model of the exoskeleton's dynamics. Considering that powered exoskeletons may be intended for use in hazardous and complex industrial environments, where elevated mental workload may lead to distraction and reduced safety, it is important to better understand and quantify the potential cognitive demands required for control of these devices. Also, given that individuals may differ in their learning rate and subjective experience of workload, it is also important to quantify how mental workload may change over adaptation and learning to use powered exoskeletons.

While the measurement of mental workload has been extensively studied for over 40 years (Young et al., 2015), a recent systematic review of over 60 articles in the field of wearable assistive technologies (Marchand et al., 2021) highlights that the measurement of mental workload for wearable assistive devices remains an important and unresolved challenge, and that there is not yet an effective, universally agreed-upon method. Characterizing mental workload in the domain of assistive technologies has been suggested to be complex (and different from other fields) due to a variety of factors such as inter-dependencies between physical and mental workload experienced by users, psychological factors related to perception/expectations of the assistive technology, adaptations over time, and the variable physical interface between the user and device that causes high inter-individual variation in device controllability. Of the various methods in the literature to measure mental workload, subjective assessments and dual-task paradigms were reported as the most commonly used methods, however, they were shown to suffer from imprecision, and a need for careful consideration of appropriate and robust physiological measures was recommended (Marchand et al., 2021).

In this context, eye-tracking is a promising technique for measuring workload, since it can provide pupillary measures (e.g. pupil dilation) that correlate with the involuntary neural response to workload (Just et al., 2003), as well as eye-movement measures (e.g. fixation rate and gaze entropy), which relate to the active modulation of visual attention to meet task demands (Sarter et al., 2007; B. Shiferaw et al., 2019). Importantly, both pupillary and gaze-behavior measures have been shown to track changes over the course of cognitive and motor-skill learning (Foroughi et al., 2017; Sailer, 2005). The versatility of eye-tracking, coupled with its increasing wearability and ubiquity (Cognolato et al., 2018) make eye-tracking a viable technology to implement in dynamic, real-world environments.

Eye-tracking measures have previously been used to quantify the demands associated with difficult balance- and coordination tasks. Pupil dilation has been shown to be significantly higher when users have had difficulty controlling their wearable assistive devices (Lindner et al., 2020; M. M. White et al., 2017; Zahabi et al., 2019; W. Zhang et al., 2016). A recent study found that a difficult standing-balance task led to increases in a pupillary index of cognitive activity, indicating higher cognitive effort (Kahya et al., 2018). Another study found that older adults, who tend to invest greater cognitive effort for postural control, had longer fixations and more constrained (less exploratory) gaze compared to younger adults (Walsh, 2021). In unsteady walking conditions, walkers have been found to lower their gaze and direct it towards the walking path (as opposed to gazing straight ahead) (Koren et al., 2022). The authors of that work stated that downward-focused gaze may reflect an attempt to anticipate the future location of one's steps, or an attempt to make better use of motion parallax and visual expansion in order to improve stability while walking. Finally, when transitioning from level ground to stairs and vice-versa, lower-limb prosthesis-users directed their gaze towards the transitional region, whereas able-bodied users did not, suggesting that the use of a lower-limb assistive device can increase the challenge of gait and lead to greater visual monitoring of key environmental features (M. Li et al., 2019).

The above findings suggest that challenging coordination, movement-planning, and postural-control tasks can influence cognitive demands and visuomotor strategies; however, no research has quantified these effects in the context of using a whole-body powered exoskeleton (EXO). Our study compared pupil dilation (PD) and gaze behavior while walking with an EXO on level ground (compared to without the EXO), changes in these measures over multiple practice sessions with the EXO, as well as potential differences between first-time and skilled EXO operators. We hypothesized that pupil dilation (PD) and fixations on the path (PF) would be higher, and stationary gaze entropy (SGE) and Fixation Rate

(FR) would be lower in the EXO condition compared to the no-EXO condition for first-time operators, and that the size of these effects would be attenuated for skilled operators.

3.2 Methods

3.2.1 Participants

A sample of 11 healthy male participants completed the study. Prior to any data collection, informed consent was obtained from each participant following procedures approved by the University Institutional Review Board. Since adequate fit in the whole-body powered exoskeleton (i.e., sufficient contact to torso and pelvis load cells) was critical for the safe operation of the exoskeleton prototype considered in this study, all the participants in the study had to be at least 1.78 m tall. Six of eleven participants were novice operators, with respective mean (SD) stature, body mass, and age of 1.8 (0.04) m, 84.4 (6.8) kg, and 36.8 (15.4). These novice participants had no experience operating the whole-body powered exoskeleton prior to data collection. Five of the eleven participants were considered to be experts, with mean (SD) stature, body mass, and age of 1.8 (0.03) m, 83.9 (8.2) kg, and 31.2 (7.8). Each of these experts had extensive experience in testing and operating the whole-body powered exoskeleton throughout its developmental phases (>3 months). No participants had any self-reported musculoskeletal injuries or disorders in the previous 12 months before the study, and we recruited only those participants who could perform the walking task without wearing corrective lenses, or agreed to wear contact lenses only.

3.2.3 Materials and Apparatus

Powered Exoskeleton and Eye Tracker

We used an early prototype version of the Guardian® XO® (EXO for brevity) developed by Sarcos Robotics, which was specifically designed for heavy industrial applications. The EXO had a mass of 158 kg and included 18 active degrees of freedom (DOFs; Figure 16). The EXO used a patented "Get-Out-Of-The-Way" control scheme with torque sensors at major body joints that allow the EXO to follow human movements and amplify human joint torques, according to pre-set gains (Jacobs et al., 2018). To understand user movement intent, user input was obtained from embedded 6-DOF force-moment load cells located at the hands, feet, torso, and pelvis locations of the EXO. Participants wore a safety helmet during each study session.

For measuring eye-activity, we used a head-worn Tobii eye-tracker (Tobii Pro glasses 2, Tobii Technology AB, Danderyd, Sweden) that sampled gaze data at 100 Hz. The eye-tracker records a video

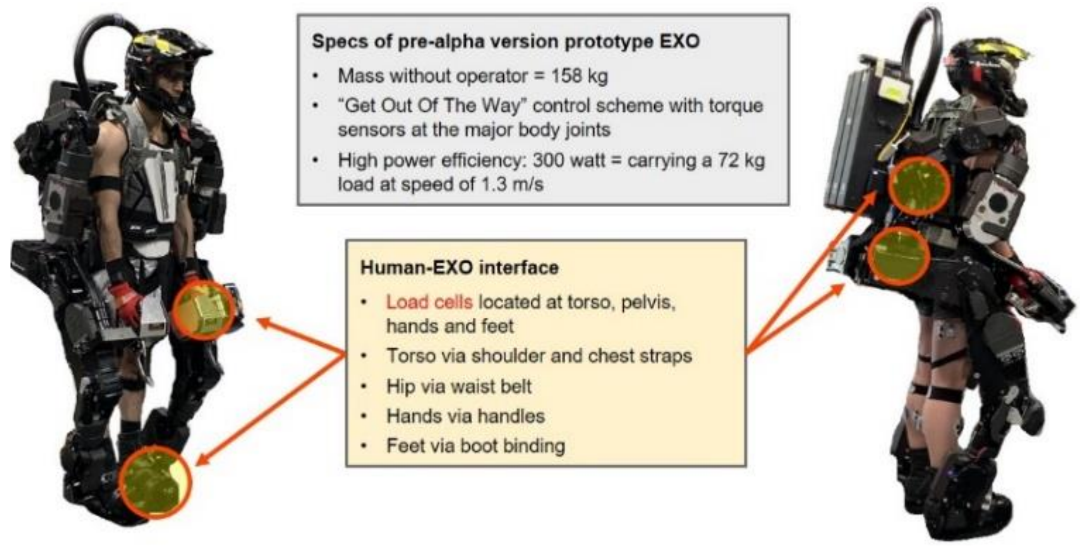


Figure 16. Pre-alpha prototype of the occupational whole-body powered exoskeleton (Guardian® XO®, Sarcos Robotics, www.sarcos.com). The red circled areas denote the human-EXO load cell (6-DOF force-moment sensor) interfaces where the EXO measures human-EXO interaction forces.

from the wearer's point of view using a forward-facing camera, at a resolution of 1920 x 1080 and a framerate of 25 Hz. The point of gaze was overlaid onto the video to identify where the participant was looking at a given point in time. To reduce loss in accuracy due to eye-tracker slippage and external vibrations, we secured the eye-tracker to the helmet by using tape (Niehorster et al., 2020).

3.2.4 Procedure

Novices performed four level-walking sessions with the EXO and one without it (No-EXO session), with the first session involving familiarization and parameter tuning (Figure 17). In this first familiarization session, participants received basic training including EXO operational and safety training, donning/doffing training, how to move around and manipulate the EXO, strategies to retain balance, how to react if the EXO accidentally fails/faults during an operation, and were also provided with training and feedback on what may be going wrong if they were unable to perform a certain task. Then, participants completed several walking trials along a level, linear, 10m gait track, during which the various tunable parameters (such as the virtual center of mass and harness adjustments) were modified until participants felt securely fitted and comfortable using the EXO. Verbal feedback from participants was used to iteratively check and further optimize the fit and tunable EXO parameters.

Following this familiarization session, three subsequent walking sessions (S1-S3) were completed. In these three walking sessions, no further specific training was provided to participants. In each walking session, data were collected when participants walked along the linear gait track at their preferred walking speed six times (i.e., six trials). Each novice's experimental session was separated from another by at least two hours, but such that all three sessions were collected within a four-day experimental period, to minimize potential fatigue development or any long-term loss in adaptation. Novices were randomly assigned to perform one no-EXO (control) condition either in sessions 1 or 3. Experts completed the same walking tasks in two sessions (EXO and no-EXO) on the same day.

Figure 17. Experiment Design. Novices performed a familiarization/training session followed by three EXO sessions with minimum separation of 2 hours, and one No-EXO session which took place on the same day as either EXO-S1 or EXO-S3. Experts performed the No-EXO and EXO sessions on the same day

3.4.5 Outcome Measures and Data Processing

Outcome metrics were computed per walking direction (~6 times per session). The analysis was also limited only to the straight-line walking phases of the task (and did not include turns at the end of the walkway to repeat a walking trial). Walking phases were marked manually using Tobii Pro Lab software

(Figure 18a) and raw data for those walking intervals were exported and analyzed. Trials with > 58% of valid eye-tracking samples were included in the analysis. The outcome metrics are described below.

a) *Pupil Dilation (PD)*: We computed baseline pupil diameter as the average pupil diameter of both eyes during two seconds of physical inactivity prior to beginning the walking trials. Pupil dilation was then computed by subtracting this baseline value from the average pupil diameter recorded during each straight-line walking trial. PD, reflecting mental effort and attention (Eckstein et al., 2017), was expected to reflect the overall mental demands associated with operating the exoskeleton.

b) *Fixation Rate (FR)*: The number of fixations occurring per second. Tobii Pro Lab software was used to classify gaze event data into fixations using an "I-VT" velocity-threshold fixation algorithm (Olsen, 2012). The default algorithm settings were used, which included linear interpolation of gaps in the data shorter than 75 ms and noise reduction using a moving median across a window of 3 samples. Gaze data were classified as fixations if the angular velocity of the eyes was below 30°/s for more than 60 ms. FR was computed by dividing the total number of fixations per single-line walk by the walking duration (in seconds). FR was expected to quantify the degree of overall visual monitoring, with a higher FR indicating greater visual monitoring.

c) *Stationary Gaze Entropy (SGE)*: SGE reflects the randomness in the overall spatial distribution of gaze fixations in the environment. To compute entropy, we first mapped fixations to a snapshot of the participant's first-person view of the walking path (Figure 18b) using the Tobii "Assisted Gaze Mapping" feature (connect.tobii.com/s/article/how-to-perform-manual-and-assisted-mapping). Any incorrectly mapped fixation locations were manually corrected. Each 1920 x 1080 pixel snapshot was then discretized into spatial bins, or areas of interest (AOIs) of 100 x 100 pixels (Figure 18d), and the proportion of fixations falling into each of these spatial bins was used to compute entropy, using the following equation adapted from (B. Shiferaw et al., 2019):

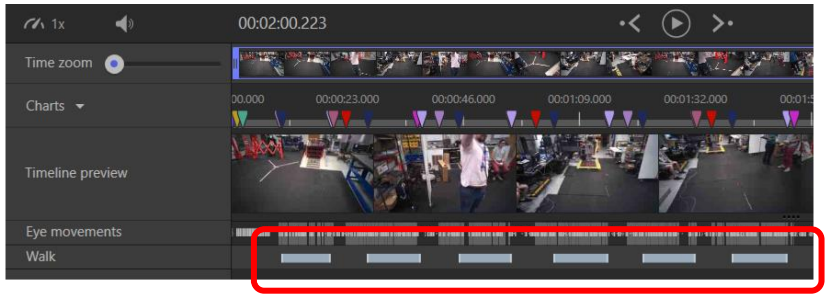
$$H_s(x) = -\sum_{i=1}^{N} p(i) \times \log_2 p(i) \qquad (1)$$

where $H_s(x)$ is the SGE value for a particular time-interval 'x' (equal to the duration of each single-direction walk), 'i' represents the successive AOI-fixations in the interval 'x' and 'p' is the proportion of fixations on the i^{th} AOI in the interval 'x'. N equals the number of unique AOIs that were fixated during the interval 'x'. SGE was normalized to the maximum theoretical entropy for each walking interval 'x', equal to $\log_2 N$ (B. Shiferaw et al., 2019).

d) *Ratio of path-focused fixations (PF)*: PF represents the ratio of fixations that were directed towards the walking path, versus straight ahead in the line of sight. To compute PF, we first generated a set of rectangular 1800 x 60 px AOI that spanned across each snapshot from top to bottom (Figure 18c). The AOIs that covered the portion of the walkway leading up to the turning point were categorized as "path-AOIs" or "D-AOIs" (these covered approximately the lower half of the snapshot) and those above the turning point were classified as "upward-AOIs" or "U-AOIs". Both types of AOIs were numbered from one to a maximum of 12 for the U-AOIs and 15 for the D-AOIs (Figure 18c). PF was computed as the ratio of all fixations located in the path-AOIs to the total number of fixations during each single-direction walk.

a) Mark walking intervals

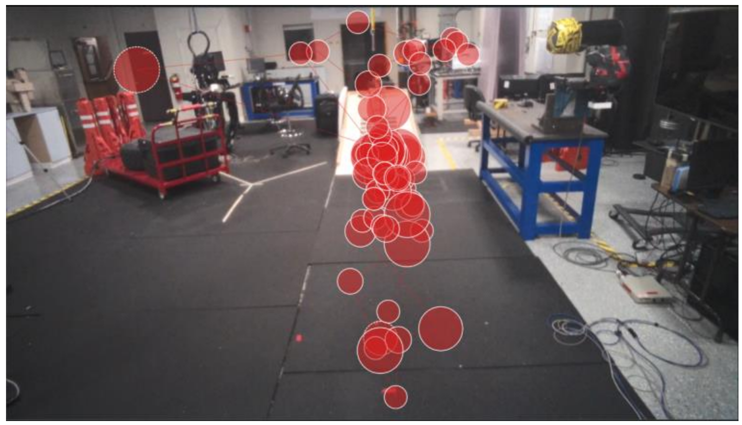
a)



b) Map gaze data

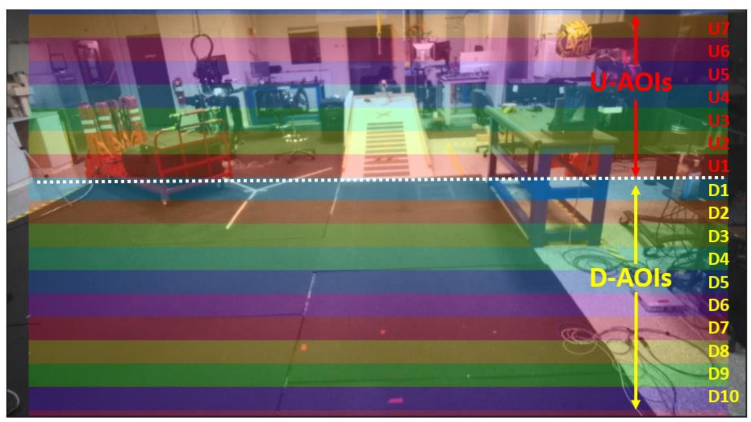
1. Generate snapshot for recording
2. Use mapping algorithm
3. Manually correct mapping errors

b)



c) AOIs for computing path-focused fixations. Fixations in “D” AOIs correspond to path-focused gaze and those in “U” AOIs indicate upward-focused gaze. The white dotted line represents the path endpoint.

c)



d) AOIs for computing entropy (illustration only). These AOIs were defined programmatically using a MATLAB script, since they were uniformly distributed across each snapshot and not tied to any environmental features.

d)

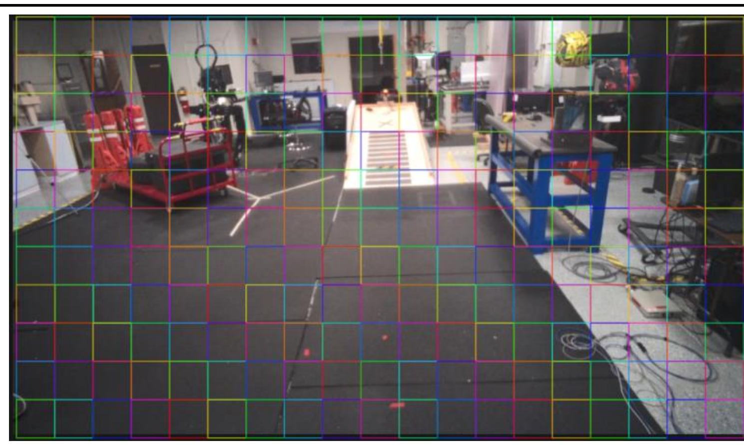


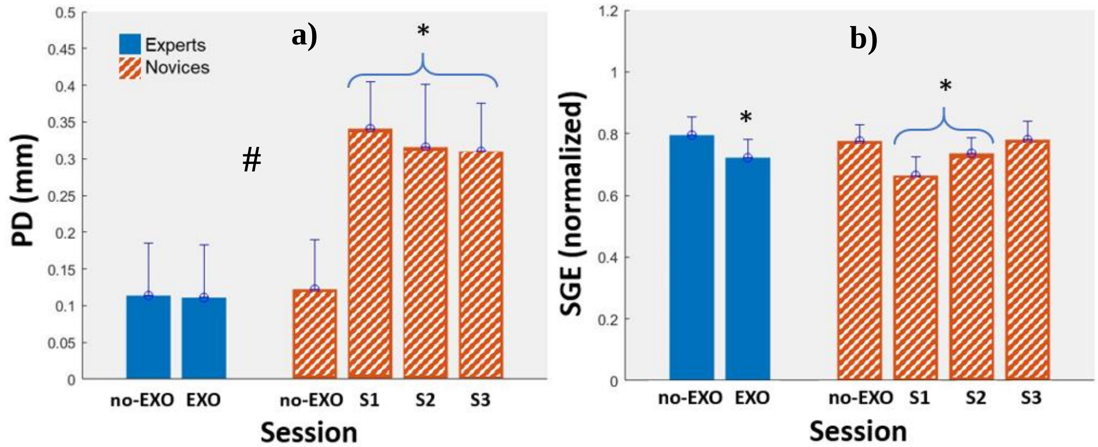
Figure 18. Data processing for a single participant, for a single walking direction

3.2.6 Statistical Analysis

A 2-way ANOVA with Session (no-EXO, EXO), Group (Novice, Expert) and their interaction was performed for each session for analyzing whether use of the exoskeleton caused significant changes in visuomotor behaviors and mental workload, and whether these changes were variable across exoskeleton-use skill level. For those outcome variables that violated the normality assumption, appropriate transformations were conducted before running the analyses. We set the significance level at $\alpha = 0.1$ to determine statistical significance in this exploratory work, and all statistical analyses were performed in JMP (SAS Institute Inc., USA).

3.3 Results

The two groups were comparable at baseline, i.e., in the no-exoskeleton condition, there were no group differences between experts and novices in any of the dependent measures. Verbal reports and observed walking performance indicated that walking in the exoskeleton was more mentally effortful than walking without. This was also supported by the finding that multiple novice participants lost balance during initial familiarization trials in the exoskeleton and appeared to take longer to walk with the exoskeleton compared to without. Gait kinematics data were from these tasks were analyzed separately and published recently (Park et al., 2023). Results for all gaze dependent measures are illustrated in figure 19, and p values and effect sizes are shown in Table 5. There was a significant *EXO x Group* interaction effect for PD: in all three EXO sessions, novices showed significantly higher PD (~200% increase compared to No-Exo), but experts did not show such an increase in PD (Figure 19a). There were small (~10%), but significant, reductions in SGE among experts as well as during the first two EXO sessions for novices, compared to the respective no-EXO conditions (Figure 19b). Both experts and novices showed significantly higher PF in the EXO condition compared to the no-EXO condition, and the *EXO x group* interaction was also significant for PF: experts showed higher PF in the EXO condition than the No-Exo condition compared to all three novice-EXO sessions (~300% increase among experts compared to ~80% increase among novices in the EXO condition compared to No-Exo; Figure 19c).



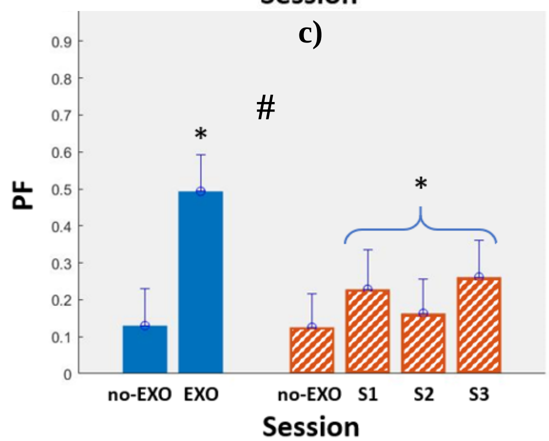


Figure 19. a) Pupil dilation (PD), **b)** Stationary Gaze Entropy (SGE), and **c)** Ratio of path-focused fixations (PF) in the EXO and no-EXO sessions for experts and novices. The bars represent least squares means, and the error bars represent 1 standard error. Note: * denotes a significant main effect of Session (i.e., exoskeleton session compared to no-exoskeleton session); # denotes significant interaction effect of *EXO x group.*

Table 5

a. Effect sizes quantified using partial eta-squared (η_p^2) and p-values of pupil dilation and gaze behavior. SGE was logit-transformed to fit the normality assumption. "G" indicates Group, and S indicates Session

Workload Measures	**p-value (effect size - η_p^2)**								
	S1			S2			S3		
	G	S	G x S	G	S	G x S	G	S	G x S
PD (mm)	.1515 (.077)	**0.0029 (0.056)**	**0.0023 (0.06)**	.5446 (.037)	**0.0638 (0.021)**	**0.0509 (0.024)**	.2019 (.095)	**0.0013 (0.08)**	**0.001 (0.084)**
SGE*	.6054 (.0765)	**0.0025 (0.32)**	.6115 (.02)	.8844 (.002)	**0.0301 (0.15)**	.6596 (.007)	.8148 (.0035)	.1347 (.1077)	.1884 (.042)
PF	.3535 (.2004)	**0.0002 (0.398)**	**0.0213 (0.19)**	.1988 (.27)	**0.0005 (0.34)**	**0.0034 (0.26)**	.361 (.166)	**<0.0001 (0.43)**	**0.0428 (0.156)**

b. F-statistics for pupil dilation and gaze behavior metrics. Bold fonts indicate significant effects ($p < 0.1$). Degrees of freedom are shown in brackets.

Workload Measures	S1			S2			S3		
	G	S	G x S	G	S	G x S	G	S	G x S
PD (mm)	F (1, 9.44) = 2.43	**F (1, 149.7) = 9.2**	**F (1, 149.7) = 9.6**	F (1, 8.9) = 0.4	**F (1, 149.1) = 3.5**	**F (1, 149.1) = 3.9**	F (1, 9.16) = 1.89	**F (1, 140.1) = 10.7**	**F (1, 140.1) = 11.3**
SGE*	F (1, 8.48) = 0.24	**F (1, 27.65) = 8.7**	F (1, 27.65) = 0.38	F (1, 9.2) = 0.002	**F (1, 30.06) = 3.16**	F (1, 30.06) = 0.27	F (1, 8.5) = 0.065	F (1, 27.5) = 1.44	F (1, 27.5) = 1.84
PF	F (1, 9.27) = 0.95	**F (1, 28.34) = 19.04**	**F (1, 28.34) = 5.94**	F (1, 9.15) = 1.9	**F (1, 29.83) = 15.35**	**F (1, 29.83) = 10.12**	F (1, 9.26) = 0.92	**F (1, 28.58) = 21.69**	**F (1, 28.58) = 4.5**

3.4 Discussion

This study explored the effects of using a whole-body powered exoskeleton on measures of mental workload and gaze behavior. Overall, our results indicated that walking in the exoskeleton was mentally demanding, particularly in terms of maintaining whole-body coordination and balance. Our hypothesis that novices would experience higher mental demand compared to experts, as reflected in PD and SGE, was partially supported. Novices' PD was significantly higher in the EXO session, whereas no differences in PD were observed among experts, between the EXO and no-EXO sessions. This suggests that walking in the exoskeleton may have imposed a greater motor-cognitive demand, but only on novices. Our task, i.e., using the EXO, was significantly physically demanding (Park et al., 2023), compared to level walking in the control condition. Prior work has shown that in tasks combining high physical and mental workload, there is a risk of observing false positives of increases in mental workload (e.g., in (Bridger et al., 2018), the increase in heart rate was due to participants maintaining a squatting position and not their mental workload) or false negatives (e.g., in (Knaepen et al., 2015), mental workload as measured by HRV was masked by the physical demands of the task). However, the selective increase of PD only among novices and not experts (despite both groups performing the same physically demanding tasks) shows that PD can be a measure of mental workload that is relatively robust to variations in concomitant physical workload. Moreover, it is possible that we may have in fact under-estimated the effects on PD, considering that novices were more susceptible to pupil-constriction driven by greater environmental light because of their greater upward-focused gaze in the EXO, compared to experts.

Next, the EXO condition led to a significant reduction in SGE from the no-EXO condition, indicating less exploration of the visual environment, potentially due to greater cognitive demands (Walsh, 2021) or an effort to improve balance (Jahn et al., 2002), and this effect was similar among experts and novices. SGE has been shown to be associated with situation awareness in emergency situations involving nuclear power plant control (Lee et al., 2022), and useful for predicting performance such as lane departures and safety of intersection crossings (Martin et al., 2021; B. A. Shiferaw, 2018), in prior work. The magnitude of change in SGE observed in our study (~10%) was also similar to those reported in earlier studies of SGE among normal vs. sleep-deprived adults driving over long durations. Thus, the changes in SGE associated with exoskeleton use observed in our study may represent meaningful differences that may influence the situational awareness and safety performance of exoskeleton operators, especially in dynamic industrial settings.

We hypothesized that PF would increase while using the exoskeleton, indicating a greater visual focus on the walking path, possibly as a strategy to improve walking stability (Koren et al., 2022). As expected, PF increased in the EXO condition compared to the no-EXO condition for both novices and experts. However, interestingly, experts showed a significantly higher proportion of downward path-fixations compared to novices (Figure 19). It is possible that the use of downward-focused gaze as a strategy to improve walking stability may be learnt over time, and novices did not develop this strategy over the course of our study. Further analysis of gaze locations on specific AOIs indicated that there was a significant ($p<0.05$) increase in the fixations on the “D1” and “U1” AOIs (for an illustration, see Figure 20) from the No-EXO to the EXO condition for both groups. Since these two AOIs were adjoining to the endpoint of the path, a higher proportion of fixations in these AOIs may indicate that in addition to gazing downward for steadiness, participants may also have been used this as a strategy to anticipatorily plan and control their movements relative to the destination endpoint while walking in the EXO.

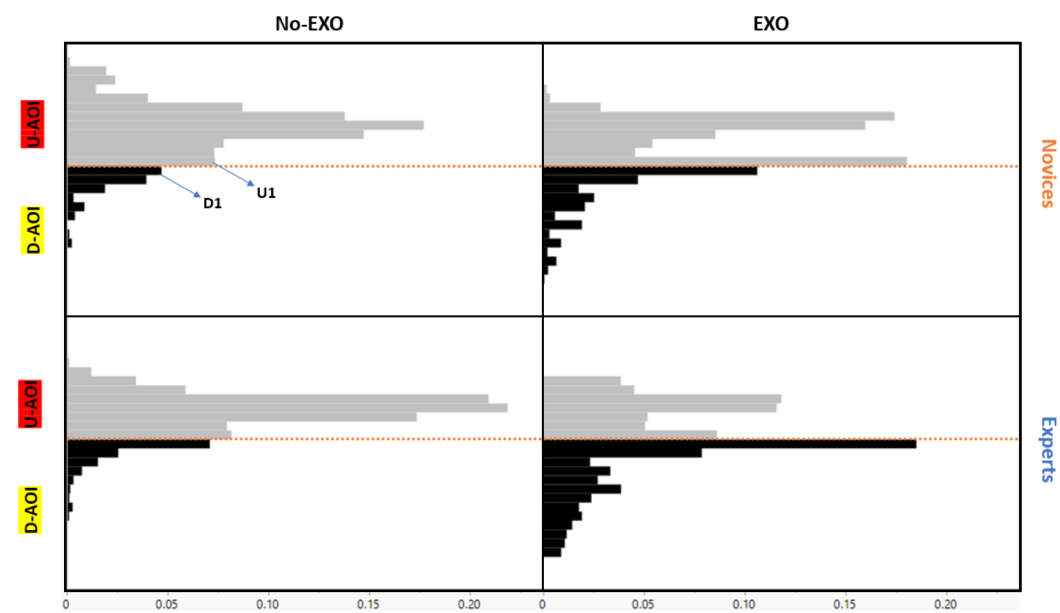


Figure 20. Ratio of fixations in upward- and downward AOI for experts and novices (S1 only). Individual bars represent the ratio of fixations in a single AOI, with black bars indicating downward-AOI and grey bars indicating upward-AOI. The orange dotted lines represent the position of the endpoint (turning point) of the walking path.

Although downward-focused gaze may improve walking stability and help in path-planning, it may also lead to reduced attention for monitoring the environment for potential hazards. In complex industrial environments, reduced attention towards the environment may put workers at greater risk of collisions with other workers and equipment. Thus, gaze behavior that serves to improve walking stability may, in fact, put the exoskeleton operator at a greater risk of collisions with surrounding objects. Future research should aim to quantify the potential impact of downward-focused gaze on safety, using methodologies such as peripheral detection tasks and dual-tasking while using powered exoskeletons.

In closing, if the mental workload induced by using an assistive device is such that the user is unable to perform a parallel task or fully exploit the capabilities of the device, this could eventually lead to its abandonment (Batavia & Hammer, 1990). Devices that are too complex to use could also negatively impact the user's acceptance of the device, and motivation to further use it: for example, prostheses users that had to compensate for the slowness and cumbersome controls of a prosthetic not only showed increased mental workload, they also abandoned the use of the prosthetic over time (Maclean & Pound, 2000). Our results indicate that while the cognitive processing demands of using a whole-body powered exoskeleton may have attenuated with increased adaptation to the device (PD), both expert and novice users of the technology experienced altered visual exploration patterns and visuomotor strategies, needed for successful performance. While this was the case for a routine everyday task like gait, how such mental demands may be influenced by using exoskeletons to perform more complex tasks needs to be understood in future work.

3.5 Limitations

A main limitation of this study was its small sample size, which was the result of a limited participant pool (company employees only, and a limited range of statures compatible with the EXO). Also, we cannot rule out the potential confounding effects of physical workload on PD with absolute certainty - it is possible that novices may have exerted greater physical effort than experts due to not having developed efficient movement patterns in the EXO. Lastly, a relatively large proportion of eye-tracker data (up to 42% per trial) were lost due to extreme downward gaze angles while turning in the exoskeleton. Thus, we could only analyze the straight-line walking paths, and not analyze potentially interesting gaze differences between novices and experts during the turning phases that may have been more challenging to execute. Future studies may use eye-trackers with wider fields of view that can better accommodate for extreme gaze angles.

3.6 Conclusion

To the best of our knowledge, this is the first study to investigate cognitive demands associated with walking in a whole-body powered exoskeleton. We found that novices' PD was significantly higher while using the exoskeleton, suggesting higher cognitive demands. The proportion of path-fixations was higher in experts, suggesting a potential stability-improving strategy that was learned over time. These results suggest that eye-tracking metrics could be used to identify expert strategies for using complex robotic devices, and also to monitor users' skill level over time as they learn to use devices such as powered exoskeletons. The ability to distinguish between expert and novice users, as well as changes in skill level, may promote the use of eye-tracking metrics as potential input variables for adaptive control strategies for powered exoskeletons.

References

1. Batavia, A. I., & Hammer, G. S. (1990). Toward the development of consumer-based criteria for the evaluation of assistive devices. *The Journal of Rehabilitation Research and Development*, *27*(4), 425. https://doi.org/10.1682/JRRD.1990.10.0425
2. Bequette, B., Norton, A., Jones, E., & Stirling, L. (2020). Physical and Cognitive Load Effects Due to a Powered Lower-Body Exoskeleton. *Human Factors*, *62*(3), 411–423. https://doi.org/10.1177/0018720820907450
3. Bridger, R. S., Ashford, A. I., Wattie, S., Dobson, K., Fisher, I., & Pisula, P. J. (2018). Sustained attention when squatting with and without an exoskeleton for the lower limbs. *International Journal of Industrial Ergonomics*, *66*, 230–239. https://doi.org/10.1016/j.ergon.2018.03.005
4. Cognolato, M., Atzori, M., & Müller, H. (2018). Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances. *Journal of Rehabilitation and Assistive Technologies Engineering*, *5*, 2055668318773991. https://doi.org/10.1177/2055668318773991
5. Cornwall, W. (2015). In pursuit of the perfect power suit. *Science*, *350*(6258), 270–273. https://doi.org/10.1126/science.350.6258.270
6. Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, *25*, 69–91. https://doi.org/10.1016/j.dcn.2016.11.001

7. Foroughi, C. K., Sibley, C., & Coyne, J. T. (2017). Pupil size as a measure of within-task learning. *Psychophysiology*, *54*(10), 1436–1443. https://doi.org/10.1111/psyp.12896
8. Heller, B. W., Datta, D., & Howitt, J. (2000). A pilot study comparing the cognitive demand of walking for transfemoral amputees using the Intelligent Prosthesis with that using conventionally damped knees. *Clinical Rehabilitation*, *14*(5), 518–522.
9. Jacobs, D. A., Koller, J. R., Steele, K. M., & Ferris, D. P. (2018). Motor modules during adaptation to walking in a powered ankle exoskeleton. *Journal of NeuroEngineering and Rehabilitation*, *15*(1), 2. https://doi.org/10.1186/s12984-017-0343-x
10. Jahn, K., Strupp, M., Krafczyk, S., Schüler, O., Glasauer, S., & Brandt, T. (2002). Suppression of eye movements improves balance. *Brain*, *125*(9), 2005–2011. https://doi.org/10.1093/brain/awf204
11. Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 56–88. https://doi.org/10.1080/14639220210159735
12. Kahya, M., Wood, T. A., Sosnoff, J. J., & Devos, H. (2018). Increased Postural Demand Is Associated With Greater Cognitive Workload in Healthy Young Adults: A Pupillometry Study. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00288
13. Kim, S., Srinivasan, D., Nussbaum, M. A., & Leonessa, A. (2021). Human Gait During Level Walking With an Occupational Whole-Body Powered Exoskeleton: Not Yet a Walk in the Park. *IEEE Access*, *9*, 47901–47911. https://doi.org/10.1109/ACCESS.2021.3068836
14. Knaepen, K., Marusic, U., Crea, S., Rodríguez Guerrero, C. D., Vitiello, N., Pattyn, N., Mairesse, O., Lefeber, D., & Meeusen, R. (2015). Psychophysiological response to cognitive workload during symmetrical, asymmetrical and dual-task walking. *Human Movement Science*, *40*, 248–263. https://doi.org/10.1016/j.humov.2015.01.001
15. Koren, Y., Mairon, R., Sofer, I., Parmet, Y., Ben-Shahar, O., & Bar-Haim, S. (2022). Vision, cognition, and walking stability in young adults. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-021-04540-w
16. Lee, Y., Jung, K.-T., & Lee, H.-C. (2022). Use of gaze entropy to evaluate situation awareness in emergency accident situations of nuclear power plant. *Nuclear Engineering and Technology*, *54*(4), 1261–1270. https://doi.org/10.1016/j.net.2021.10.022
17. Li, M., Zhong, B., Liu, Z., Lee, I.-C., Fylstra, B. L., Lobaton, E., & Huang, H. H. (2019). Gaze Fixation Comparisons Between Amputees and Able-bodied Individuals in Approaching Stairs and Level-ground Transitions: A Pilot Study. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3163–3166. https://doi.org/10.1109/EMBC.2019.8857388
18. Lindner, H., Hill, W., Norling Hermansson, L., & Lilienthal, A. J. (2020). *Cognitive load in learning to use a multi-function hand*. MEC20, Fredericton, New Brunswick, Canada, August 10-13, 2020. (Symposium canceled). http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-84900
19. Maclean, N., & Pound, P. (2000). A critical review of the concept of patient motivation in the literature on physical rehabilitation. *Social Science & Medicine (1982)*, *50*(4), 495–506. https://doi.org/10.1016/s0277-9536(99)00334-2
20. Marchand, C., De Graaf, J. B., & Jarrassé, N. (2021). Measuring mental workload in assistive wearable devices: A review. *Journal of NeuroEngineering and Rehabilitation*, *18*(1), 160. https://doi.org/10.1186/s12984-021-00953-w
21. Martin, M. S., Huard-Nicholls, B., & Johnson, A. P. (2021). Gaze and pupil size variability predict difficulty-level and safe intersection crosses in a driving simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1). https://trid.trb.org/view/1892017
22. Miller, W. C., Speechley, M., & Deathe, B. (2001). The prevalence and risk factors of falling and fear of falling among lower extremity amputees. *Archives of Physical Medicine and Rehabilitation*, *82*(8), 1031–1037. https://doi.org/10.1053/apmr.2001.24295

23. Morgan, S. J., Hafner, B. J., & Kelly, V. E. (2016). The effects of a concurrent task on walking in persons with transfemoral amputation compared to persons without limb loss. *Prosthetics and Orthotics International*, *40*(4), 490–496. https://doi.org/10.1177/0309364615596066
24. Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T. C., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods*, *52*(3), 1140–1160. https://doi.org/10.3758/s13428-019-01307-0
25. Olsen, A. (2012). *The Tobii I-VT Fixation Filter*. 21.
26. Park, H., Kim, S., Nussbaum, M. A., & Srinivasan, D. (2022). Effects of using a whole-body powered exoskeleton during simulated occupational load-handling tasks: A pilot study. *Applied Ergonomics*, *98*, 103589. https://doi.org/10.1016/j.apergo.2021.103589
27. Park, H., Kim, S., Nussbaum, M. A., & Srinivasan, D. (2023). A pilot study investigating motor adaptations when learning to walk with a whole-body powered exoskeleton. *Journal of Electromyography and Kinesiology*, *69*, 102755. https://doi.org/10.1016/j.jelekin.2023.102755
28. Sailer, U. (2005). Eye-Hand Coordination during Learning of a Novel Visuomotor Task. *Journal of Neuroscience*, *25*(39), 8833–8842. https://doi.org/10.1523/JNEUROSCI.2658-05.2005
29. Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data. *Human Factors*, *49*(3), 347–357. https://doi.org/10.1518/001872007X196685
30. Shiferaw, B. A. (2018). Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *SCIENTIfIC REPOrTS*, 10.
31. Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews*, *96*, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007
32. Walsh, G. S. (2021). Visuomotor control dynamics of quiet standing under single and dual task conditions in younger and older adults. *Neuroscience Letters*, *761*, 136122. https://doi.org/10.1016/j.neulet.2021.136122
33. White, M. M., Zhang, W., Winslow, A. T., Zahabi, M., Zhang, F., Huang, H., & Kaber, D. B. (2017). Usability Comparison of Conventional Direct Control Versus Pattern Recognition Control of Transradial Prostheses. *IEEE Transactions on Human-Machine Systems*, *47*(6), 1146–1157. https://doi.org/10.1109/THMS.2017.2759762
34. Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17. https://doi.org/10.1080/00140139.2014.956151
35. Zahabi, M., White, M. M., Zhang, W., Winslow, A. T., Zhang, F., Huang, H., & Kaber, D. B. (2019). Application of Cognitive Task Performance Modeling for Assessing Usability of Transradial Prostheses. *IEEE Transactions on Human-Machine Systems*, *49*(4), 381–387. https://doi.org/10.1109/THMS.2019.2903188
36. Zhang, W., White, M., Zahabi, M., Winslow, A. T., Zhang, F., Huang, H., & Kaber, D. (2016). Cognitive workload in conventional direct control vs. Pattern recognition control of an upper-limb prosthesis. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 002335–002340. https://doi.org/10.1109/SMC.2016.7844587

Chapter 4. Mental Workload and Gaze Behavior of Younger and Older Adults during Robot-assisted Object Co-manipulation

Abstract

Physically-coupled robots (cobots) are anticipated to provide significant benefits to older industrial workers, by augmenting their physical abilities, reducing the harmful impact of strenuous physical work, and enabling older workers to remain engaged in the workforce for an extended period of time. However, the potential cognitive impact and learning rates associated with using a cobot with complex motor control dynamics, among older adults, is not well-studied. We conducted a study to investigate the differences in performance and mental workload between younger (18-35 years) and older (45-70 years) adults, while they learned to perform a virtual object pick-and-place task using a cobot. Task difficulty was manipulated by changing the joint stiffness of the robot to increase motor-coordination demands, and by enabling or disabling objects in the virtual task environment to introduce the possibility of collisions with those objects. Eye-tracking measures were used to quantify mental workload and visuomotor strategies. Younger and older adults showed similar performance, learning rates and perceived workload, but eye-tracking measures suggested compensatory effort or changes in visual strategies in older adults for maintaining performance. Pupil dilation (PD) increased with robot-difficulty and reduced over learning for both groups, although older adults exhibited higher levels of PD in the initial trials, indicating high mental effort. Fixation rate (FR) increased for both younger and older adults when object-collisions were introduced. Stationary gaze entropy (SGE) increased, and gaze transition entropy (GTE) reduced with task difficulty, and the drop in GTE was higher in older adults when collisions were introduced. Fixations on the manipulated object reduced, and fixations on the robot increased for both age groups when task difficulty increased. Generally, measures of gaze behavior were more sensitive to the likelihood of collisions, and PD was more sensitive to robot-difficulty in both groups. This work provides evidence that older adults may use compensatory visuomotor strategies to exhibit similar performance to younger adults when using cobots to perform moderately difficult tasks, with consequentially higher mental workload. The study also demonstrates the utility of eye-tracking measures to detect age-related differences in workload and visuomotor strategies

during short-term adaptations to cobots. Future research should investigate the generalizability of these results over more diverse tasks and longer-term adaptations.

Keywords: Aging, Motor Learning, Gaze Transition Entropy, Working Memory

4.1 Introduction

An important component of developing robust measures of mental workload is determining the extent to which these measures may be sensitive across diverse populations that may differ in age, gender, and cognitive characteristics. Age is a particularly important and relevant variable in the context of motor skill learning and physical human-robot interaction. Based on relevant findings from the literature, older adults may need to exert greater mental effort for planning their movements (Coats et al., 2016), and they may also engage in greater conscious movement control (Seidler et al., 2010), thus leading to an overall greater mental workload compared to younger adults in motor control tasks. Older adults may also exhibit poorer motor performance and learning due to a lower working memory capacity or the inability to effectively engage their working memory resources (Seidler et al., 2012). Specifically, age-related reduction in working memory capacity and spatial abilities were found to be associated with performance decrements for older adults while using a robotic manipulator (Paperno et al., 2019). Thus, although one of the leading arguments in favor of developing industrial collaborative robots is that they may be able to diversify the workforce and help older adults re-enter the workforce later in their lives (Calzavara et al., 2020), there is little research investigating the potential mental demands that older adults may experience while using cobots.

Theoretical models of cognitive and neural aging have been proposed to explain age-related changes in executive functioning: The HAROLD (Hemispheric Asymmetry Reduction in Older Adults), and CRUNCH (Compensation-Related Utilization of Neural Circuits Hypothesis) models are two such models hypothesizing that task difficulty levels produce significant age-related differences in mental workload (and performance) due to differential neural activation (Cabeza, 2002; Reuter-Lorenz & Cappell, 2008). These models state that while age-related compensatory activation of neural circuits can be efficient at lower levels of task difficulty, as task demands increase, a resource ceiling might be reached, leading to insufficient processing and a decrement in performance relative to young adults. Thus, older adults may progress from over-activation at lower levels of task demand to under-activation at higher levels of task demand, due to a resource ceiling. These models have been supported by empirical work in visuo-spatial working memory tasks (Nagel et al., 2009; Piefke et al., 2012; Vermeij et al., 2014), but they have been unexplored in the context of motor-cognitive tasks. Furthermore, some work in specific skilled populations (such as pilots) has also shown that some of the cognitive deficits associated with aging are offset by older adults either exerting themselves to a greater extent, or by greater experience, i.e., long-term skill acquisition over years of experience (Causse et al., 2019; Lassiter et al., 1997).

In terms of mental workload assessments during task performance, age-related differences in the cost of dual-tasking have been extensively reported (Bock & Schneider, 2002; K. Z. H. Li et al., 2005; McDowd et al., 1991; Riby et al., 2004). Previous research in driving and basic motor-control has also reported some notable differences in eye-tracking measures between younger and older adults in response to task demands. In an object pick-and-place task, older adults fixated more frequently on future target locations, indicating that they required more advance visual information from the environment to plan their movements (Coats et al., 2016). The same study also found that when task demands increased (by making the pick-up location unstable), older adults gazed at the pick-up location for a longer duration than younger adults. In a simulated driving task, older adults exhibited a greater reduction in GTE compared to younger adults when they were distracted by a secondary visuo-spatial task (Schieber & Gilland, 2008). Lastly, a study found that older adults' pupils may be less sensitive to changes in workload, possibly due to age-related degeneration of the pupil dilation muscles, which can reduce overall pupil size and dilation response (Van Gerven et al., 2004). This study employed a working-memory-based task to manipulate workload and included adults that were quite old (mean 69, SD 4 years), and it is unclear whether similar trends in pupillary response may be observed among aging individuals (~>50 years) performing other tasks.

The aims of our study were to investigate the differences in performance and mental demands between young and old adults using an industrial cobot, to perform virtual object-manipulation tasks. We designed a task that allowed us to study the multidimensionality of workload in pHRI, i.e. the varied influences of the difficulty associated with controlling a robot, as well as the inherent task-difficulty (in terms of monitoring for collisions, performing complex movement patterns, and dealing with time- and precision-demands). The object manipulation tasks were designed to be virtual in this study since physically manipulating a variety of objects for prolonged durations while employing a cobot (that was difficult to control) presented various practical concerns related to safety and physical fatigue, especially for the older adults. There were also logistical concerns related to organizing a variety of task scenarios presenting different target and collision arrangements that changed dynamically on a trial-to-trial basis. Hence, similar to study 1, participants performed timed tasks that required the use of a bimanual cobot to pick and place virtual objects at different target locations. This study manipulated workload by varying the control-difficulty of the robot, as well as the complexity of the task environment (presence or absence of other virtual objects that required collision avoidance). Several eye-tracking metrics such as pupil dilation, fixation rate, fixations in different areas of interest (AOI), and gaze entropies were computed. Additionally, we measured participants' working memory capacity (WMC) using the operation-span (Ospan) and Symmetry Span (SymSpan) tasks and included both scores as covariates in our model. The following hypotheses were generated and tested in this study:

H1. Task performance and learning rates were expected to vary according to task difficulty, with no significant age differences (based on prior literature, we expected older adults to be able to perform just as well as younger adults, but with a greater cost, as reflected by mental workload).

H2. However, older adults were expected to experience elevated mental workload in higher task difficulties (task x group interaction), as evidenced by higher ratings on NASA-TLX, higher pupil dilation (PD), fixation rate (FR), stationary gaze entropy (SGE) and gaze transition entropy (GTE). Measures were expected to change in the direction of reduced workload over the course of learning.

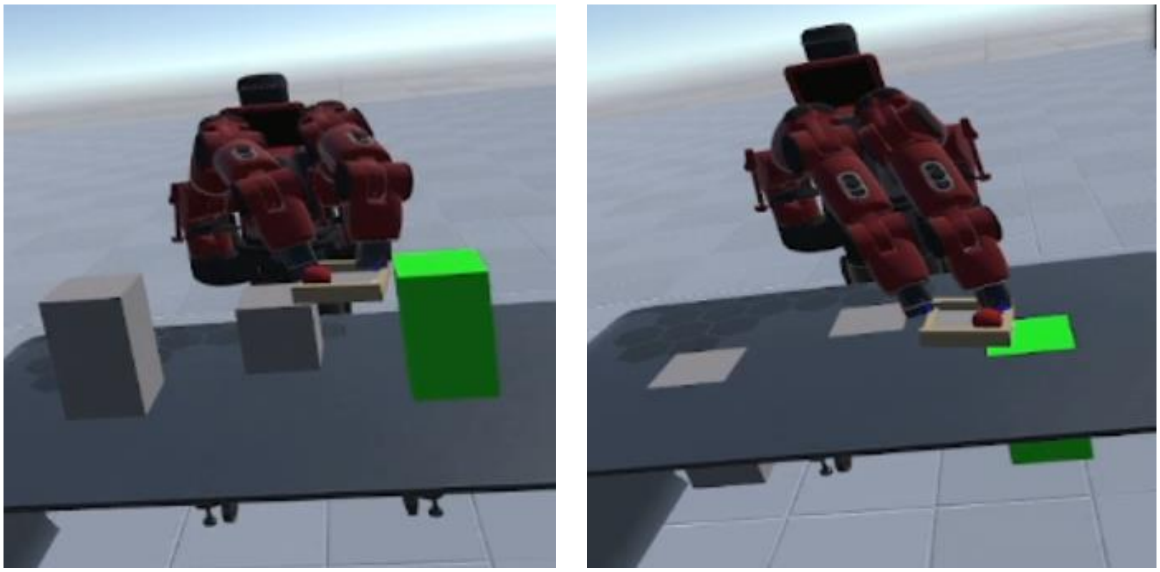
In addition, we also aimed to explore whether robot control or task environmental difficulty had selective effects on the different age groups, and any potential age-related differences in visual attention on different areas of interest, to infer whether younger and older adults utilized different strategies to accomplish the tasks.

4.2 Methods

4.2.1 Study Task and Experimental Setup

The experimental setup was largely similar to that used in Study 1, with some modifications (figure 8). Briefly, participants performed a timed bimanual object-manipulation task using a Rethink Robotics Baxter collaborative robot under two levels of robot control difficulty. All task components, namely, a virtual plate with a ball inside located on a virtual table, three target locations (towers in the form of solid gray blocks), and a digital twin of the Baxter robot were visualized in Virtual Reality (VR) using an HTC Vive Pro Eye headset (*VIVE Pro Eye Overview | VIVE United States*, 2022). Participants were asked to pick up the virtual plate and transfer it to the highlighted green location (any one of the three targets chosen randomly) within 10 seconds, while avoiding collisions and ensuring placement accuracy. Some modifications were introduced in this study, compared to Study 1 – to control for the effects of physical workload and fatigue, especially for the older adult group, the duration of each trial was reduced from 5 minutes to 3 minutes. Additionally, new end-effectors for the robot were designed and fitted with ergonomic bicycle handles to help reduce physical exertion in participants' wrists and make the interaction with virtual objects more naturalistic. Lastly, we modified the virtual environment such that two task environment difficulty conditions were presented – in one condition, three target towers were visible (participants completed a transfer to one of the three towers following a target-highlight cue, requiring them to avoid potential collisions with the other towers), and in the other condition, the targets were collapsed in height and presented at table height, so as to eliminate the possibility of collisions (Figure 21). These two conditions, referred to as "Col" (where the target towers were visible and collisions were possible) and

"NoCol" (where the targets were collapsed and collisions were not possible) respectively allowed us to distinguish between the effects of robot control difficulty vs. task-environmental collision-monitoring difficulty on gaze behavior.



a) b)

Figure 21. a) Targets in the Up/Col condition, where collisions were possible b) Targets in the Down/NoCol condition where the possibility of collisions was eliminated

4.2.2 Participants

We recruited a sample of 14 older adults (45-70 years; 10F and 4M) with females' mean age (SD) = 56 (4.64) years and males' mean age (SD) = 51 (4.08) years, and 15 younger adults (18-35 years; 7 F and 8M) with females' mean age (SD) = 26 (5.9) years and males' mean age (SD) = 28 (3.6) years from the local community and university in Clemson, SC. These age ranges were chosen to maximize age-related differences between the two groups, based on evidence that cognitive abilities, such as memory, reasoning, spatial visualization and speed of processing decline after the age of 40 (Paperno et al., 2019; Salthouse, 2009). One of the participants in the older group elected to stop the study mid-way, claiming to have attention-deficit-hyperactive-disorder (ADHD) because of which the task felt too "tedious", so data for that participant was not included in the analysis, resulting in a final sample size of 13 older adults. In terms of inclusion criteria, participants were required to be able to perform common daily tasks such as housework or desk organization without the use of corrective lenses, since corrective lenses could interfere with the eye-tracking equipment, and to be free of any recent history of musculoskeletal disorders (past 12 months). We excluded individuals with a history of migraine, vertigo and epilepsy, since these conditions can increase the susceptibility to VR sickness. In addition, we requested participants to limit their consumption of caffeine or nicotine 2 hours prior, and alcohol or sedative drugs 24 hours prior to the experiment to control for influences on pupil diameter (Matthews et al., 2015), and to not wear any makeup (especially mascara) to avoid any undesirable effects on eye tracking accuracy (Holmqvist et al., 2011).

4.2.3 Experiment Design and Protocol

This study employed a mixed-factor design, with difficulty (manipulated variable) and trial number (observed variable) as within-subjects variables, and age-group as the between-subjects variable. Overall task difficulty was manipulated in two different ways – robot-difficulty and target-difficulty. Robot-difficulty was manipulated by changing the degree of "match" between the joint impedances of the two arms of the robot, with two levels, low difficulty robot (LDR; matched impedances) and high difficulty robot (HDR; mis-matched impedances). The mis-matched condition was expected to increase motor-coordination demands and consequently increase the difficulty of balancing the ball on the virtual plate and the demand associated with collision-avoidance. Secondly, within each robot-difficulty, the three target pillars switched between the down/NoCol and up/Col conditions. The NoCol condition was expected to be lower in difficulty, since the possibility of collisions was eliminated. Trial number (with six levels) was included as an observed independent variable to test the effect of practice/learning on our dependent measures.

Participants completed a first familiarization session followed by four experimental blocks, each comprised of three 3-minute trials. The first and second blocks were performed in the order of LDR followed by the HDR condition, and the order of difficulty (LDR vs. HDR) was counterbalanced for the third and the fourth blocks. After all four blocks were completed, we asked participants a few questions about their overall experience doing the task over time, potential strategies they employed, and where they focused their attention during the task. Lastly, we assessed both younger and older participants' working memory capacity (WMC).

4.2.4 Data Collection and Processing

Demographic data (age, gender) and any prior VR experiences were collected from each participant prior to beginning the study. The operation span (OSpan) and symmetry span (SymSpan) tasks were collected as measures of WMC (Foster et al., 2015). The OSpan task requires participants to remember and recall a sequence of numbers while also solving a series of math operations (Stone & Towse, 2015). The math operation appears immediately after each digit in the to-be-remembered sequence and acts as a distractor. The SymSpan task is similar in structure to the OSpan task, the difference being that the to-be-remembered items are locations of blue squares that appear in a 4×4 grid of potential locations, and the distractor task involves judging whether a displayed shape is symmetrical or not along its vertical axis. We followed the scoring technique as detailed in (Conway et al., 2005) and computed WMC scores as the proportion of correctly-recalled items out of the total number of items presented to the participant in each separate WMC task. The final WMC score was computed as the average of the Ospan and SymSpan score for each participant. A free and open-source version of the WMC task (Stone & Towse, 2015) designed using the Training and Testing Tool (Tatool) computer software (von Bastian et al., 2013) was administered to participants, and task data were exported in .csv format using the available scripts written for the R statistical software (Stone & Towse, 2015).

A custom Unity script was used to record pupil size and 3-D coordinates of the gaze point at 90 Hz from the eye-tracker embedded within the VR headset. When the participant looked at any VR object (also referred to as an area of interest or AOI), the gaze point intersected with the surface of the object, thus registering as a 'hit' on the AOI. Each eye-tracking sample was associated with an AOI-hit. The script also recorded the 3-D coordinates of the virtual plate, ball, and the virtual Baxter's end-effectors, as well as the instantaneous timer value corresponding to transfer completion. NASA-TLX ratings were obtained after each trial to assess perceived workload. Fixation rate (FR), median pupil dilation (PD), stationary gaze entropy (SGE), and gaze transition entropy (GTE) were computed for each transfer within each trial, using procedures described in Study 1 (page 41). The ratios of fixations on specific AOIs, i.e., the plate (which included the ball), the robot-arms, and the top surfaces of the targets, were also computed for each transfer, to understand the relative attentional focus directed towards these AOI.

4.2.5 Outcome Measures

FR was computed by dividing the number of fixations in each transfer by the duration of the transfer. We used PD, FR, SGE and GTE as eye-tracking measures of mental workload and visual monitoring, and the raw NASA-TLX rating (R-TLX) as a measure of perceived workload. The R-TLX was computed as the average rating on all NASA-TLX subscales (Hart, 2006). Performance was quantified as the ratio of successful transfers per trial.

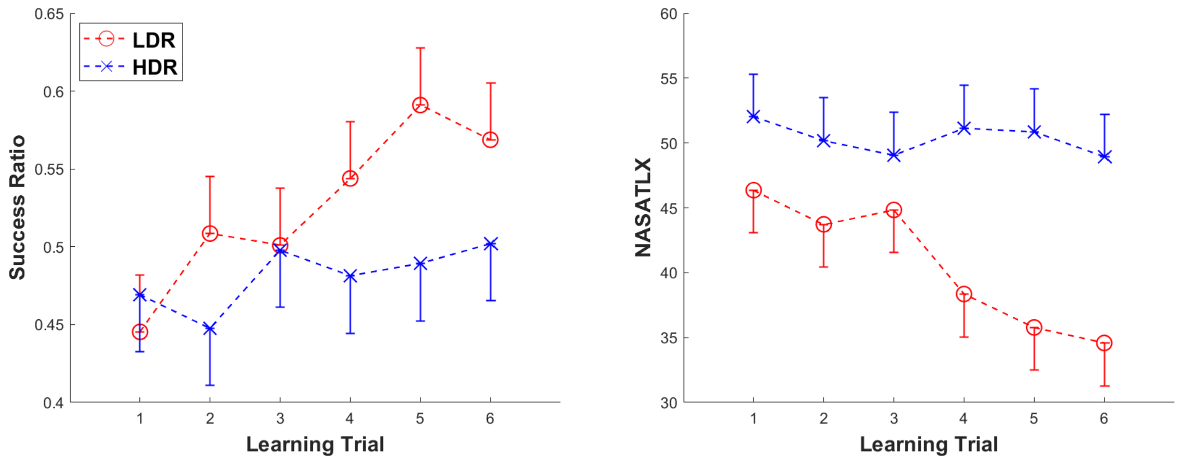
4.2.6 Statistical Analysis

Since task difficulty was influenced by the robot-condition (LDR vs HDR) as well as the target-position (NoCol, Col), we created a "Difficulty" variable with four levels (1: LDR-NoCol; 2: HDR-NoCol; 3: LDR-Col; 4: HDR-Col). Additionally, since we were interested in studying the changes in eye-tracking metrics in each difficulty condition, we combined the transfers performed in each difficulty condition in each trial to create a "Trial" variable (six levels for each of the four difficulty conditions). Separate analyses of variance were conducted to evaluate the effects of the independent variables on the three classes of measures (performance, perceived workload, and eye-tracking measures). Specifically, the ANOVA for eye tracking measures included Difficulty (LDR-NoCol, HDR-NoCol, LDR-Col, HDR-Col), Group (younger; older), and Trial (1 to 6) and their interactions as independent variables, and the ANOVA for performance and perceived workload included the same factors except that difficulty was replaced with 'robot difficulty' (LDR vs. HDR). Lastly, participants' WMC scores (Ospan and SymSpan) were included as covariates in all models. Partial eta-squared (η_p^2) was used to quantify effect sizes. The significance level was set at $\alpha = 0.05$, and all statistical analyses were performed in JMP (SAS Institute Inc., USA).

4.3 Results

4.3.1 Performance and Perceived Workload

The ratio of successes generally increased from trial 1 to trial 6 (Figure 22a), but the increase was statistically significant only in the LDR condition. Perceived workload was significantly higher in the HDR condition than in the LDR condition, across both age-groups. Post-hoc tests and significant interactions indicated that workload reduced in the LDR condition from trial 1 to trial 6, but remained high in the HDR condition. The Ospan covariate had a significant effect on perceived workload. These results are presented in Table 6.



a) b)

Figure 22. **a)** Performance and **b)** Perceived Workload over the course of six learning trials for the two robot-difficulty conditions (LDR, HDR) pooled across all subjects. Individual data points represent least squares means, and the error bars represent 1 standard error.

Table 6

a) p-values and effect sizes (quantified using partial eta-squared) for performance and perceived workload measures. Bold fonts indicate significant effects (p<0.05). Columns with covariates include p-values only.

Dependent Measures	p-value (Effect Size - η_p^2)								
	Condition	Trial	Difficulty x Trial	Group	Group x Condition	Group x Trial	Group x Condition x Trial	Covariate	
								Ospan	SymSpan
Success Ratio	.4336 (.002)	**.0006 (.073)**	**.0439 (.039)**	.9891 (3.21e-6)	.4584 (.002)	.7263 (.0098)	.8892 (.006)	.3186	.3731
Nasa-TLX	**.0265 (.017)**	**.0005 (.075)**	**.0034 (.059)**	.5471 (.008)	.7308 (.0004)	.7236 (.0098)	.871 (.006)	**.0362**	.3912

b) F-statistics for performance and perceived workload measures. Bold fonts indicate significant effects (p<0.05). Degrees of freedom are shown in brackets.

Dependent Measures	Condition	Trial	Difficulty x Trial	Group	Group x Condition	Group x Trial	Group x Condition x Trial	Covariate	
								Ospan	SymSpan
Success Ratio	F (1, 285) = 0.61	**F (5, 285) = 4.5**	**F (5, 285) = 2.3**	F (1, 33.05) = 0.38	F (1, 285) = 0.55	F (5, 285) = 0.56	F (5, 285) = 0.34	F (1, 23.99) = 1.04	F (1, 23.99) = 0.8
Nasa-TLX	**F (1, 287) = 4.97**	**F (5, 286.5) = 4.62**	**F (5, 286.5) = 3.63**	F (1, 32.4) = 0.01	F (1, 287) = 0.12	F (5, 286.5) = 0.57	F (5, 286.5) = 0.37	**F (1, 25.2) = 4.9**	F (1, 25.3) = 0.76

4.3.2 Eye-tracking measures

Eye-tracking measures largely showed expected effects due to robot-difficulty and learning, but the novel age- and target-difficulty manipulations in this study resulted in some key differences. Significance values and effect sizes for all eye-tracking measures are shown in Table 7. Wherever significant 3-way interaction effects were found, e.g. for PD and plate-fixations, results were plotted separately for younger and older age groups, to help visualize the three-way interactions. For significant higher-order interactions, specific post hoc contrasts were performed, guided by our research questions and hypotheses, with a focus on age differences.

Table 7

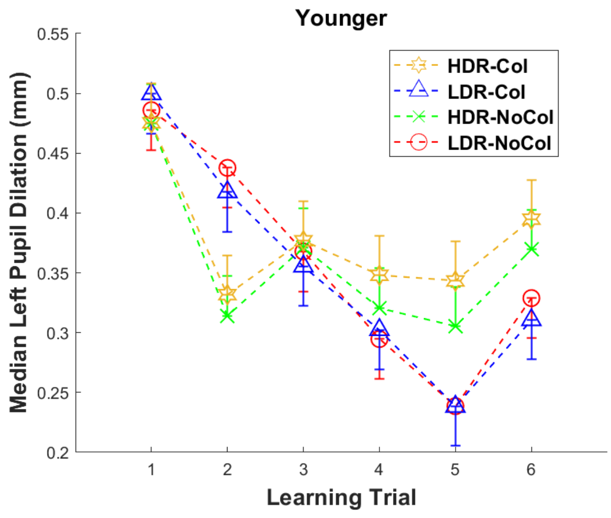
a) p-values and effect sizes (quantified using partial eta-squared) for eye-tracking measures. Bold fonts indicate significant effects (p<0.05). Columns with covariates include p-values only.

Dependent Measures	p-value (Effect Size - η_p^2)								
	Difficulty	Trial	Difficulty x Trial	Group	Group x Difficulty	Group x Trial	Group x Trial x Difficulty	Covariate	
								Ospan	SymSpan
PD (mm)	.2492 (.0003)	**<.0001 (.087)**	**<.0001 (.012)**	.78 (.00027)	**.0002 (.0015)**	**<.0001 (.006)**	**<.0001 (.013)**	.5298	**.0240**
FR (/sec)	**<.0001 (.008)**	**.0386 (.001)**	**.0058 (.0026)**	.8863 (.00001)	.06 (.0006)	.5695 (.0003)	.2962 (.001)	.9938	.2128
SGE	**.0089 (.001)**	.1651 (.0007)	.118 (.002)	.3082 (.0004)	.0736 (.0006)	.5934 (.00036)	.0693 (.002)	.9356	.8468
GTE	**<.0001 (.007)**	.9973 (.00003)	.1149 (.0019)	.842 (.00004)	**.0474 (.0007)**	.5248 (.000366)	.351 (.0014)	.2001	.5492
Plate-Fixations	**<.0001 (.03)**	.2215 (.0005)	**<.0001 (.005)**	.5685 (.0004)	**.0004 (.0015)**	**.02 (.001)**	**<.0001 (.004)**	.9317	.6381
Robot-Fixations	**<.0001 (.013)**	**.0001 (.002)**	**<.0001 (.006)**	.2944 (.00088)	.2072 (.00036)	.2607 (.0005)	.065 (.002)	.6589	.6580

b) F-statistics for eye-tracking measures. Bold fonts indicate significant effects (p<0.05). Degrees of freedom are shown in brackets

Dependent Measures	Difficulty	Trial	Difficulty x Trial	Group	Group x Difficulty	Group x Trial	Group x Trial x Difficulty	Covariate	
								Ospan	SymSpan
PD (mm)	F (3, 12561) = 1.36	**F (5, 12561) = 238.7**	**F (15, 12561) = 10.25**	F (1, 24.94) = 0.26	**F (3, 12561) = 6.5**	**F (5, 12561) = 14.6**	**F (15, 12561) = 11.13**	F (1, 23.98) = 0.4	**F (1, 23.99) = 5.8**
FR (/sec)	**F (3, 12562) = 33.26**	**F (5, 12562) = 2.33**	**F (15, 12562) = 2.16**	F (1, 29.24) = 0.2	F (3, 12562) = 2.5	F (5, 12562) = 0.77	F (15, 12562) = 1.16	F (1, 23.8) = 0.0001	F (1, 23.87) = 1.64
SGE	**F (3, 10279.7) = 3.84**	F (5, 10278.4) = 1.56	F (15, 10279.3) = 1.43	F (1, 34.6) = 0.91	F (3, 10280.1) = 2.3	F (5, 10278.5) = 0.74	F (15, 10279.4) = 1.58	F (1, 23.6) = 0.007	F (1, 23.8) = 0.04
GTE	**F (3, 11389) = 28.37**	F (5, 11388.6) = 0.07	F (15, 11389) = 1.45	F (1, 27.8) = 0.03	**F (3, 11389.1) = 2.63**	F (5, 11388.6) = 0.84	F (15, 11389) = 1.1	F (1, 24) = 1.74	F (1, 24) = 0.37
Plate-Fixations	**F (3, 12561.8) = 120.6**	F (5, 12561.5) = 1.4	**F (15, 12561.8) = 4.2**	F (1, 26.9) = 0.15	**F (3, 12561.9) = 6.11**	**F (5, 12561.5) = 2.69**	**F (15, 12561.8) = 3.63**	F (1, 24) = 0.008	F (1, 24) = 0.23
Robot-Fixations	**F (3, 12561.9) = 120.6**	**F (5, 12561.5) = 1.4**	**F (15, 12561.9) = 4.19**	F (1, 27.5) = 0.15	F (3, 12562) = 6.11	F (5, 12561.5) = 2.69	F (15, 12561.9) = 3.63	F (1, 23.9) = 0.007	F (1, 23.9) = 0.23

PD showed a significant 3-way interaction effect of age, task difficulty, and learning trial (table 7). While PD significantly reduced over trials for both older and younger adults (Figure 23), there were specific differences with difficulty and learning that were explored using post hoc contrasts. Within the four



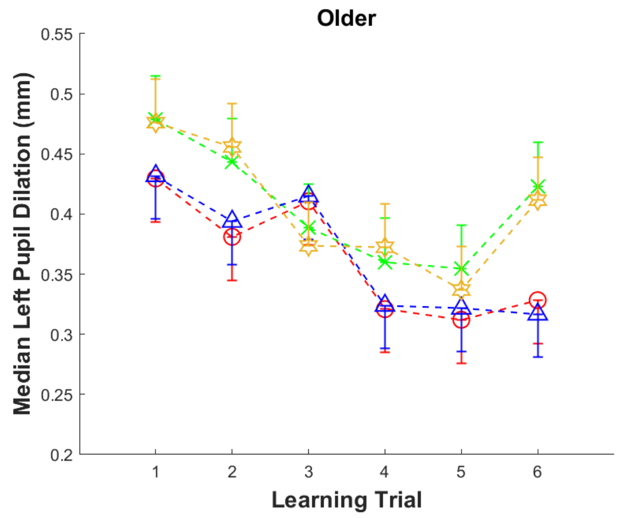


Figure 23. PD for **a)** Younger and **b)** Older groups over the course of six learning trials for each difficulty condition (LDR-NoCol; HDR-NoCol; LDR-Col; HDR-Col) pooled across all subjects. Individual data points represent least squares means, and the error bars represent 1 standard error.

difficulty levels, PD did not vary significantly with collision vs. no-collision conditions (i.e., Col vs. NoCol), but did show significant differences between the two different levels of robot difficulty (i.e., LDR vs. HDR). In the low robot-difficulty conditions (i.e., conditions LDR-NoCol and LDR-Col), the younger group's PD showed a greater reduction than the older group in the first three trials compared to the last three trials. In the high robot-difficulty conditions (i.e., conditions HDR-NoCol and HDR-Col), there was no significant age difference, however, both groups showed greater reduction in PD in the first three, compared to the last three trials. This indicated that there were greater adaptations (and resultant changes in PD) during the initial set of trials than in later trials, and this phenomenon was different between young and old groups depending on task difficulty. Also, in the final (6^{th}) trial, PD continued to be elevated in the HDR conditions compared to the LDR conditions for both age groups. Of the two working memory covariates, SymSpan showed a significant effect on PD.

FR showed a significant interaction effect of difficulty and trial, but no significant age effects (table 7). FR was significantly higher in the high target-difficulty conditions (LDR-Col and HDR-Col) compared to the low target-difficulty conditions (LDR-NoCol and HDR-NoCol) (Figure 24b). In contrast to PD that showed significant differences across robot difficulty levels, FR was not significantly different across different levels of robot-difficulty (i.e., LDR vs. HDR). A post hoc contrast revealed that the difference in FR between the NoCol and Col conditions was greater in trial 1 compared to trial 6, indicating that FR in the two conditions became more similar over time. When considering the changes across trials in each of

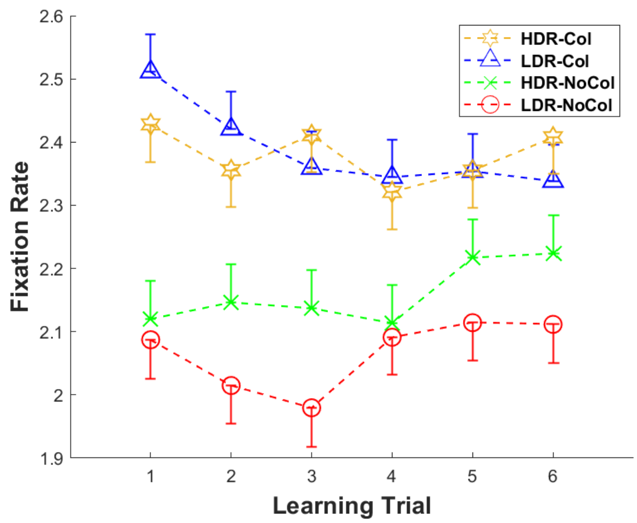


Figure 24. FR over the course of six learning trials for each difficulty condition (LDR-NoCol; HDR-NoCol; LDR-Col; HDR-Col) pooled across all subjects. Individual data points represent least squares means, and the error bars represent 1 standard error.

the four difficulty levels, we found that FR reduced over time (trials) only for the LDR-Col condition, and stayed at the same level for all other difficulties. There were no significant effects of the covariates on FR.

SGE showed a significant effect of difficulty (table 7). SGE significantly increased in the three higher-difficulty conditions compared to the lowest difficulty condition (LDR-NoCol) across both groups. GTE showed a significant interaction effect of difficulty and age. The post hoc tests indicated that GTE significantly reduced with increasing task difficulty for both groups, and the older adults experienced a greater drop in GTE in the collision (Col) conditions compared to the no-collision (NoCol) conditions. There were no statistically significant effects of the covariates on SGE and GTE.

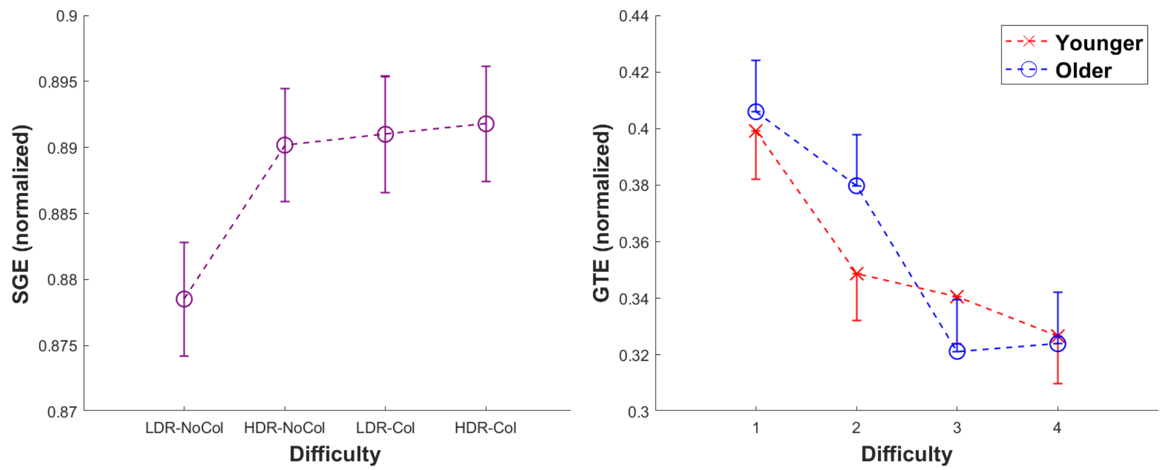
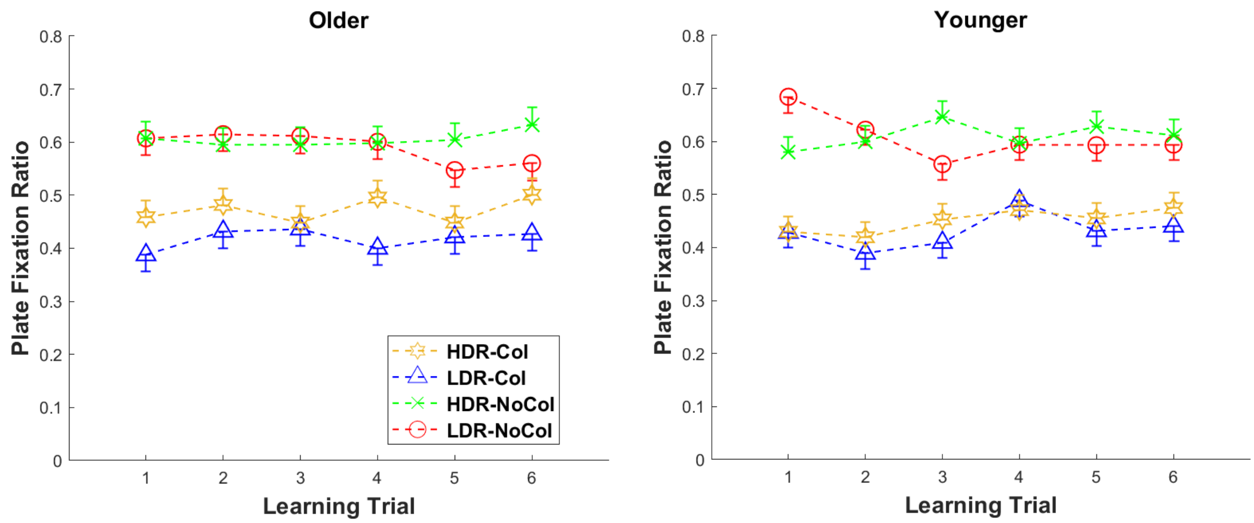


Figure 25. **a)** SGE for each difficulty condition and **b)** GTE for each difficulty condition, separated by group. Individual data points represent least squares means, and the error bars represent 1 standard error.

For plate-fixations, a significant three-way interaction effect of difficulty, trial and age was observed. Plate-fixations were significantly fewer (meaning that participants looked away from the plate) in the Col conditions compared to the NoCol conditions, for both older and younger groups; and were not different with different levels of robot difficulty. Post hoc contrasts indicated that for both age groups, there were more plate fixations in the NoCol conditions compared to the Col conditions in the first trial, but over time (by the 6th trial), this difference in plate fixations between NoCol and Col conditions had reduced, i.e., participants were able to direct their gaze more towards the plate even in the conditions where collisions were expected, as they became more familiar with the tasks.



a) **b)**

Figure 27. Plate-fixations for **a)** Younger and **b)** Older groups over the course of six learning trials for each difficulty condition (LDR-NoCol; HDR-NoCol; LDR-Col; HDR-Col) pooled across all subjects. Individual data points represent least squares means, and the error bars represent 1 standard error.

Robot-fixations were significantly greater in the Col conditions compared to the NoCol conditions for both age groups. Robot-fixations were also greater in the HDR compared to the LDR conditions, only in trial 1. Robot-fixations reduced over trials, but only for the two HDR conditions. The covariates did not have a significant effect on any AOI measures.

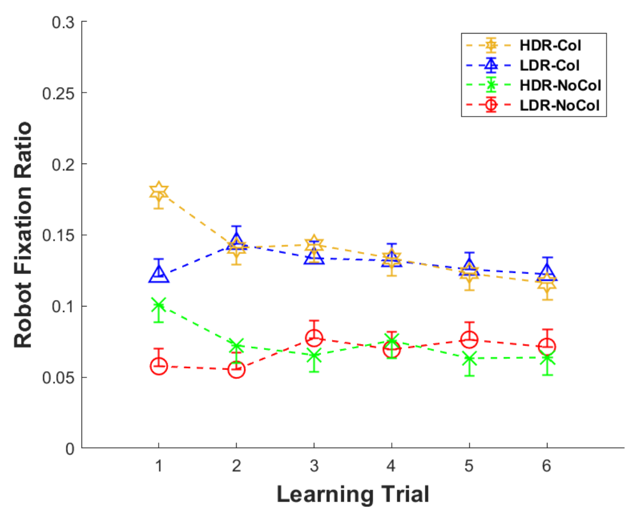


Figure 26. Robot-fixations over the course of six learning trials for each difficulty condition (LDR-NoCol; HDR-NoCol; LDR-Col; HDR-Col) pooled across all subjects. Individual data points represent least squares means, and the error bars represent 1 standard error.

4.4 Discussion

This study explored the differences in performance, perceived mental workload and eye-tracking metrics between younger and older adults under different types of task demands in physical human-robot interaction. Overall, our results indicated that participants learned the task and improved their performance over time, but that the rate of these improvements was significantly higher in the easier robot condition (LDR). Across all trials, the HDR condition was perceived to be more difficult than the LDR condition, and only the LDR condition was perceived to become less difficult over time. As expected, there were no significant age differences in performance. Furthermore, older and younger adults also did not report significantly different perceived mental workload across the different task manipulations. This largely aligns with prior work showing that in tracking tasks requiring accuracy as well as speed, older adults were impaired only on speeded tasks, but not in tasks that focused on accuracy (Etnier & Landers, 1998; McNay & Willingham, 1998). Prior work has also shown that while age differences existed in performance of some motor tasks with task difficulty (older adults exhibited greater errors in more difficult tasks), the rates of adaptation (improvements in task performance with time) was not significantly different across age (Bock & Schneider, 2002). Although the HDR condition was perceived as more difficult, our results indicate that older adults probably compensated (potentially using greater mental effort), to achieve similar levels of performance as the younger group; however, this was not perceived by them as they didn't indicate higher levels of perceived mental workload.

When considering the eye-tracking metrics, Pupil Dilation (PD) reduced over trials for both groups, indicating that it was sensitive to changes in workload over learning. Additionally, the rate of reduction of PD was different for different robot-conditions. Specifically, PD showed a larger drop from trial 1 to trial 6 in the LDR condition compared to the HDR condition for both groups. These trends aligned with perceived workload ratings, which further suggests the potential of PD as an index of task difficulty and learning. PD was also sensitive to group differences that were not captured by NASA-TLX ratings in this particular context. Specifically, younger adults' PD reduced at a significantly quicker rate than older adults'

PD in the first three trials of the LDR condition, suggesting that older adults may have sustained their mental effort for longer, at least during initial learning. Overall, PD was more sensitive to changes in learning rather than changes in task difficulty in this specific context, which aligns with our findings from Study 1.

Among the task-difficulty manipulations, PD was only sensitive to changes in robot-difficulty but not target-difficulty. The sensitivity to robot-difficulty but not target-difficulty is somewhat surprising, since multi-object tracking demands (which were expected in the high target-difficulty condition), have been shown to increase PD in other cognitive domains such as air-traffic control (Ahlstrom & Friedman-Berg, 2006). Some possible explanations for this are that: (i) the target difficulty manipulation in our collision conditions was not significantly more demanding than the no-collision conditions (as compared to the complexity of air traffic control), or (ii) that the robot difficulty manipulation caused a greater difference in cognitive effort that produced changes in PD. The latter explanation is supported by the reasoning that the higher robot difficulty (HDR) condition created a need for forming a more complex mental model for successful task performance, which may have produced a greater cognitive effort. Finally, it is also possible that there was a higher level of physical effort needed for controlling the robot in the HDR condition that may have also contributed to the increase in pupil dilation (Zénon et al., 2014).

Fixation rate (FR) increased in the high target-difficulty condition, likely due to the increased monitoring demand associated with path-planning and avoiding collisions with the targets. This effect of target-difficulty was somewhat consistent throughout the six trials, although it was slightly smaller in the sixth trial compared to the first trial, suggesting some learning effects. Considering that the targets switched between the high- and low- difficulty conditions (Col vs. NoCol) on a transfer-by-transfer basis, FR may be a useful measure of instantaneous changes in environmental complexity in pHRI. There were no visual monitoring differences evident between the two age-groups as reflected by FR.

SGE increased, and GTE decreased with increases in overall task difficulty, which aligned with our findings from Study 1. The increase in SGE indicated greater variation in the AOIs that were monitored, due to changes in task complexity driven by both the mis-matched robot arms and the need for collision avoidance. In the easiest condition, it might have been sufficient for participants to direct their gaze mostly towards the plate; however, as difficulty increased, gaze fixations likely spread to the robot and the targets in addition to the plate, thus increasing SGE. The increase in SGE is said to reflect greater "bottom-up" gaze behavior which is driven primarily by changes in the visual complexity of the environment (B. Shiferaw et al., 2019). Although a greater variety of AOIs was monitored (as indicated by SGE), the reduction in GTE indicated that the gaze transitions between those AOI became more stereotypical and ordered, similar to our finding in Study 1. Interestingly, compared to younger adults, older adults experienced a greater drop in GTE when target-difficulty increased, possibly due to age-related reduction in the ability to actively modulate attention (Gazzaley et al., 2005; B. Shiferaw et al., 2019).

In terms of AOI-based metrics, fixations were directed largely towards the plate throughout all trials; however, as the task became more difficult, participants turned their gaze away from the plate and towards the robot. This finding aligns with the increase in SGE in the more difficult conditions. Robot fixations increased due to both robot-difficulty and target-difficulty, suggesting that participants may have monitored the robot to better understand its dynamics and form a mental model, as well as for collision-avoidance. However, it is possible that mental-model formation occurred relatively quickly, since robot-difficulty effects were significant only in the first trial. A group difference was found in plate-fixations – older adults' gaze was diverted away from the plate to a significantly greater extent in the LDR-Col condition compared to the HDR-Col condition, even though the robot was easier to operate in the LDR condition. Based on observations made during the study, participants were able to transfer the plate significantly more quickly towards the targets in the LDR-Col condition, which, in turn, made it more likely that the ball would roll off the plate. These ball-drops in the LDR-Col condition may have diverted fixations away from the plate. It is possible that older adults had greater difficulty optimizing their movement speed for avoiding ball-drops, and continued to be distracted by the ball rolling off the plate. Such distractions and other intermittent events may cause instantaneous changes in gaze behavior, and although these changes may not correspond to a learned cognitive strategy, they may still be valuable sources of information of momentary changes in mental/visuomotor effort based on task demands (Aronson & Admoni, 2018).

Considering the WMC covariates, only the SymSpan score had a statistically significant effect on PD. This aligns with findings from past studies which showed that individual differences in spatial working memory capacity may influence motor task performance and learning rates (Anguera et al., 2011). Specifically, this work has proposed that the spatial component of working memory can store motor error information on a per-movement basis, which is then used to correct future movements with an updated mental model, leading to performance improvements over time.

Overall, learning in visuo-motor tasks can reflect distinctly different processes: while on one hand, improvements over time may reflect better mental model formation (or perceptual-motor recalibration), improvements in performance may also reflect movement corrections based just on sensory feedback, i.e., cognitive decisions to aim past a perceived target position, and other so-called ‘strategic processes’. Prior work examining age-related adaptations to visuo-motor tasks have reported that older adults did not differ from younger adults in mental model formation, but that there were significant age-related differences in the strategic processes utilized for performance (Fernández-Ruiz et al., 2000; McNay & Willingham, 1998; Roller et al., 2002). Our findings of age differences in specific gaze behavior metrics such as GTE support this idea that younger adults may have utilized such strategic processes to a greater extent than older adults, to try and improve performance in the higher-difficulty conditions.

4.5 Limitations and Future Work

Our age groups were closer to each other in age compared to previous studies that found overall group differences in eye-tracking workload measures (Coats et al., 2016; Heintz Walters et al., 2021; Schieber & Gilland, 2008). We did not recruit adults older than 70, since we aimed to align our sample age with the typical age of industrial workers. However, an important anticipated benefit of using cobots is that they may empower older adults well beyond retirement age to perform household tasks with minimal assistance, and even allow them to even potentially re-enter the workforce. Thus, future research may be conducted with participants older than our study sample, with due precautions taken to minimize physical workload.

This study investigates only short-term adaptations to a cobot, and it is not yet known how long-term adaptation to a cobot may affect the measures collected in this study, and the nature of any additional age-related differences in these measures that may emerge over time. Future research should focus on long-term adaptations to a cobot in a variety of tasks.

The relative contributions of foveal and peripheral vision are task-specific, and can significantly influence overall gaze patterns. Some participants mentioned that they used their "ambient vision" to determine the target to which they were supposed to transfer the plate. Thus, we cannot rule out that peripheral vision may have played a significant role in our task, not only for target-monitoring but potentially also for robot- and plate-monitoring. The situations in which peripheral vision may have played a role, or whether there were any age-differences, cannot be determined in this study. Older adults are known to experience decrements in their useful-field-of-view (UFOV) which can hinder the acquisition of visual information (B. Sekuler, 2000) and potentially lead to changes in visual monitoring strategies. Future work should account for these differences, and investigate the potential impact of UFOV on workload and gaze strategies in pHRI.

Lastly, our VR-based task environment could not reproduce the sensory feedback associated with manipulating a physical object, leading participants to solely rely on their vision for performing the task, as opposed to tactile and force-feedback. A more physically-realistic task may amplify or reduce age- or task-related effects on eye-tracking measures. The generalizability of our results to pHRI co-manipulation tasks with real, physical objects is yet to be verified.

4.6 Conclusion

This study found that older adults learned to perform a physical human-robot-interaction task just as well as younger adults, and that they may have compensated for any age-related deficits by applying greater mental effort. Although self-reported workload measures did not support this finding in our specific task and context, eye-tracking measures provided evidence that there may have been differences in workload, learning rate, and strategies across age groups. Eye-tracking measures were also diagnostic of different types of task demands.

PD was sensitive to robot-difficulty and learning, and measures of gaze behavior were generally more sensitive to target-difficulty in this study. PD and GTE reflected age-differences in learning rate and visual monitoring respectively. Overall age-group differences were not observed. FR as a context-independent measure was sensitive to instantaneous changes in target-difficulty, which highlights its use as a potential candidate for tracking or predicting changes in task environment when the task context is unknown. AOI-based measures were more sensitive to target-difficulty than FR, although they are, by definition, dependent on context and require knowledge of gaze location in the external world, thus limiting their applicability. In pHRI tasks, the use of both pupillary and gaze-behavior measures can provide multidimensional information about workload and visual behaviors due to changes in task difficulty and learning.

References

1. Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623–636. https://doi.org/10.1016/j.ergon.2006.04.002
2. Anguera, J. A., Reuter-Lorenz, P. A., Willingham, D. T., & Seidler, R. D. (2011). Failure to Engage Spatial Working Memory Contributes to Age-related Declines in Visuomotor Learning. *Journal of Cognitive Neuroscience*, *23*(1), 11–25. https://doi.org/10.1162/jocn.2010.21451
3. Aronson, R. M., & Admoni, H. (2018). *Gaze for Error Detection During Human-Robot Shared Manipulation*. Proceedings of RSS ’18 Towards a Framework for Joint Action Workshop.
4. B. Sekuler, P. J. B., Mortimer Mamelak, Allison. (2000). Effects of Aging on the Useful Field of View. *Experimental Aging Research*, *26*(2), 103–120. https://doi.org/10.1080/036107300243588
5. Bock, O., & Schneider, S. (2002). Sensorimotor adaptation in young and elderly humans. *Neuroscience & Biobehavioral Reviews*, *26*(7), 761–767. https://doi.org/10.1016/S0149-7634(02)00063-5
6. Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: The HAROLD model. *Psychology and Aging*, *17*, 85–100. https://doi.org/10.1037/0882-7974.17.1.85
7. Calzavara, M., Battini, D., Bogataj, D., Sgarbossa, F., & Zennaro, I. (2020). Ageing workforce management in manufacturing systems: State of the art and future research agenda. *International Journal of Production Research*, *58*(3), 729–747. https://doi.org/10.1080/00207543.2019.1600759

8. Causse, M., Chua, Z. K., & Rémy, F. (2019). Influences of age, mental workload, and flight experience on cognitive performance and prefrontal activity in private pilots: A fNIRS study. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-44082-w
9. Coats, R. O., Fath, A. J., Astill, S. L., & Wann, J. P. (2016). Eye and hand movement strategies in older adults during a complex reaching task. *Experimental Brain Research*, *234*(2), 533–547. https://doi.org/10.1007/s00221-015-4474-7
10. Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772
11. Etnier, J. L., & Landers, D. M. (1998). Motor performance and motor learning as a function of age and fitness. *Research Quarterly for Exercise and Sport*, *69*(2), 136–146. https://doi.org/10.1080/02701367.1998.10607679
12. Fernández-Ruiz, J., Hall, C., Vergara, P., & Díaz, R. (2000). Prism adaptation in normal aging: Slower adaptation rate and larger aftereffect. *Cognitive Brain Research*, *9*(3), 223–226. https://doi.org/10.1016/S0926-6410(99)00057-9
13. Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236. https://doi.org/10.3758/s13421-014-0461-7
14. Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature Neuroscience*, *8*(10), Article 10. https://doi.org/10.1038/nn1543
15. Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909
16. Heintz Walters, B., Huddleston, W. E., O'Connor, K., Wang, J., Hoeger Bement, M., & Keenan, K. G. (2021). The role of eye movements, attention, and hand movements on age-related differences in pegboard tests. *Journal of Neurophysiology*, *126*(5), 1710–1722. https://doi.org/10.1152/jn.00629.2020
17. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press. http://lup.lub.lu.se/record/1852359
18. Lassiter, D. L., Morrow, D. G., Hinson, G. E., Miller, M., & Hambrick, D. Z. (1997). Expertise and age effects on pilot mental workload in a simulated aviation task. In *Designing for an aging population: Ten years of human factors/ergonomics research* (pp. 226–230). Human Factors and Ergonomics Society.
19. Li, K. Z. H., Krampe, R. Th., & Bondar, A. (2005). An Ecological Approach to Studying Aging and Dual-Task Performance. In *Cognitive limitations in aging and psychopathology* (pp. 190–218). Cambridge University Press. https://doi.org/10.1017/CBO9780511720413.009
20. Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The Psychometrics of Mental Workload: Multiple Measures Are Sensitive but Divergent. *Human Factors*, *57*(1), 125–143. https://doi.org/10.1177/0018720814539505
21. McDowd, J., Vercruyssen, M., & Birren, J. E. (1991). Aging, divided attention, and dual-task performance. In *Multiple Task Performance*. CRC Press.
22. McNay, E. C., & Willingham, D. B. (1998). Deficit in learning of a motor skill requiring strategy, but not of perceptuomotor recalibration, with aging. *Learning & Memory*, *4*(5), 411–420. https://doi.org/10.1101/lm.4.5.411
23. Nagel, I. E., Preuschhof, C., Li, S.-C., Nyberg, L., Bäckman, L., Lindenberger, U., & Heekeren, H. R. (2009). Performance level modulates adult age differences in brain activation during spatial working memory. *Proceedings of the National Academy of Sciences*, *106*(52), 22552–22557. https://doi.org/10.1073/pnas.0908238106

24. Paperno, N., Rupp, M. A., Parkhurst, E. L., Maboudou-Tchao, E. M., Smither, J. A., Bricout, J., & Behal, A. (2019). Age and Gender Differences in Performance for Operating a Robotic Manipulator. *IEEE Transactions on Human-Machine Systems*, *49*(2), 137–149. https://doi.org/10.1109/THMS.2019.2890855
25. Piefke, M., Onur, Ö. A., & Fink, G. R. (2012). Aging-related changes of neural mechanisms underlying visual-spatial working memory. *Neurobiology of Aging*, *33*(7), 1284–1297. https://doi.org/10.1016/j.neurobiolaging.2010.10.014
26. Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Current Directions in Psychological Science*, *17*, 177–182. https://doi.org/10.1111/j.1467-8721.2008.00570.x
27. Riby, L., Perfect, T., & Stollery, B. (2004). The effects of age and task domain on dual task performance: A meta-analysis. *European Journal of Cognitive Psychology*, *16*(6), 863–891. https://doi.org/10.1080/09541440340000402
28. Roller, C. A., Cohen, H. S., Kimball, K. T., & Bloomberg, J. J. (2002). Effects of normal aging on visuo-motor plasticity. *Neurobiology of Aging*, *23*(1), 117–123. https://doi.org/10.1016/S0197-4580(01)00264-0
29. Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023
30. Schieber, F., & Gilland, J. (2008). Visual Entropy Metric Reveals Differences in Drivers' Eye Gaze Complexity across Variations in Age and Subsidiary Task Load. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(23), 1883–1887. https://doi.org/10.1177/154193120805202311
31. Seidler, R. D., Bernard, J. A., Burutolu, T. B., Fling, B. W., Gordon, M. T., Gwin, J. T., Kwak, Y., & Lipps, D. B. (2010). Motor control and aging: Links to age-related brain structural, functional, and biochemical effects. *Neuroscience & Biobehavioral Reviews*, *34*(5), 721–733. https://doi.org/10.1016/j.neubiorev.2009.10.005
32. Seidler, R. D., Bo, J., & Anguera, J. A. (2012). Neurocognitive Contributions to Motor Skill Learning: The Role of Working Memory. *Journal of Motor Behavior*, *44*(6), 445–453. https://doi.org/10.1080/00222895.2012.672348
33. Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews*, *96*, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007
34. Stone, J., & Towse, J. (2015). A Working Memory Test Battery: Java-Based Collection of Seven Working Memory Tasks. *Journal of Open Research Software*, *3*(1), Article 1. https://doi.org/10.5334/jors.br
35. Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, *41*(2), 167–174. https://doi.org/10.1111/j.1469-8986.2003.00148.x
36. Vermeij, A., van Beek, A. H. E. A., Reijs, B. L. R., Claassen, J. A. H. R., & Kessels, R. P. C. (2014). An exploratory study of the effects of spatial working-memory load on prefrontal activation in low- and high-performing elderly. *Frontiers in Aging Neuroscience*, *6*, 303. https://doi.org/10.3389/fnagi.2014.00303
37. *VIVE Pro Eye Overview | VIVE United States*. (2022). https://www.vive.com/us/product/vive-pro-eye/overview/
38. von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*(1), 108–115. https://doi.org/10.3758/s13428-012-0224-y
39. Zénon, A., Sidibé, M., & Olivier, E. (2014). Pupil size variations correlate with physical effort perception. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00286

Chapter 5. Conclusion

The primary goal of this work was to investigate the ability of eye-tracking measures to characterize workload and visuomotor strategies in physical human-robot interaction (pHRI). This work provides evidence for eye tracking measures to elucidate workload and visuomotor strategies broadly, by encompassing interaction with an industrial collaborative robot as well as an intelligent exoskeleton, considering task difficulty in terms of robot control and environmental constraints, explicitly considering learning and adaptations in each setting, and including novices, experts, and age-related differences. We focused on co-manipulation tasks using a cobot and an exoskeleton, since co-manipulation is a novel application of robotics, wherein little is known about how continuous force-interactions between a human and robot, and the need to control the robot and any potential lack of synergy between human and robot, may all impact mental workload and visual attention. Furthermore, the gradual adaptations in performance in each setting, and the resulting changes in mental workload were also studied. Specifically, we investigated 1) The sensitivity and reliability of eye-tracking measures, and their ability to predict performance in an upper-limb pHRI task under varying workload 2) The motor-coordination demands associated with using a whole-body powered exoskeleton for level walking, and potential differences between expert- and novice exoskeleton-users using eye-tracking measures 3) The differences in workload and gaze behavior between younger and older adults while performing a pHRI co-manipulation task. Overall, our results suggest that it is possible and feasible to measure the unique motor-cognitive demands associated with using cobots via eye-tracking metrics. Multiple eye-tracking metrics can distinguish between different types of mental demands and also change over the course of learning to use a cobot, although they generally appear to be less sensitive to motor-control demands compared to the demands of more cognitive work. Moreover, eye-tracking measures also differentiate between novice- and expert-users and highlight age differences in mental workload and visuomotor strategies.

5.1 Summary of major results

In chapter 2, we investigated the sensitivity and reliability of eye-tracking metrics to the difficulty associated with controlling a cobot, as well as their ability to predict task performance. We found that pupil dilation (PD) and stationary gaze entropy (SGE) were most sensitive to, and reliably associated with robot-difficulty, and that PD was more sensitive and reliable at detecting changes over learning, compared to robot-difficulty in our study context. AOI metrics were more sensitive to robot-difficulty than any of the eye-tracking workload metrics, suggesting that robot-difficulty had a strong effect on visual attention and strategies. Although AOI measures may be highly task-specific, and they are not direct measures of workload, they can provide valuable information regarding instantaneous changes in gaze behavior and strategies based on task-demands. Eye-tracking measures were also able to predict performance with greater accuracy compared to the study independent variables, thus supporting their ability to not only indicate workload, but also to predict successes and failures on a per-trial basis. Taken together, these results suggest that a combination of eye-tracking measures can provide multidimensional information regarding workload, strategies, and performance in a pHRI task under different robot-difficulty conditions.

In chapter 3, we investigated the differences in workload and visual attention while walking in a whole-body powered exoskeleton compared to walking freely, as well as the potential differences between expert- and novice exoskeleton-users. Lastly, we also studied the changes in gaze behavior of novices over multiple practice sessions in the exoskeleton. In addition to PD and SGE as measures of workload and visual attention, we also computed the ratio of path-focused fixations (PF; as opposed to fixations directed upwards in the field of view), as a measure specific to the demands associated with maintaining stability while walking. We found that experts and novices showed similar PD, SGE and PF in the free-walking condition. While walking in the exoskeleton, however, novices' PD increased, indicating higher mental demand for novices compared to experts. SGE reduced in the exoskeleton for both novices and experts, suggesting suppression of eye-movements in order to maintain balance, and novices' SGE appeared to increase over sessions. Lastly, PF increased for both experts and novices while using the exoskeleton, although experts' PF increased to a greater extent, indicating that greater PF may be a visual strategy that is acquired over long-term learning. This work provided initial evidence that mental workload increased and gaze behavior was altered while using an exoskeleton, and that eye-tracking measures may be useful indicators of expertise. Future research should aim to measure the actual impact of compensatory gaze strategies such as downward-focused gaze on situation awareness and detection of environmental hazards.

Finally, in chapter 4, we investigated the differences in workload between younger and older adults while they performed a co-manipulation task with a cobot under different levels of robot- and environmental-difficulty. Initial results aligned with those from study 1, in terms of the sensitivity of eye-tracking measures to workload and learning. More importantly, however, we found age differences in workload that were not captured by perceived workload metrics and performance – in initial trials, PD reduced significantly for younger compared to older adults, indicating that older adults may have sustained cognitive effort for a longer duration. Gaze transition entropy (GTE) reduced with task difficulty for both groups, although the reduction was significantly more for older adults, possibly due to a reduced ability to modulate attention in older adults. Lastly, different types of workload manipulations affected eye-tracking measures differently. Only robot-difficulty had a significant effect on PD, and environmental complexity was more likely to have a significant effect on gaze-behavior measures. The effect of environmental complexity on gaze behavior metrics was more consistent over trials, whereas PD changed significantly over the course of learning. These results indicate that pupillary- and gaze behavior metrics were diagnostic of different types of task demands, and that a combination of these metrics may be more beneficial to workload assessment in real-world situations where sources of demand may be highly variable.

5.2 Overall Limitations

There are some general limitations to this work. Firstly, although our use of virtual reality (VR) for simulating the task environment in the first and last studies afforded multiple benefits, such as maintaining a continuous rate of presentation of stimuli, reducing the potential for physical workload to confound the effects on mental workload on the dependent measures, allowing for variation in task demands, and more convenient measurement of AOI-based metrics, we did not provide participants with the full sensory feedback (e.g., tactile) that is associated with manipulating objects. The lack of tactile and force feedback from a physical object may have influenced gaze behavior, and the ability of eye-tracking measures to generalize to real-world pHRI tasks is yet to be confirmed. However, several measures demonstrated expected changes in study 2 which involved a physical task in the real-world, providing some indication of the generalizability of the eye-tracking metrics used in this work. Secondly, all our studies investigated learning over the short term, with only the second study considering longer-term skill acquisition. How eye-tracking workload measures may vary over multiple days or weeks of learning to use a cobot is a critical issue for future research, which will provide further understanding regarding whether eye-tracking measures are viable and reliable indicators of cognitive state during steady-state task performance. Thirdly, the tasks designed for studies 1 and 3 precluded us from computing anticipatory gaze measures such as eye-hand span that have also been shown to be strongly related with motor-skill learning in upper-limb goal-directed tasks. Since our tasks did not require participants to anticipate locations of objects or targets based on memory, i.e., the manipulated object appeared in the same location only after the previous step was completed, and the target sequence was randomized and unpredictable, measures of anticipatory gaze behavior such as eye-hand span were not applicable in this context. An important question for future research is how cobot-difficulty and motor-skill learning may influence the anticipatory gaze movements, and how these may change over the course of learning. Lastly, in our final study, potential age differences may have been attenuated, due to our age groups being relatively closer in age compared to other studies employing extreme-groups designs, and the older adult group being imbalanced in terms of gender. Future research should consider the ability of eye-tracking measures to characterize workload across a more diverse set of tasks and individuals.

5.3 Overall Applications

Generally, our findings indicate that eye-tracking measures can index different components of mental workload in pHRI – the physiological response to workload (via PD), as well as visuomotor strategies (FR, SGE, AOI-metrics) in response to task demands. Moreover, eye-tracking metrics, especially PD, were also sensitive to learning. Future adaptive control algorithms that may seek to employ these metrics to understand cognitive state, may use PD to infer gradual learning or changes in workload due to robot-difficulty, and FR or robot-fixations may be used to infer user intent (e.g. an effort to avoid collisions). On the other hand, real-time cognitive state inference based on eye-tracking metrics may be somewhat more challenging during real-world tasks, due to the need to identify and classify AOI in the real-world. Eye-tracking data analysis for our work on the whole-body exoskeleton involved a considerable amount of manual post-processing; however, techniques based on computer vision to detect real-world objects or kinematic sensors to identify gaze direction based on head- and body-orientation may help circumvent these requirements.

We found that the sensitivity of eye-tracking metrics to cognitive demands was somewhat dependent on the stage of practice, and these differences varied across age. For example, in study 3, we found that the main effect of task difficulty on PD was significant only in the last three trials, i.e. PD was initially similar in both the easy and difficult conditions but reduced at different rates, such that it was significantly different across task difficulties by the end of the session. We anticipate that the sensitivity of PD to workload will vary at different points in time over skill acquisition, and the rates of these changes may vary across age groups and other cognitive and physical individual differences. Long term adaptations may again reduce the sensitivity of PD, as workload in difficult conditions reduces to the level of workload in the easier conditions. Thus, it is possible that PD is most sensitive to workload during intermediate stages of learning. It is important that future adaptive algorithms consider such fluctuations in sensitivity over time. The ability of eye-tracking measures to predict performance with better accuracy than task variables (Study 1, Chapter 2) is an encouraging result for robot-adaptation based on these measures. Since eye-tracking measures can represent the intent to perform an action, they can provide advance information regarding upcoming successes or failures that can inform robot-adaptation strategies.

An ongoing research effort pertains to determining task-cobot fit, and knowledge regarding the array of tasks for which cobots may be used for continues to evolve. Since eye-tracking can only convey information about foveal, and not peripheral vision, it is possible that they may only be effective as workload measures in goal-directed pHRI tasks that demand high precision. Some participants from studies 1 and 3 stated that they monitored the target locations using their "ambient vision". Such use of ambient/peripheral vision may reduce dependence on active monitoring, and reduce the utility of gaze behavior measures in particular. Future work aiming to study eye-tracking measures in pHRI may consider designing experimental tasks based on precision-tasks such as welding or drilling, which are likely to demand greater monitoring and more focused fixation patterns.

Certain precautions may need to be taken by future researchers who wish to study gaze patterns in VR, and are particularly interested in classification of eye-movements such as fixations. We discovered that the VR system used in this study (HTC Vive Pro Eye) had inbuilt software features that processed the gaze signal in a way that caused head-rotation to influence eye-gaze position. This would have introduced noise into the eye-tracking position signal, had our study required large head movements. Unfortunately, disabling this feature required the purchase of a paid software package. Thus, to reduce the need for large head movements, the virtual robot was positioned farther away from the operator in VR, compared to its distance from the operator in the real world. Researchers may need to be vigilant of such instances in which VR manufacturers may pre-process eye-tracking data in a way that may introduce bias or noise.

The work described in chapter 3 is especially novel for a few reasons. Firstly, very little work has quantified the impact of stability demands on gaze behavior in healthy populations during free-walking, i.e. walking without needing to visually identify and step on specific footholds. The use of a whole-body powered exoskeleton provided a unique opportunity and an applied context within which to study the impact of coordination and stability demands on gaze behavior. The observed reduction in SGE and increase in path-focused fixations in both novices and experts indicated that using a powered exoskeleton can continue to produce compensatory gaze behaviors even after long-term use. In fact, downward gaze could be viewed as a maladaptation to the demands of using an exoskeleton. This has significant implications for operator training and long-term assessment of visual attention and industrial safety.

The questions raised in chapter 3, regarding whether downward-focused gaze may negatively impact attention and situation awareness, also apply to the studies in chapters 2 and 4 involving object co-manipulation. In those studies, it was found that some participants would accidentally transfer the plate to an incorrect target location, and in most cases, that location was one that was previously highlighted but resulted in a failed transfer. This observation suggests that our task demands may have caused some inattention or change blindness due to mental or visual distraction. Participants may have directed their visual attention towards the incorrect target, not realizing that another target was actually highlighted. Thus, sub-optimal gaze patterns may detract from the ability to detect hazards or perform other secondary tasks. This effect may be exacerbated in older groups due to a reduced ability to control attention. An opportunity for future research is to investigate how the mental demands of using a cobot and specific gaze behaviors (e.g. downward gaze, or reduced GTE) may lead to inattention and loss of safety in complex, multitasking environments.

In summary, the findings from this work provide evidence for the potential of eye-tracking metrics as continuous, near-real-time indices of workload and learning in pHRI. Additionally, this work provides a methodological foundation for future pHRI studies to build upon in terms of study and task-design, potential use of novel technologies such as virtual reality (VR) to investigate motor performance, and future research directions regarding mental workload and human performance in pHRI. In the long term, findings from this research can guide the design of control algorithms that use eye-tracking measures as inputs in order to enable cobots to adapt to the operator's real-time cognitive state.