Multimodal Evaluation of Cognitive Load During Prosthetic Device Use in Activities of Daily Living

Lorena Velásquez[1*], Gabriella Goytizolo[1], Hasan Ayaz[2], Jeremy D. Brown[1,3]

[1*]Department of Mechanical Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, 21218, MD, USA.
[2]School of Biomedical Engineering, Science and Health Systems, Drexel University, 3141 Chestnut Street, Philadelphia, 19104, PA, USA.
[3]Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, 725 N. Wolfe Street, Baltimore, 21205, MD, USA.

*Corresponding author(s). E-mail(s): lvelasq3@jhu.edu;
Contributing authors: ggoytiz1@jhu.edu; ayaz@drexel.edu; jdelainebrown@jhu.edu;

Abstract

Purpose: The cognitive burden associated with upper-limb prosthesis use remains a barrier to functional adoption, yet existing evaluations often rely on single-modality measures. This study investigates a multimodal framework for assessing cognitive load during activities of daily living. **Methods:** Twelve participants without limb difference performed standardized tasks using their natural hand and a tendon-driven prosthesis in a within-subject design. Cognitive workload was assessed using task performance metrics, subjective ratings from the NASA Task Load Index (NASA-TLX), eye-tracking measures including gaze behavior and pupillary dynamics, and prefrontal cortical activation measured with functional near-infrared spectroscopy (fNIRS). **Results:** Prosthesis use was associated with increased cognitive demand across modalities, although modality-specific responses differed by task. Subjective workload ratings and prefrontal cortical activation consistently indicated greater workload during prosthesis use, with neural activation becoming more widespread as task complexity increased. In contrast, pupillary activity differentiated conditions during the simpler block transfer and checker transfer tasks but not during the more visually dynamic carton transfer and pour task. Performance measures alone did not fully capture these differences, highlighting complementary information provided by physiological and visual attention measures. **Conclusion:** A multimodal framework provides a more comprehensive assessment of cognitive load than any single measure alone, offering valuable insights into how cognitive demand manifests across behavioral, perceptual, and biological domains . This approach enables improved evaluation of prosthetic systems and may inform the design of devices that reduce user burden and enhance functional performance.

Keywords: Upper-limb prosthesis, Cognitive load, Multimodal assessment, Activities of daily living

Introduction

Of the estimated 1 million Americans living with upper-limb difference (17% of the 5.6 million individuals with limb loss in the United States), only 27–50% choose to use prosthetic devices to assist with activities of daily living (ADL) [1, 2]. These devices lack individualized design and tunable control strategies that can adapt to a user's unique physiology and preferences in real time to capture the form, function, and effort of the residual limb. This lack of control, exacerbated by current challenges in sensory feedback, makes prosthetic devices physically, and more importantly, cognitively demanding to use, resulting in consistently high abandonment rates [3, 4]. Understanding how to capture cognitive burden in prosthetic device use provides valuable insights that can directly inform their development, ensuring that the trajectory and implementation align with the user's needs [5, 6].

Early evaluations of cognitive load sought to capture the mental effort associated with daily tasks through self-report instruments. The National Aeronautics and Space Administration-Task Load Index (NASA-TLX), originally developed to assess workload in aviation, evaluates mental, physical, and temporal demand, as well as effort, performance, and frustration [7]. More recently, it has been adapted to assess cognitive burden during prosthetic use through the Prosthesis Task Load Index (PROS-TLX), which accounts for the unique challenges of prosthetic device operation [8]. Although TLX-based assessments provide rich insight into the user's subjective experience of workload and have demonstrated sensitivity to differences in task complexity and device design, self-report measures cannot continuously track real-time changes in cognitive load and may be influenced by recall bias or individual interpretation [5, 9].

More comprehensive task evaluation strategies have been implemented to evaluate cognitive load during task execution, including time to completion, error rate, and dual-task interference. These metrics serve as proxies for cognitive effort, based on the assumption that tasks requiring greater mental resources are performed more slowly, with more errors, or with degraded performance on a secondary task. While useful, such measures cannot disentangle whether performance limitations arise from mechanical restrictions of the device, user behavioral strategies, or underlying cognitive demand [5, 10, 11]. The combination of self-report and performance-based cognitive evaluations has shown merit for assessing cognitive load in upper-limb prostheses; however, both self-report and task evaluation methods are discrete evaluations that occur after the task has been performed, which may not capture the continuous, time-varying nature of mental effort. These external measures of cognitive load also do not capture underlying neurophysiological behavior [12].

Other approaches to evaluating cognitive load rely on continuous physiological measurements acquired during task execution, including heart rate, eye tracking, electrodermal activity, and respiratory rate [5]. Eye-tracking has emerged as an important objective tool for assessing attentional allocation and workload-related visual behavior associated with cognitive load during prosthesis use. Measures such as fixation duration and gaze allocation provide insight into how users distribute their attention between the device, task objects, and the surrounding environment [13, 14]. Increased fixation times, gaze switching, and pupil size have been associated with greater attention demand, reflecting increased mental workload have been associated with greater attentional demand and increased workload-related processing [15]. In prosthesis studies, eye-tracking has revealed characteristic differences between natural hand users and prosthesis users, including a tendency for the latter to fixate more heavily on their hands and to exhibit longer gaze delays when coordinating object manipulation [16, 17]. Though these metrics provide continuous, real-time assessment of attentional strategies, they can be highly sensitive to motion artifacts, environmental factors such as illumination, and the underlying or inherent physical characteristics of participants.

In contrast, brain imaging techniques are less susceptible to environmental fluctuations and can provide a more robust method for measuring physiological signals associated with cognitive load. With higher spatial resolution than electroencephalogram (EEG) and greater temporal resolution than Functional Magnetic Resonance Imaging (FMRI), Functional near-infrared spectroscopy (fNIRS) provides a complementary window into cognitive workload by measuring task-evoked changes in cortical oxygenation [18–20]. By monitoring hemodynamic activity in the prefrontal cortex, fNIRS has been shown to reliably index changes in prefrontal engagement associated with cognitive workload across a variety of motor and control tasks [21]. Unlike self-report post-task measures, fNIRS can capture dynamic fluctuations in mental effort, and unlike eye tracking, it reflects internal neural activation rather than behavioral manifestations of attention. In prosthetics research, fNIRS has demonstrated sensitivity to differences in device control strategies and task complexity, highlighting the elevated mental demand associated with operating advanced myoelectric prostheses [12, 22, 23]. The non-invasive nature of fNIRS and its compatibility with real-world tasks make it a promising neuroergonomic tool for evaluating the cognitive costs of prosthesis use. However, while informative, fNIRS alone does not provide insight into the behavioral strategies or subjective perceptions that shape user experience.

Despite the clear value of each of these approaches, studies of upper-limb prosthesis use have largely treated cognitive load assessment as a single-modality problem. Performance-based metrics provide coarse indicators of task difficulty but cannot disentangle cognitive contributions from mechanical or motor limitations. Self-report instruments such as NASA-TLX and PROS-TLX capture user perception but lack temporal resolution and are subject to recall bias. Eye-tracking reveals attentional strategies but cannot directly quantify neural effort. At the same time, fNIRS provides neural correlates of workload but does not account for how users allocate attention or perceive their experience. By contrast, multimodal approaches that integrate subjective, behavioral, and neural measures have already shown merit in other domains, including driving, aviation, and collaborative learning, where they explain more variance in workload and task performance than any single measure alone [24–27]. However, there has been limited adoption of multi-modal cognitive load approaches in prosthetics research, particularly for the upper-extremity. A multimodal approach would provide a more holistic evaluation of the cognitive demands imposed by prosthesis use, yielding insights that cannot be obtained from any individual measure in isolation. While prior studies have evaluated cognitive workload during prosthesis use using performance-based metrics, subjective workload assessments, fNIRS, or eye-tracking measures individually, few studies have examined how these complementary measures converge and diverge across activities of daily living with differing motor and visual demands. The present study addresses this gap by directly comparing behavioral, subjective, pupillary, gaze-based, and neural indices of workload within the same participants and experimental task framework.

In this manuscript, we employ a multimodal cognitive load assessment framework to assess nuanced differences in cognitive demand during prosthetic device use. We asked participants to perform three activities of daily living (ADL) tasks of varying complexity with their natural hand while cognitive workload was assessed using three complementary modalities: subjective perception via the NASA-TLX questionnaire, pupillary dilations via eye-tracking glasses, and prefrontal cortex activation via fNIRS imaging. We then asked the same participants to repeat the experiment using a research prosthesis under identical conditions, enabling a direct comparison between natural hand and prosthesis use. We hypothesize that no single measurement modality can robustly capture all aspects of cognitive load during upper-limb prosthesis use. Rather, we posit that cognitive demand is inherently multidimensional and is best characterized through a multimodal evaluative framework that integrates task performance, subjective workload assessment, neural measures, and behavioral indicators of visual attention. Within this framework, each modality is expected to capture distinct but complementary dimensions of cognitive load, including perceived effort, executive control, and attentional allocation.

Specifically, we anticipate that prosthesis use will be associated with reduced task performance relative to the natural hand across activities of daily living, providing important context for interpreting cognitive workload. We further expect subjective workload ratings obtained through the NASA-TLX to differentiate between natural hand and prosthesis use and to scale with task demands, reflecting perceived mental, physical, and temporal effort. At the neural level, we hypothesize that prefrontal cortex activation will be greater during prosthesis use and will exhibit task-dependent patterns consistent with increasing executive and monitoring demands. At the behavioral level, we expect pupillary activity derived from eye-tracking to reflect task-dependent attentional demands during prosthesis use, with sensitivity to tasks requiring active visual monitoring. Finally, we hypothesize that task complexity will modulate both the magnitude and distribution of cognitive load across performance, subjective, neural, and behavioral measures, yielding complementary and modality-specific signatures of cognitive demand.

In what follows, we describe the experimental protocol and detail the multimodal data collection framework, including subjective workload assessment, eye-tracking of visual attention through pupillary activity, and fNIRS-based quantification of prefrontal cortex activation. We then present the results of these analyses, organized to assess both task-complexity effects and differences between the natural hand and prosthesis use. Beyond demonstrating the value of a multimodal approach, this study provides new insight into how different cognitive load measures converge and diverge across task contexts. By first evaluating task performance, subjective workload, neural activation, and behavioral measures in a cohort of participants without limb loss, this work establishes a baseline for understanding the sensitivity and task dependence of these complementary metrics. This controlled framework enables systematic comparison across modalities and tasks, providing a necessary reference point for future investigations involving individuals with limb loss, where factors such as embodiment, training history, and long-term adaptation introduce additional sources of variability. This study intentionally employed participants without limb loss using a simulated prosthesis to establish a controlled baseline for evaluating the sensitivity and task dependence of multimodal workload measures. Accordingly, the findings should be interpreted primarily as evidence regarding modality-specific responses to prosthesis-mediated task demands, rather than as direct estimates of cognitive workload in experienced prosthesis users. Importantly, we recognize that comprehensive multimodal assessment may not be feasible in all experimental or clinical settings, and therefore, our findings enable informed selection of cognitive load metrics by highlighting which modalities are most sensitive to specific task demands. In this way, the present work offers both a framework for multimodal cognitive load evaluation and practical recommendations for assessing cognitive demand when only a subset of measurement modalities are available.

Methods

Participants

N=12 participants without limb loss (8 female, 4 male; age 22 ± 1.5 years, 10 right-handed) were recruited from the adult population of Johns Hopkins University and Hospital System in accordance with a study protocol approved by the Johns Hopkins Medicine IRB (#00147458). All participants provided informed consent before participation in the study. Experimental procedures were approved by the Johns Hopkins Medicine Institutional Review Board (IRB #00147458), and carried out in accordance with the Declaration of Helsinki and with the relevant guidelines and regulations set forth by the Johns Hopkins Medicine Institutional Review Board. Participants were compensated at a rate of $15 per hour, and the total experimental session lasted approximately 150 minutes. Participants had no prior exposure to prosthetic devices. Participants had normal or corrected-to-normal vision and the ability to wear contact lenses on the day of the experiment. Each participant provided informed consent before participating in the study.

Experimental Procedure

In the first phase of the experiment, participants used their right hand to complete nine one-minute trials of three ADL-inspired tasks: block transfer, checker transfer and stack, and carton transfer and pour. During task performance, participants wore Pupil Core eye-tracking glasses and a 16-optode fNIRS device. Between trials, participants completed the NASA-TLX survey to capture subjective workload. The second phase proceeded under identical conditions, with participants employing the Tendon Actuated Modular Prosthesis (TAMP), a trans-radial research prosthesis, instead of their natural hand. Within each condition, trial order was randomized and counterbalanced across participants to mitigate ordering effects. The experiment was conducted in two 60-minute sessions separated by a minimum of 48 hours to minimize learning effects.

Experimental Hardware

Devices used in the experiment include the Pupil Core Eye-Tracking system, a 16-optode Model 2000s fNIRS device, and the Tendon Actuated Modular Prosthesis (TAMP). All input and output data streams were controlled through a Quanser Q8 USB DAQ sampling at a 1 kHz sample rate, using QUARC real-time software in MATLAB/Simulink 2021a (Mathworks).

Pupil Core Eye-tracking Glasses

The Pupil Core eye-tracking system (Pupil Labs GmbH, Berlin, Germany) recorded binocular gaze data and pupil diameter during task performance. The system samples eye images at 200 Hz and includes a forward-facing scene camera to capture the participant's visual field. Prior to each trial, a standard calibration procedure was performed to ensure accurate gaze estimation.

Functional Near-Infrared Spectroscopy (fNIRS)

Cortical activity was measured using a wearable 16-optode Model 2000s fNIRS system (fNIR Devices, LLC) positioned over the participant's prefrontal cortex as described by Ayaz et. al, in [28]. The optode layout was selected to target bilateral dorsolateral and medial prefrontal regions associated with executive function and cognitive workload. The system records changes in oxygenated (HbO) and deoxygenated (HbR) hemoglobin concentrations during task performance.

Tendon Actuated Modular Prosthesis (TAMP)

TAMP is a 3D-printed modular prosthesis. The custom socket has a simulated prosthesis structure designed to be worn by individuals without limb loss on the right arm. As seen in Fig. 1, the device consists of two interlocking pieces: 1) a shield that obscures the wearer's hand and integrates the custom-designed anthropomorphic hand, and 2) a socket featuring an open lattice structure with integrated mountings for two rotary DC motors (Maxon RE30) positioned at the proximal end of the device. The prosthesis is actuated using two primary tendons that spool around the motor shaft and extend

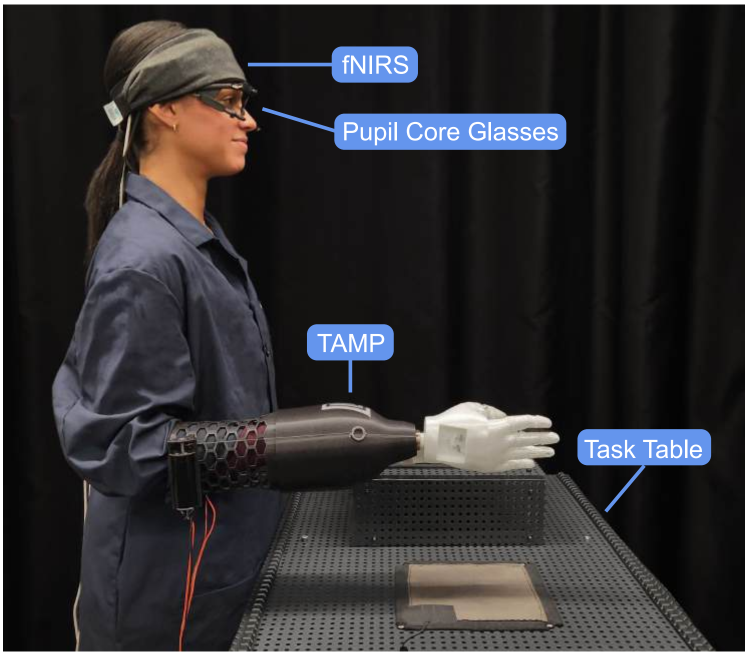

Fig. 1 Experimental setup used for simultaneous multimodal data collection. Participants wore an fNIRS headband and Pupil Core eye-tracking glasses while performing tabletop tasks with the Tendon-Actuated Modular Prosthesis (TAMP) at a standardized task table.

along the device, terminating at the base of the Whipple-tree. Secondary end effector-specific tendons route through either side of each finger, initiating at the fingertip and terminating at the first stage of the whipple tree. The control scheme employs a Hill-Type Muscle model-based First-In-First-Out Finite State Machine (FIFO FSM) to map sEMG signals to tendon actuation [29]. Each participant undergoes an individualized calibration process to define flexor and extensor activation thresholds, ensuring robust signal differentiation. More details of the device design and control can be found in [30]. Together, these systems enabled simultaneous acquisition of multimodal data during task performance.

Cognitive Load Measures

Cognitive load was assessed using three complementary modalities capturing subjective perception (NASA-TLX), neural activity (fNIRS), and behavioral indicators of visual attention (eye-tracking), providing a multidimensional evaluation of workload-related behavioral, perceptual, and neural responses during task performance.

NASA Task Load Index (NASA-TLX)

Subjective workload was assessed using the NASA Task Load Index (NASA-TLX), a validated instrument that evaluates perceived workload across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. The standard NASA-TLX was selected over the Prosthesis-Specific TLX (PROS-TLX), as the latter is designed for device-specific evaluation and is not applicable to the natural hand conditions included in this study. Participants completed the NASA-TLX questionnaire after each trial. Weighted TLX scores were computed to quantify overall cognitive workload, enabling comparison across tasks and conditions.

Functional Near-Infared Spectroscopy (fNIRS)

Cortical activation was measured using a wearable functional near-infrared spectroscopy (fNIRS) Model 2000s system (fNIR Devices, LLC) configured with 16 optodes positioned over the forehead to image the prefrontal cortex. The optode layout was selected to target bilateral, dorsolateral, and medial prefrontal regions commonly associated with executive function and cognitive workload. The fNIRS system recorded changes in oxygenated (HbO) and deoxygenated (HbR) hemoglobin concentrations during task performance, providing a non-invasive measure of task-evoked neural activity. Comprehensive procedural details, including optode placement, mechanical stabilization, and signal quality verification, are described in section 1 of the supplementary document. fNIRS data was processed using FNIRSoft, following established pre-processing procedures [31]. Specifically, signals were corrected for motion artifacts and low-frequency drift and low-pass filtered to remove high-frequency physiological noise. Hemodynamic responses were segmented by trial and task condition, and mean HbO changes within each prefrontal region of interest were used as neural indicators of cognitive demand.

Eye-Tracking and Pupillary Activity

Eye-tracking data was collected using Pupil Core eye-tracking glasses (Pupil Labs GmbH, Berlin, Germany), which record binocular gaze data and pupil diameter during task performance. The system samples eye images at 200 Hz and includes a forward-facing scene camera to capture the participant's visual field. Before each trial, a standard calibration procedure was performed for each participant to ensure accurate gaze estimation. Participants with corrected-to-normal vision wore contact lenses during data collection to ensure unobstructed pupil tracking. Eye-tracking glasses were worn continuously throughout each task trial for both natural hand and prosthesis conditions and were adjusted to ensure comfort for each participant. Raw eye-tracking data was exported and post-processed to extract pupil diameter and gaze metrics. Samples with confidence values below 0.6 were excluded to mitigate the effects of blinks, occlusions, and tracking loss. Cognitive load was quantified using the Low/High Index of Pupillary Activity (LHIPA), which captures fluctuations in pupil diameter associated with mental effort [32].

Gaze Behavior

In addition to pupillary measures, gaze behavior was analyzed to characterize visual attention strategies, including the distribution of gaze across task-relevant regions providing a behavioral measure of visual attention to complement pupillary and neural indices of cognitive load. Gaze data were segmented into task-relevant regions of interest, including the end effector (or natural hand), task objects, and the surrounding workspace. The proportion of time spent fixating within each region was computed for each trial to quantify the distribution of visual attention. These gaze metrics were used to provide contextual interpretation of pupillary responses, particularly in tasks requiring dynamic visual monitoring. Detailed descriptions of gaze segmentation and analysis procedures are provided in section 2 of the supplementary document.

Experimental Tasks

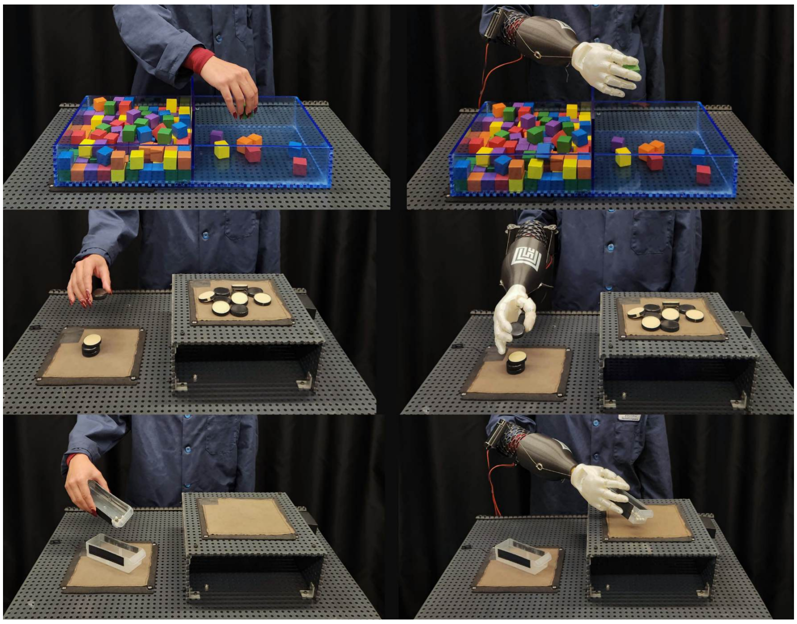

Fig. 2 A participant performing the block transfer, checker transfer and stack, and carton transfer and pour tasks using either the natural hand or prosthesis.

Each experimental task was performed for three one-minute trials by each participant.

Box and Blocks

The Box and Block Test is commonly used to assess unilateral manual dexterity in rehabilitative settings [33, 34]. Although the test is traditionally validated in a seated position, pilot studies indicated that participants using the prosthesis experienced greater comfort when standing. During the task, participants stood at a table facing a rectangular box divided into two equal square compartments by a central partition. 150 colored wooden blocks, each measuring 2.5 cm^3, were placed in one compartment (see Fig. 2). Participants were instructed to transfer as many blocks as possible, one at a time, from the right to the left compartment within the one-minute time frame. This task required the use of a pinch-style grasp.

Checker Transfer and Stack

The checker transfer and stack task draws inspiration from the thin-profile transfer tasks of the Southampton Hand Assessment Procedure and the Jebsen-Taylor Hand Function Test checker-stacking task [35–37]. This task integrates the precise pinch or lateral grip required to lift thin objects with the fine motor control necessary for transferring and stacking checkers. Participants interacted with 30 checkers (38 mm in diameter and 10 mm thick) placed on the elevated target surface of a tiered transfer table (see Fig. 2). Participants were instructed to transfer and stack as many checkers as possible onto the lower surface of the table within the time limit.

Carton Transfer and Pour

The carton transfer and pour task requires participants to use a power grasp to transfer a carton filled with weighted marbles (50 g, 75 g, 100 g, 5 mm diameter) from one side of a tiered transfer table and pour them into an identical empty carton on the opposite side. Participants are instructed to repeat this transfer and pouring process as many times as possible within the time limit. This task introduces several layers of complexity beyond a standard grasp-and-transfer exercise. It involves lifting objects of varying mass between two different heights and pouring the contents from one container to another. The objects used are rectangular cartons, designed to mimic real-life cartons, such as those featured in the Southampton Hand Assessment Procedure [37]. The dynamic components of the task challenged participants to adapt to changes in weight distribution while maintaining pouring accuracy. By incorporating varying object masses and dynamic weight shifts, the task assesses both transfer accuracy and the ability to handle variable loads under realistic conditions.

NASA Task Load Index (NASA-TLX)

The NASA Task Load Index (TLX) is a validated tool for assessing subjective workload [7]. It considers mental, physical, and temporal demands, along with perceived performance, effort, and frustration [12]. In this study, the NASA-TLX compares workload scores and perceived performance across tasks between the natural hand and TAMP.

Metrics

The following metrics were used to analyze cognitive loading and task performance across trials.

of Blocks Transferred

The total quantity of blocks that a participant moves from one side of the box to the other during a trial of the box and blocks task.

of Checkers Transferred

The total quantity of checkers that a participant successfully moves from one side of the transfer table to the other during a trial of the checker transfer and stack task.

of Checkers Stacked

The total number of checkers that a participant successfully stacks on top of one another during a trial of the checker transfer and stack task.

of Carton Transfers

The total count of successful transfers completed during a trial of the carton transfer and pour task. A successful transfer is defined as the uninterrupted movement of a carton from one side of the tiered table to the other.

Carton Pour Accuracy

The proportion of marbles successfully transferred during each pour relative to the total marble count, producing an average accuracy score for each trial of the carton transfer and pour task.

NASA-TLX

Weighted TLX scores determined increases in cognitive loading across tasks as described in [12].

Prefrontal Cortex Activation

Functional near-infrared spectroscopy (fNIRS) was used to measure task-evoked changes in prefrontal cortical hemodynamics associated with workload-related neural activity. fNIRS quantifies relative changes in oxygenated (HbO) and deoxygenated (HbR) hemoglobin concentration through near-infrared light absorption and neurovascular coupling mechanisms. The fNIR Devices Model 2000 system (fNIR Devices, LLC) was configured with 16 optodes positioned over the prefrontal cortex to monitor hemodynamic activity during task performance. Raw optical intensity signals acquired at 730 and 850 nm were processed in FNIRSoft using standard preprocessing procedures, including low-pass finite impulse response (FIR) filtering (Hamming window, order 100, and cutoff frequency 0.1 Hz), sliding window motion artifact rejection (SMAR) algorithm for signal quality inspection and exclusion of unstable channels prior to hemodynamic conversion via the modified Beer–Lambert Law [31]. Hemodynamic responses were segmented by trial and condition, and cortical activation was quantified as the mean change in oxygenated hemoglobin (ΔHbO) averaged within each prefrontal region of interest (ROI). Due to susceptibility of inferior optodes to infrared contamination from the eye-tracking system, only superior-row optodes corresponding to bilateral dorsolateral and medial prefrontal regions were included in statistical analyses.

Low/High Index of Pupillary Activity (LHIPA)

For each eye, LHIPA is computed via wavelet decomposition (sym16), extracting and normalizing high and low-frequency coefficients, then calculating their ratio. Pupillary activity events are identified using modulus maxima and a universal threshold. Unlike Duchowski's per-sample normalization, which can be biased by uneven sampling after confidence filtering, we normalized by time, resulting in an output of LHIPA values for each eye, their mean, and their absolute difference (for consistency checks). Measures are z-standardized within each participant across tasks, enabling task-level comparisons. Higher z-scored LHIPA values were interpreted as reflecting increased pupillary activity associated with attentional and cognitive demand, while recognizing that pupil dynamics may also be influenced by visual, luminance, and oculomotor factors.

Statistical Analysis

All statistical analyses were conducted in RStudio (v2.1). Data were evaluated for normality and homogeneity of variance prior to parametric testing. Random-intercept linear mixed-effects models were implemented with participant included as a random effect to account for repeated measures across trials. Fixed-effect estimates are reported with corresponding 95% confidence intervals derived using Wald estimation procedures. Effect magnitude for fixed effects in linear mixed-effects models was quantified using standardized beta coefficients (β_{std}), calculated by normalizing fixed-effect estimates to the standard deviation of the outcome variable. For task performance analyses, separate task-specific random intercept linear mixed-effects models were fit for each performance metric. Performance outcomes were modeled as a function of end-effector condition and trial number. For carton transfer and pour analyses, carton weight was additionally included as a fixed effect.

For NASA-TLX analyses, separate task-specific linear mixed-effects models were fit for each TLX subscale, including mental demand, physical demand, temporal demand, performance, effort, and frustration. Subjective workload measures were modeled as a function of end-effector condition and trial number, with participant included as a random intercept. To account for repeated testing across ROIs, false discovery rate (FDR) correction was applied within each task.

For eye-tracking analyses, separate random intercept linear mixed-effects models were fit for LHIPA by task. Eye-tracking measures were modeled as a function of end-effector condition and trial number, with participant included as a random intercept.

For fNIRS analyses, separate task-specific linear mixed-effects models were fit for each prefrontal region of interest (ROI), with mean HbO response modeled as a function of end-effector condition and trial number. To account for repeated testing across ROIs, the Benjamini–Hochberg false discovery rate (FDR) correction was applied within each task. Post hoc Tukey corrections were applied for pairwise comparisons where appropriate. Statistical significance was defined at $p < 0.05$.

In addition, supplementary exploratory multimodal covariate analyses were conducted to evaluate potential shared variance between pupillary and neural workload measures. Representative ROI-specific mixed-effects HbO models were re-fit with

LHIPA included as a fixed-effect covariate, and nested models were compared using likelihood ratio tests (see Supplemental Analysis 2.2).

Results

Task Metrics

Graphical representations of task metric results can be found in section 3 of the supplementary document.

of Blocks Transferred

Participants completed a significantly greater number of block transfers when using the natural hand compared to the prosthesis ($\beta =$ - 42.40, CI[-47.0, -37.0], SE = 2.36, p < 0.001, $\beta_{std} = 1.78$). Across both conditions, trial number did not have a significant effect on block transfer performance ($\beta =$ - 0.31, CI[-1.7, 1.10], SE = 0.71, p = 0.67, $\beta_{std} = 0.01$).

of Checkers Transferred

For the checker transfer task, participants transferred significantly more checkers when using the natural hand than with the prosthesis ($\beta =$ -35.86, CI[-39.1, -32.7], SE = 1.62, p < 0.001, $\beta_{std} = 1.89$). Trial number did not have a significant effect on checker transfer performance ($\beta =$ -0.05, CI[-1.2, 1.10], SE = 0.56, p = 0.925, $\beta_{std} = 0.002$).

of Checkers Stacked

For the checker stacking task, the number of checkers successfully stacked was significantly higher when using the natural hand than when using the prosthesis ($\beta =$ -36.49, CI[-39.8, -33.2], SE = 1.67, p < 0.001, $\beta_{std} = 1.89$). Trial number had no significant effect on stacking performance across conditions ($\beta =$ -0.21, CI[-1.3, 0.88], SE = 0.55, p = 0.708, $\beta_{std} = 0.01$).

of Carton Transfers

Participants completed a significantly greater number of carton transfers when using the natural hand than with the prosthesis ($\beta =$ -20.41, CI[-22.6, -18.2], SE = 1.12, p < 0.001, $\beta_{std} = 1.74$). Trial number ($\beta = 1.22$, CI[0.77, 1.70], SE = 0.22, p < 0.001, $\beta_{std} = 0.10$) had a significant effect on the number of carton transfers completed, while carton weight ($\beta =$ -1.05, CI[-3.79, 1.61], SE = 1.36, p = 0.44, $\beta_{std} = 0.17$) did not have a significant effect on carton transfers completed.

Carton Pour Accuracy

Similarly, the number of carton pours was significantly higher when using the natural hand compared to the prosthesis ($\beta =$ -19.52, CI[-37.6, -1.44], SE = 9.22, p = 0.03, $\beta_{std} = 0.41$). Both trial number ($\beta =$ -1.67, CI[-7.30, 3.95], SE = 2.87, p = 0.56, $\beta_{std} = 0.035$) and carton weight ($\beta =$ -3.48, CI[-30.97, 12.7], SE = 11.38, p = 0.76, $\beta_{std} = 0.07$) had no significant effect on carton pour accuracy.

NASA-TLX

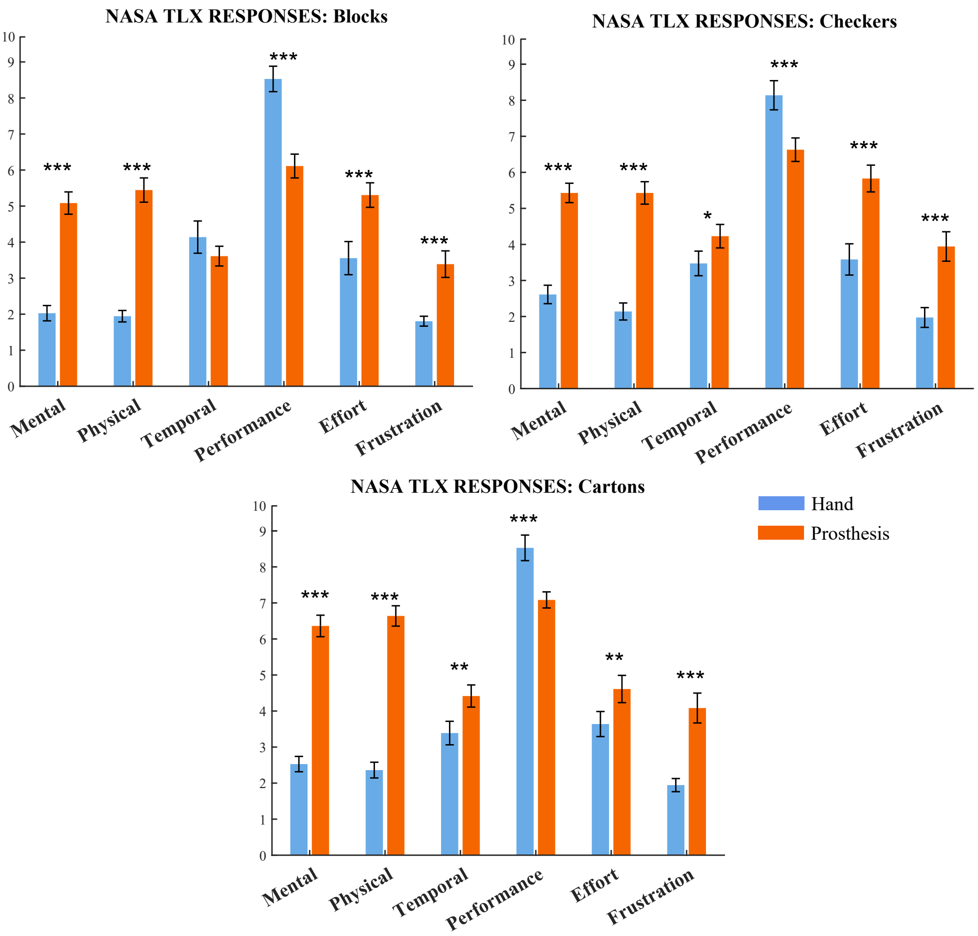

Fig. 3 NASA-TLX subjective workload ratings across tasks and conditions. Mean NASA-TLX subscale ratings (Mental Demand, Physical Demand, Temporal Demand, Perceived Performance, Effort, and Frustration) are shown for the block transfer, checker transfer and stack, and carton transfer and pour tasks. Blue bars represent the natural hand, and orange bars correspond to the prosthesis condition. Bars represent model-estimated marginal means from the linear mixed-effects model; error bars indicate $\pm$ 1 standard error of the estimated means. Asterisks denote statistically significant differences between conditions based on linear mixed-effects modeling (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Subjective workload ratings differed significantly between the natural hand and the prosthesis across all tasks (Fig. 3). During the block transfer task, participants reported significantly lower mental demand (β = -3.05, CI[-3.6,-2.5], SE = 0.28,

$p < 0.001$, $\beta_{std} = 1.39$), physical demand (β = -3.5, CI [-3.9, -3.0],SE = 0.25, $p < 0.001$, $\beta_{std} = 1.48$), effort (β = -1.75, CI[-2.6,-0.8], SE = 0.48, $p < 0.001$, $\beta_{std} = 0.68$), and frustration (β = -1.58, CI[-2.1,-1.0], SE = 0.29, $p < 0.001$, $\beta_{std} = 0.86$), as well as higher perceived performance (β = 2.42, CI[1.7, 3.1], SE = 0.36, $p < 0.001$, $\beta_{std} = 1.01$) when using the natural hand compared to TAMP. No significant difference between conditions was observed for temporal demand (β = -0.52, CI[-0.11,1.6], SE = 0.32, p = 0.11, $\beta_{std} = 0.24$). Trial number had no significant effect on any TLX responses.

Similar trends were observed during the checker transfer and stacking task. Participants reported significantly lower mental demand (β = -2.80, CI[-3.3, -2.3], SE = 0.25, $p < 0.001$, $\beta_{std} = 1.33$), physical demand (β = -3.20, CI[-3.9, -2.7], SE = 0.32, $p < 0.001$, $\beta_{std} = 1.41$), effort (β = -2.27, CI[-3.2, -1.3], SE = 0.49, $p < 0.001$, $\beta_{std} = 0.85$), and frustration (β = -1.95, CI[-2.6, -1.3], SE = 0.35, $p < 0.001$, $\beta_{std} = 0.85$), along with higher perceived performance (β = 1.56, CI[0.8, 2.3], SE = 0.36, $p < 0.001$, $\beta_{std} = 0.67$) when using the natural hand compared to the prosthesis. In contrast to the block transfer task, temporal demand differed significantly between conditions, with higher ratings during prosthesis use (β = 0.79, CI[0.1, 1.4], SE = 0.33, $p < 0.001$, $\beta_{std} = 0.40$). Trial number did not significantly influence subjective workload measures.

During the carton transfer and pour task, participants again reported significantly lower mental demand (β = -3.85, CI[-4.5, -3.3], SE = 0.30, $p < 0.001$, $\beta_{std} = 1.56$), physical demand (β = -4.32, CI[-4.8, -3.8], SE = 0.26, $p < 0.001$, $\beta_{std} = 1.65$), and frustration (β = -2.20, CI[-2.9, -1.4], SE = 0.39, $p < 0.001$, $\beta_{std} = 1.00$), as well as higher perceived performance (β = 1.60, CI[1.0, 2.2], SE = 0.30, $p < 0.001$, $\beta_{std} = 0.82$) when using the natural hand compared to TAMP. Temporal demand was significantly higher during prosthesis use (β = 0.93, CI[0.3, 1.6], SE = 0.33, $p < 0.01$, $\beta_{std} = 0.50$). Perceived effort remained significantly lower for the natural hand (β = -0.92, CI[-1.7, -0.2], SE = 0.39, $p < 0.001$, $\beta_{std} = 0.42$) and trial number had a significant effect on performance ratings (β = 0.25, CI[0.04, 0.47], SE = 0.10, $p < 0.05$, $\beta_{std} = 0.042$).

Prefrontal Cortex Activation

Prefrontal cortex (PFC) activation differed between the natural hand and the prosthesis across tasks. During the block transfer task, oxygenated hemoglobin (HbO) levels were significantly higher during prosthesis use in the right medial PFC (β = 0.71, CI[0.18, 1.25], SE = 0.27, $p < 0.05$, $\beta_{std} = 0.56$) and right dorsolateral PFC (β = 0.52, CI[0.12, 0.92], SE = 0.20, $p < 0.05$, $\beta_{std} = 0.55$). No significant differences were observed in the left dorsolateral (β = 0.05, CI[-0.46, 0.56], SE = 0.36, p = 0.85, $\beta_{std} = 0.04$) or left medial PFC (β = 0.04, CI[-0.12, 0.95], SE = 0.02, p = 0.18, $\beta_{std} = 0.34$), and trial order did not significantly affect any region of PFC activation.

During the checker transfer and stacking task, HbO levels were significantly elevated during prosthesis use in the right medial (β = 0.39, CI[0.04, 0.74], SE = 0.18, $p < 0.05$, $\beta_{std} = 0.54$), right dorsolateral (β = 0.47, CI[0.09, 0.86], SE = 0.20, $p < 0.05$, $\beta_{std} = 0.53$), and left dorsolateral (β = 0.58, CI[0.10, 1.05], SE = 0.24, $p < 0.05$, $\beta_{std} = 0.54$). No significant difference was observed in the left medial PFC (β = 0.43, CI[-0.04, 0.90], SE = 0.24, p = 0.08, $\beta_{std} = 0.41$), and trial number had no significant effect.

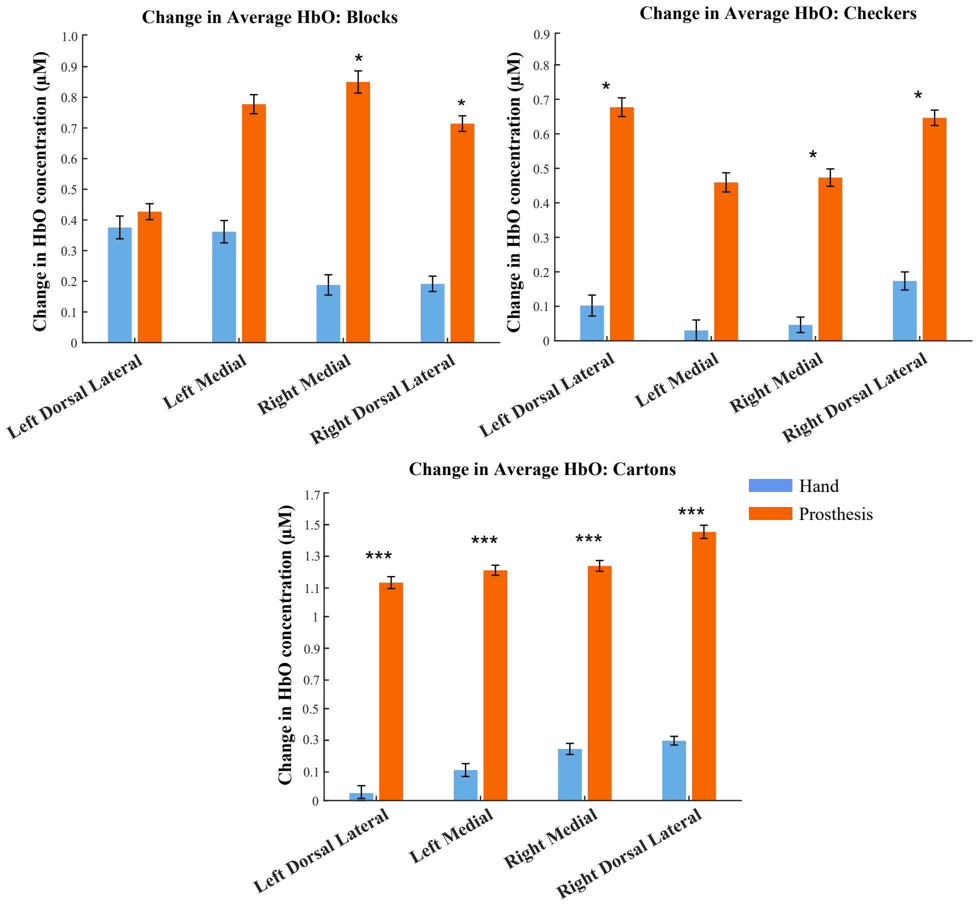

Fig. 4 Task-evoked prefrontal cortex activation measured via fNIRS. Mean changes in oxygenated hemoglobin (HbO) concentration are shown for four prefrontal cortex (PFC) regions during the block transfer, checker transfer and stack, and carton transfer and pour tasks. Blue bars represent the natural hand, and orange bars represent the prosthesis condition. Bars represent model-estimated marginal means from the linear mixed-effects model; error bars indicate ± 1 standard error of the estimated means. Statistical significance between conditions is indicated by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

In the carton transfer and pour task, HbO levels were significantly higher during prosthesis use compared to the natural hand across all PFC regions, including the left dorsolateral ($\beta = 1.14$, CI[0.57, 1.72], SE = 0.29, $p < 0.001$, $\beta_{std} = 0.80$), left medial ($\beta = 1.09$, CI[0.6, 1.57], SE = 0.25, $p < 0.001$, $\beta_{std} = 0.87$), right medial ($\beta = 0.96$, CI[0.56, 1.35], SE = 0.20, $p < 0.001$, $\beta_{std} = 0.81$), and right dorsolateral PFC ($\beta = 1.14$, CI[0.73, 1.54], SE = 0.21, $p < 0.001$, $\beta_{std} = 0.93$). Trial number had a significant effect on activation in the left medial PFC ($\beta = -0.15$, CI[-0.24, -0.05], SE = 0.07, $p < 0.01$, $\beta_{std} = 0.03$).

Low/High Index of Pupillary Activity (LHIPA)

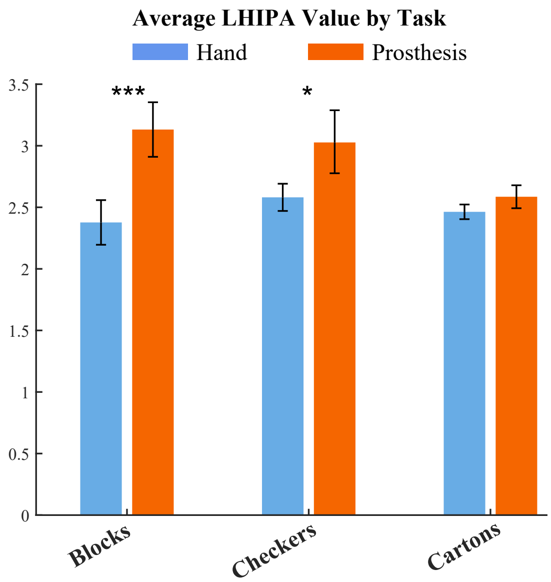

Fig. 5 Mean LHIPA values z-standardized within participant, and are shown for the block transfer, checker transfer and stack, and carton transfer and pour tasks. Blue bars correspond to the natural hand and orange bars correspond to the prosthesis condition. Bars represent model-estimated marginal means from the linear mixed-effects model; error bars indicate ± 1 standard error of the estimated means. Asterisks denote statistically significant differences between conditions based on linear mixed-effects modeling (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Mean binocular LHIPA was used in all statistical analyses. LHIPA values were significantly lower with the natural hand than they were with the prosthesis for the block transfer ($\beta = 0.75$, CI[0.44, 1.07], SE = 0.1, $p < 0.001$, $\beta_{std} = 0.73$) and checker transfer and stack ($\beta = 0.44$, CI[0.09, 0.79], SE = 0.17, $p < 0.05$, $\beta_{std} = 0.55$) tasks. There was no difference in LHIPA values between the natural hand and the prosthesis in the carton transfer and pour task ($\beta = 0.09$, CI[-0.17, 0.37], SE = 0.13, p = 0.47, $\beta_{std} = 0.10$). Trial had no effect on LHIPA value across conditions and tasks.

In addition to LHIPA, gaze allocation analysis indicated task-dependent differences in visual attention strategies that contextualize pupillary responses. Participants using the prosthesis allocated more fixation events toward the end effector, whereas natural hand use involved minimal to no end-effector monitoring and a higher focus on the object throughout the task (see Supplementary Material Section 2, Fig. S2).

Discussion

Across all tasks, participants demonstrated significantly higher task performance when using their natural hand compared to the prosthetic device. These performance differences spanned a range of motor demands, including gross manipulation, fine motor precision, dynamic stabilization, and tasks requiring both accuracy and coordinated control. While expected, these disparities provide critical context for interpreting cognitive load measures, as they reflect both the mechanical limitations of the prosthesis and the increased attentional and executive resources required to successfully complete tasks using a prosthetic device [3, 4].

Consistent with our hypotheses, prosthesis use was associated with increased cognitive demand across all evaluated modalities relative to natural hand use. Importantly, however, the observed differences depended on task complexity, and the measurement modality influenced our ability to detect specific dimensions of cognitive demand rather than the demand itself. This reinforces the value of a multimodal evaluation framework [24–26].

Subjective workload ratings derived from the NASA Task Load Index revealed consistently higher mental and physical demand, greater perceived effort, increased frustration, and lower perceived performance during prosthesis use across all tasks. These findings align with the intended design of the TLX, in which individual subscales probe complementary dimensions of workload, including mental demand, physical demand, emotional response, effort expenditure, and perceived task success.

Notably, temporal demand, defined within the TLX as how hurried or rushed the pace of the task felt, exhibited a clear task-dependent scaling effect. During the simplest task, the block transfer task, participants did not report a significant difference in temporal demand between the natural hand and prosthesis conditions, despite marked differences in objective performance. However, as task complexity increased in the checker transfer and stacking task, temporal demand ratings diverged, with participants reporting a greater sense of time pressure during prosthesis use. This divergence became more pronounced in the most complex task, the carton transfer and pour task, where temporal demand ratings further separated between conditions even when the objective task duration remains constant.

Crucially, all task trials were fixed at one minute, indicating that differences in temporal demand did not reflect objective time constraints but rather participants' subjective perception of task pacing. Simpler, repetitive tasks may feel less rushed or even monotonous, whereas more complex tasks require sustained attention, problem-solving, and continuous monitoring. These demands can compress the subjective experience of time and increase the sensation of being hurried, even when the objective task duration remains constant. In this way, temporal demand appears particularly sensitive to task complexity during prosthesis use, capturing an experiential dimension of cognitive load that extends beyond raw task timing and experimental conditions.

Neural activation patterns measured via fNIRS mirrored these task-dependent trends. Across all tasks, prosthesis use elicited greater prefrontal cortex activation than natural hand use, consistent with increased executive control, monitoring, and decision-making demands [20, 21]. Importantly, the spatial extent of activation scaled with task complexity. During the block transfer task, significant differences were limited to the right prefrontal cortex, whereas more complex tasks elicited significant activation across bilateral dorsolateral and medial prefrontal regions.

This progression suggests a shift from localized executive engagement during simpler tasks to more globally distributed prefrontal recruitment as task demands increase. The convergence between fNIRS-derived neural activation and subjective workload ratings provides physiological grounding for perceived workload, particularly for temporal demand [12]. In this context, fNIRS findings are consistent with the increased executive oversight and sustained attentional demands that may contribute to participants perceiving tasks as more rushed despite fixed durations, as increased prefrontal engagement reflects heightened executive oversight and sustained attentional demands during prosthesis use.

In contrast to subjective and neural measures, behavioral indicators derived from pupillary activity exhibited a uniquely different pattern. Differences in pupillary activity between natural hand and prosthesis use were largest during the simplest task and progressively diminished with increasing task complexity. While significant differences were observed during the block transfer and checker transfer and stacking tasks, no significant difference was detected during the carton transfer and pour task.

This divergence likely reflects task-dependent adaptation of visual attention strategies rather than a reduction in cognitive demand. During the block transfer task, participants using their natural hand commonly adopted a visually static strategy, directing their gaze toward the next target block without visually tracking its movement once grasped. This behavior is supported by the supplementary gaze allocation analysis (Fig. S2), which shows that participants allocated a greater proportion of fixation events to searching for the next block than to any other region in the scene. This resulted in a relatively stable visual scene and reduced the need for continuous visual monitoring. In contrast, prosthesis use required more frequent visual monitoring of the end effector, contributing to greater differences in pupillary activity between conditions. As shown in Fig. S2, participants using their natural hand exhibited minimal fixation allocation toward the end effector (i.e., natural hand) across tasks, whereas in the prosthesis condition, the end effector became one of the most frequently fixated regions in the scene.

As task complexity increased, participants were required to visually track objects as they moved, monitor dynamic changes in object state, and continuously assess spatial relationships within the task environment. In the carton transfer and pour task, visual attention was distributed across multiple dynamic elements, including shifting object mass, container orientation, and pouring accuracy. Under these conditions, both natural hand and prosthesis users exhibited high levels of visual engagement, reducing between-condition differences in pupillary activity.

Importantly, this does not indicate a reduction in cognitive load during complex tasks. Rather, it suggests that pupillary activity may lose sensitivity as a proxy for cognitive load in highly dynamic, visually demanding motor tasks. In such contexts, pupil dilation may reflect reactive visual processing rather than sustained cognitive effort. Prior work has demonstrated the robustness of pupillary measures in controlled hand–eye coordination tasks [13–15]; however, tasks involving whole-body interaction within physical, non-simulated environments may decouple pupil dynamics from cognitive load in meaningful ways [5].

Additionally, exploratory multimodal covariate analysis was additionally performed to assess potential shared variance between pupillary and prefrontal workload responses. Inclusion of LHIPA as a covariate in representative ROI-specific HbO models did not significantly improve model fit, suggesting that the pupillary and neural workload signatures may reflect distinct task-related processes (see Supplementary Analysis 2.2).

Taken together, these findings demonstrate that different modalities capture distinct and task-dependent aspects of cognitive demand [24–26]. Subjective workload and neural measures converged in their sensitivity to both prosthesis use and task complexity, while pupillary activity revealed shifts in visual attention strategies that were not directly proportional to overall cognitive load. Rather than representing a limitation, this divergence highlights the value of multimodal assessment in disentangling perceptual, behavioral, and neural components of cognitive effort.

Based on the task-dependent sensitivities observed across modalities, the present findings suggest that the selection of cognitive load measures should be guided by the complexity and visual structure of the motor task under study. For simple, repetitive manipulation tasks (e.g., grasp-and-transfer), subjective workload instruments like the NASA Task Load Index may be sufficient to detect differences in cognitive demand between conditions. In visually consistent, highly controlled tasks with repeated measures, eye-tracking metrics may also provide useful insight, particularly when visual strategies remain stable across trials. However, in situations where cognitive load may vary throughout a task and be influenced by task structure, fatigue, or impaired self-assessment, including fNIRS can provide critical physiological context by revealing underlying neural effort that may not be fully captured through self-report measures alone. For complex, dynamic tasks that involve continuous adaptation, variable object properties, or rich visual environments, a multimodal approach that combines subjective workload assessment and neural measures is recommended. In these contexts, eye-tracking may offer valuable information regarding visual attention strategies but should not be relied upon as a standalone indicator of cognitive load, as its sensitivity may diminish when visual monitoring becomes inherently demanding. Together, these considerations support the informed, task-aware selection of cognitive load metrics when comprehensive multimodal evaluation is not feasible.

This study represents a foundational step toward multimodal characterization of cognitive load during upper-limb prosthesis use. A primary limitation is the use of participants without limb loss employing a mock prosthesis, which does not capture the lived experience, embodiment, or long-term adaptation of individuals with limb loss. However, this controlled cohort was intentionally selected to assess the sensitivity, reliability, and task dependence of multimodal cognitive load measures before clinical translation with real prosthesis users. Future work will extend this framework to prosthesis users to investigate how embodiment, training history, and sensory integration shape multimodal cognitive load signatures. Additional studies should investigate eye-tracking learning effects over longer timescales. Because all participants completed the natural-hand condition before the prosthesis condition, end-effector effects could not be fully dissociated from session-order effects. Although task trials were randomized within each condition and trial number was included as a covariate in the statistical models, future studies should counterbalance end-effector order to more definitively separate device-related effects from learning, fatigue, habituation, and other session-dependent influences.Together, these efforts will enable refinement of multimodal evaluation strategies for both experimental and clinical prosthesis assessment.

Conclusion

This study demonstrates that cognitive load during upper-limb prosthesis use is multifaceted and cannot be fully captured by any single measurement modality. Across activities of daily living, prosthesis use was associated with increased cognitive demand relative to the natural hand, but the expression of this demand varied systematically with task complexity and evaluation method. Subjective workload and neural measures converged in their sensitivity to prosthesis-related effort and task demands, while behavioral indicators of visual attention revealed task-dependent shifts in attentional strategy rather than overall changes in cognitive load. Together, these findings underscore the importance of multimodal cognitive load assessment and provide practical guidance for selecting appropriate evaluation tools based on task characteristics. By establishing a controlled baseline framework for multimodal evaluation, this work lays the foundation for future studies in clinical populations and supports the development of prosthetic systems and assessment strategies that minimize cognitive burden and enhance functional usability.

Declarations

0.0.1 Ethics approval

All participants provided informed consent before participation in the study. Experimental procedures were approved by the Johns Hopkins Medicine Institutional Review Board (IRB #00147458).

0.0.2 Consent for publications

All participants provided informed consent for publication as outlined by their participant consent and authorization form approved by the Johns Hopkins Medicine Institutional Review Board (IRB #00147458).

0.0.3 Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

0.0.4 Conflict of interest

fNIR Devices, LLC manufactures the optical brain imaging instrument and licensed IP and know-how from Drexel University. Dr. Ayaz was involved in the technology development and thus offered a minor share in the startup firm fNIR Devices, LLC. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

0.0.5 Funding

This work is funded by the National Science Foundation (grant# 2146206).

0.0.6 Authors' contribution

L.V., H.A, and J.D.B. conceived the experiment, L.V. and G.G conducted the experiment, and L.V. and G.G. analyzed the results with support by H.A. All authors reviewed the manuscript.

0.0.7 Acknowledgments

The authors would like to thank Dr. Yigit Topoglu for his insight and guidance on the implementation of fNIRs, and the structure of the experimental protocol.

0.0.8 Supplementary information

Supplementary results, analysis, and experimental setup details are provided in the supplemental document. A supplementary video is also available, which details the experimental tasks and highlights differences in gaze behavior.

References

[1] Ziegler-Graham, K., MacKenzie, E.J., Ephraim, P.L., Travison, T.G., Brookmeyer, R.: Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050. Archives of Physical Medicine and Rehabilitation **89**(3) (2008) https://doi.org/10.1016/j.apmr.2007.11.005

[2] Caruso, M., Harrington, S.: Prevalence of Limb Loss and Limb Difference in the United States: Implications for Public Policy. Technical report, Avalere Health, Washington, DC (February 2024)

[3] Biddiss, E., Chau, T.: Upper limb prosthesis use and abandonment: A survey of the last 25 years (2007). https://doi.org/10.1080/03093640600994581

[4] Young, B.H.: The Bionic-Hand Arms Race: High-Tech Hands are Complicated, Costly, and Often Impractical. IEEE Spectrum **59**(10) (2022) https://doi.org/10.1109/MSPEC.2022.9915629

[5] Marchand, C., De Graaf, J.B., Jarrassé, N.: Measuring mental workload in assistive wearable devices: a review (2021). https://doi.org/10.1186/s12984-021-00953-w

[6] Park, J., Zahabi, M.: Cognitive Workload Assessment of Prosthetic Devices: A Review of Literature and Meta-Analysis (2022). https://doi.org/10.1109/THMS.2022.3143998

[7] Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology **52**(C) (1988) https://doi.org/10.1016/S0166-4115(08)62386-9

[8] Parr, J.V.V., Galpin, A., Uiga, L., Marshall, B., Wright, D.J., Franklin, Z.C., Wood, G.: A tool for measuring mental workload during prosthesis use: The Prosthesis Task Load Index (PROS-TLX). PLoS ONE **18**(5 MAY) (2023) https://doi.org/10.1371/journal.pone.0285382

[9] Hart, S.G.: NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, . an Francisco, CA, USA, 16–20 October 2006; Sage Publications : Los Angeles, CA, USA, **50** (2006)

[10] Raveh, E., Friedman, J., Portnoy, S.: Evaluation of the effects of adding vibrotactile feedback to myoelectric prosthesis users on performance and visual attention in a dual-task paradigm. Clinical Rehabilitation **32**(10) (2018) https://doi.org/10.1177/0269215518774104

[11] Thomas, N., Ung, G., Ayaz, H., Brown, J.D.: Neurophysiological Evaluation of Haptic Feedback for Myoelectric Prostheses. IEEE Transactions on Human-Machine Systems **51**(3) (2021) https://doi.org/10.1109/THMS.2021.3066856

[12] Thomas, N., Miller, A.J., Ayaz, H., Brown, J.D.: Haptic shared control improves neural efficiency during myoelectric prosthesis use. Scientific Reports **13**(1) (2023) https://doi.org/10.1038/s41598-022-26673-2

[13] Krejtz, K., Duchowski, A.T., Niedzielska, A., Biele, C., Krejtz, I.: Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. PLoS ONE **13**(9) (2018) https://doi.org/10.1371/journal.pone.0203629

[14] Iqbal, S.T., Zheng, X.S., Bailey, B.P.: Task-evoked pupillary response to mental workload in human-computer interaction. In: Conference on Human Factors in Computing Systems - Proceedings (2004). https://doi.org/10.1145/985921.986094

[15] Hess, E.H., Polt, J.M.: Pupil size in relation to mental activity during simple problem-solving. Science **143**(3611) (1964) https://doi.org/10.1126/science.143.3611.1190

[16] Zahabi, M., White, M.M., Zhang, W., Winslow, A.T., Zhang, F., Huang, H., Kaber, D.B.: Application of Cognitive Task Performance Modeling for Assessing Usability of Transradial Prostheses. IEEE Transactions on Human-Machine Systems **49**(4) (2019) https://doi.org/10.1109/THMS.2019.2903188

[17] Zhang, W., White, M., Zahabi, M., Winslow, A.T., Zhang, F., Huang, H., Kaber, D.: Cognitive workload in conventional direct control vs. pattern recognition

control of an upper-limb prosthesis. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings (2017). https://doi.org/10.1109/SMC.2016.7844587

[18] H, A., B, O., K, I., P.A, S., R, M., R., P.: Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: Empirical examples and a technological development. Fronteirs in Human Neuroscience **7**, 1–13 (2013) https://doi.org/10.3389/fnhum.2013.00871

[19] JA, M., A, C., AE, K., MD, Z., H, A.: Mental workload assessment by monitoring brain, heart, and eye with six biomedical modalities during six cognitive tasks. Fronteirs in Neuroergonomics **5**(1345507) (2024) https://doi.org/10.3389/fnrgo.2024.1345507

[20] McKendrick, R., Mehta, R., Ayaz, H., Scheldrup, M., Parasuraman, R.: Prefrontal Hemodynamics of Physical Activity and Environmental Complexity during Cognitive Work. Human Factors **59**(1) (2017) https://doi.org/10.1177/0018720816675053

[21] Fishburn, F.A., Norr, M.E., Medvedev, A.V., Vaidya, C.J.: Sensitivity of fNIRS to cognitive state and load. Frontiers in Human Neuroscience **8**(1 FEB) (2014) https://doi.org/10.3389/fnhum.2014.00076

[22] Sattar, N.Y., Kausar, Z., Usama, S.A., Farooq, U., Shah, M.F., Muhammad, S., Khan, R., Badran, M.: fNIRS-Based Upper Limb Motion Intention Recognition Using an Artificial Neural Network for Transhumeral Amputees. Sensors **22**(3) (2022) https://doi.org/10.3390/s22030726

[23] Möller, S., Ramstrand, N., Hagberg, K., Rusaw, D.: Cortical brain activity in transfemoral or knee-disarticulation prosthesis users performing single- and dual-task walking activities. Journal of Rehabilitation and Assistive Technologies Engineering **7** (2020) https://doi.org/10.1177/2055668320964109

[24] Broadbent, D.P., D'Innocenzo, G., Ellmers, T.J., Parsler, J., Szameitat, A.J., Bishop, D.T.: Cognitive load, working memory capacity and driving performance: A preliminary fNIRS and eye tracking study. Transportation Research Part F: Traffic Psychology and Behaviour **92** (2023) https://doi.org/10.1016/j.trf.2022.11.013

[25] Wang, P., Wang, H., Zhang, H.: Cognitive Workload Assessment in Aerospace Scenarios: A Cross-Modal Transformer Framework for Multimodal Physiological Signal Fusion. Multimodal Technologies and Interaction **9**(9) (2025)

[26] Zakeri, Z., Arif, A., Omurtag, A., Breedon, P., Khalid, A.: Multimodal Assessment of Cognitive Workload Using Neural, Subjective and Behavioural Measures in Smart Factory Settings. Sensors (Basel, Switzerland) **23**(21) (2023) https://doi.org/10.3390/s23218926

[27] Y, L., H, A., PA, S.: Multisubject "Learning" for Mental Workload Classification Using Concurrent EEG, fNIRS, and Physiological Measures. Frontiers in Human Neuroscience **11**(389) (2017) https://doi.org/10.3389/fnhum.2017.00389

[28] Ayaz, H., A., S.P., M., C.A.a.I., K., I., B., O.: Using MazeSuite and Functional Near Infrared Spectroscopy to Study Learning in Spatial Navigation. Journal of Visualized Experiments (56) (2011) https://doi.org/10.3791/3443

[29] L., P., D., C., H.H., H.: Myoelectric Control Based on a Generic Musculoskeletal Model: Toward a Multi-User Neural-Machine Interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering **26**(7) (2018) https://doi.org/10.1109/TNSRE.2018.2838448

[30] Velásquez, L., Brown, J.D.: Understanding the Utility of State-Based Haptic Feedback in Tendon-driven Anthropomorphic Prostheses. IEEE Transactions on Neural Systems and Rehabilitation Engineering (2025) https://doi.org/10.1101/2025.02.27.25323021

[31] Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B.: Optical brain monitoring for operator training and mental workload assessment. NeuroImage **59**(1) (2012) https://doi.org/10.1016/j.neuroimage.2011.06.023

[32] Duchowski, A.T., Krejtz, K., Gehrer, N.A., Bafna, T., Bækgaard, P.: The Low/High Index of Pupillary Activity. In: Conference on Human Factors in Computing Systems - Proceedings (2020). https://doi.org/10.1145/3313831.3376394

[33] Mathiowetz, V., Volland, G., Kashman, N., Weber, K.: Adult norms for the Box and Block Test of manual dexterity. The American journal of occupational therapy. : official publication of the American Occupational Therapy Association **39**(6) (1985) https://doi.org/10.5014/ajot.39.6.386

[34] Haverkate, L., Smit, G., Plettenburg, D.H.: Assessment of body-powered upper limb prostheses by able-bodied subjects, using the Box and Blocks Test and the Nine-Hole Peg Test. Prosthetics and Orthotics International **40**(1) (2016) https://doi.org/10.1177/0309364614554030

[35] Sigirtmac, I., Oksuz, C.: Investigation of reliability, validity, and cutoff value of the Jebsen-Taylor Hand Function Test. Journal of Hand Therapy **34**(3) (2021) https://doi.org/10.1016/j.jht.2020.01.004

[36] Krinis, S.L.L., Chadwell, A., Kenney, L., Smit, G.: A quantitative comparison between the mHand Adapt passive adjustable hand prosthesis and its predecessor, the Delft Self-Grasping Hand. PLoS ONE **19**(3 March) (2024) https://doi.org/10.1371/journal.pone.0300469

[37] Vasluian, E., Bongers, R.M., Reinders-Messelink, H.A., Burgerhof, J.G.M., Dijkstra, P.U., Van Der Sluis, C.K.: Learning effects of repetitive administration of

the southampton hand assessment procedure in novice prosthetic users. Journal of Rehabilitation Medicine **46**(8) (2014) https://doi.org/10.2340/16501977-1827

Supplementary Information

for

Multimodal Evaluation of Cognitive Load During Prosthetic Device Use in Activities of Daily Living

Lorena Velásquez, Gabriella Goytizolo, Hasan Ayaz, Jeremy D. Brown

1 Supplementary Instrumentation and Experimental Setup

Figure S1: Sequential procedure for fitting the fNIRS headband and eye-tracking glasses prior to data collection. (A) Initial placement of the fNIRS headband over the participant's forehead. (B) Alignment of the headband relative to anatomical landmarks. (C) Securing and tensioning the headband to ensure stable contact. (D) Placement of an elastic cap to further stabilize optodes during data collection. (E) Placement and adjustment of the eye-tracking glasses.

Before data collection, the fNIRS headband (Model 2000s, fNIR Devices, LLC) was positioned over the participant's prefrontal cortex following standardized placement procedures (Fig. B4). The lower edge of the headband was aligned approximately just above the participant's eyebrows, ensuring coverage of bilateral dorsolateral and medial prefrontal regions. Alignment was confirmed relative to midline anatomical landmarks (glabella and nasion) to maintain consistent optode positioning across participants. Care was taken to smooth down the hair on the head and the eyebrows to prevent any strands from interfering with fNIRS optodes. This was particularly important on the edges of the device, as they aligned with the temples, often the narrowest portion of the forehead. Optodes interacting with unmovable obstacles were noted before further calibration. With these placements, you are tailoring the fit to the person's anatomy, and you aim for the best that you can do. The goal is to preserve as many optodes as possible.

The headband was secured with its integrated Velcro tensioning system and adjusted to ensure firm yet comfortable contact between optodes and the skin. An additional headband was placed on top of the fNIRS band to block ambient light, paying special attention to blocking light on the top of the band, where the optodes meet the skin, and tucking beneath the band to prevent light exposure above the eyebrow. To minimize motion artifacts, the wire leads of the fNIRS optode were fixed behind the participant's head using Velcro. Participants were asked to perform small head movements during setup to confirm stability before moving on to the task. Participants were also asked about any tightness or pressure points along the band to ensure it was secure but not uncomfortable to wear.

Signal quality was verified and calibrated using FNIRSoft before each experimental block. Channels were inspected for acceptable light intensity levels, minimal saturation, and low ambient light interference. Particular attention was paid to optodes along the superior edge of the band and near the brow line, where loose skin contact or improper shielding could introduce noise. If channels exhibited poor contact, elevated noise, or unstable baselines, the headband tension and ambient light shielding were adjusted, and the optodes were repositioned as needed before proceeding. Data collection began only after stable hemodynamic signals were confirmed across all channels.

Participants were then asked to put on the pupil core glasses as they would any set of glasses. The eye-tracking cameras were then adjusted to capture the participant's pupils. Because the Pupil Core eye-tracking system emits near-infrared light, care was taken to minimize optical interference between systems. The fNIRS headband was positioned above the eye-tracking frame, with the additional shielding used to block out light. During setup, real-time signal inspection would capture and attempt to correct the effects of eye-tracker-based light contamination in fNIRS channels through adjusting signal tuning gains in fNIRSoft. This sequential fitting and validation procedure ensured stable multimodal acquisition throughout task performance. Eye-tracking and fNIRS calibration were performed with the participant standing, head facing down toward the task table to maintain consistency with the head position used during the experiment.

Channels with unstable levels of light contamination were noted. Data was collected from all optodes, but optodes with more than 20% data corruption due to unstable light exposure were excluded from the final dataset. This led to analysis being conducted only on the top row of optodes spanning all four brain regions, as the bottom row was affected by near-infrared light.

2 Supplementary Analysis

2.1 Gaze Allocation

To complement pupillary measures of cognitive load, gaze behavior was analyzed to characterize visual attention strategies during task performance. These analyses were designed to quantify how participants distributed their visual attention across task-relevant elements, including the prosthetic end effector (or natural hand), task objects, and the surrounding workspace. While gaze metrics were not used as a primary measure of cognitive load, they provide important contextual information for interpreting pupillary responses and task-dependent attentional demands.

2.1.1 Data Acquisition and Pre-processing

Eye-tracking data was collected using the Pupil Core system (Pupil Labs GmbH, Berlin, Germany), which records binocular gaze position and pupil diameter at 200 Hz. Raw gaze data were exported and processed offline in the Pupil systems' associated processing software (PupilPlayer). Samples with confidence values below 0.6 were excluded to mitigate the effects of blinks, occlusions, and tracking loss. The remaining data were temporally aligned with task segments corresponding to individual trials.

2.1.2 Regions of Interest (ROI) Definition

To quantify gaze allocation, the visual scene was segmented into task-relevant regions of interest (ROIs). ROIs were defined based on consistent spatial features of the experimental setup and included:

- **End Effector**: the participant's hand (natural or prosthetic)
- **Task Objects**: objects actively manipulated during each task (e.g., blocks, checkers, cartons)
- **Target Area**: destination locations for object placement or transfer
- **Other**: all remaining areas within the scene not directly involved in manipulation

ROI boundaries were defined in the image plane of the scene camera and were applied consistently across trials. Specific ROI labels for each task are provided in the corresponding supplementary figures.

2.1.3 Gaze Allocation Metrics

Gaze behavior was quantified as the proportion of fixation events directed toward each predefined region of interest (ROI) during a trial. Fixations were identified using a dispersion-based algorithm with a threshold of 1.5° and a duration window of 100–4000 ms, selected to capture both brief and sustained fixations while maintaining spatial precision in a visually complex environment. For each trial, ROI labels were manually assigned to fixation events based on their spatial location within the scene. Gaze allocation was then computed as the proportion of total fixation events directed toward each ROI, obtained by normalizing the number of fixations within each region by the total fixation count for that trial. These proportions were averaged across trials within each task and condition. This metric provides a summary of visual attention distribution across task-relevant elements without relying on fixation duration, which can be sensitive to task dynamics and environmental variability.

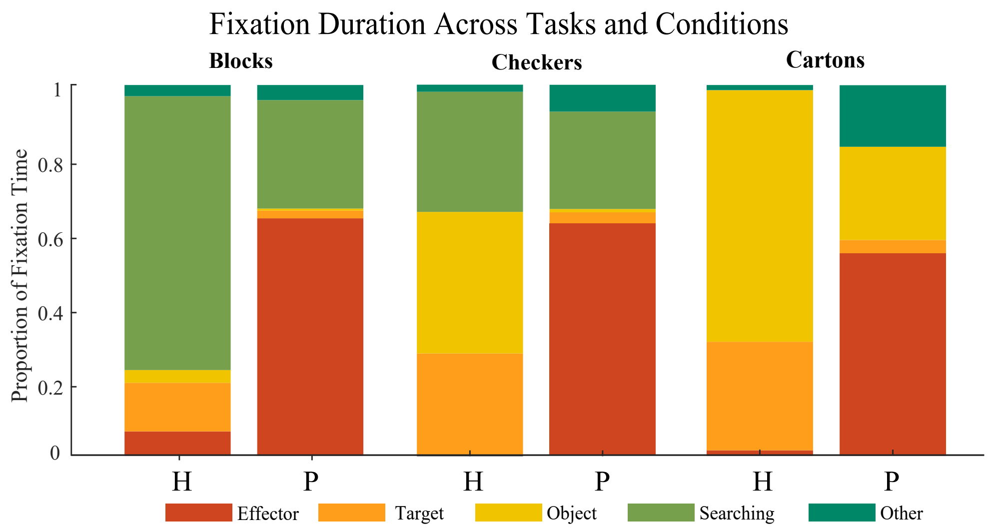

Figure S2: Proportion of fixation events directed toward predefined regions of interest (ROIs), including the end effector, target area, task objects, searching, and other regions, during the block transfer, checker transfer and stack, and carton transfer and pour tasks across both end effector conditions, natural hand (H) and prosthesis (P). Values represent the mean proportion of fixation events averaged across trials and participants for each task and condition.

2.2 Exploratory Multimodal Covariate Analysis

To further explore potential interactions between workload modalities, exploratory covariate analysis was conducted between the two biologically derived measures collected in this study: pupillary dynamics (LHIPA) and prefrontal hemodynamic response (Δ HbO). This comparison was motivated by the alignment observed between fNIRS and subjective workload measures, while eye-tracking metrics demonstrated more task-dependent divergence.

We evaluated whether pupillary workload metrics explained additional variance in the neural response measured through fNIRS. Representative ROI-specific mixed-effects models were evaluated for the block transfer task, which demonstrated significant condition-related effects in both fNIRS and eye tracking, and the weighted transfer task, which demonstrated significant effects in fNIRS but not eye tracking. Nested models were compared using likelihood ratio tests to assess whether inclusion of LHIPA as a covariate improved model fit for Δ HbO responses. Base models included condition and trial as fixed effects with participant as a random intercept, while covariate models additionally included LHIPA. Inclusion of LHIPA did not significantly improve model fit in either task condition (all $p > 0.69$), and condition-related increases in prefrontal activation remained significant following inclusion of the pupillary metric.

3 Supplementary Results

3.1 Task Metrics

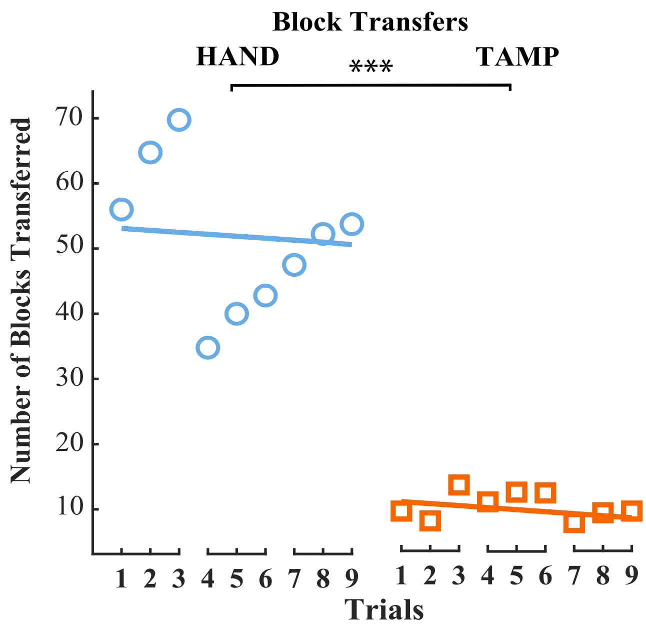

Figure S3: The average number of block transfers for each condition across trials, where the individual data points represent the average for each trial (for all participants in each condition), and the solid lines indicate the model's prediction. * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

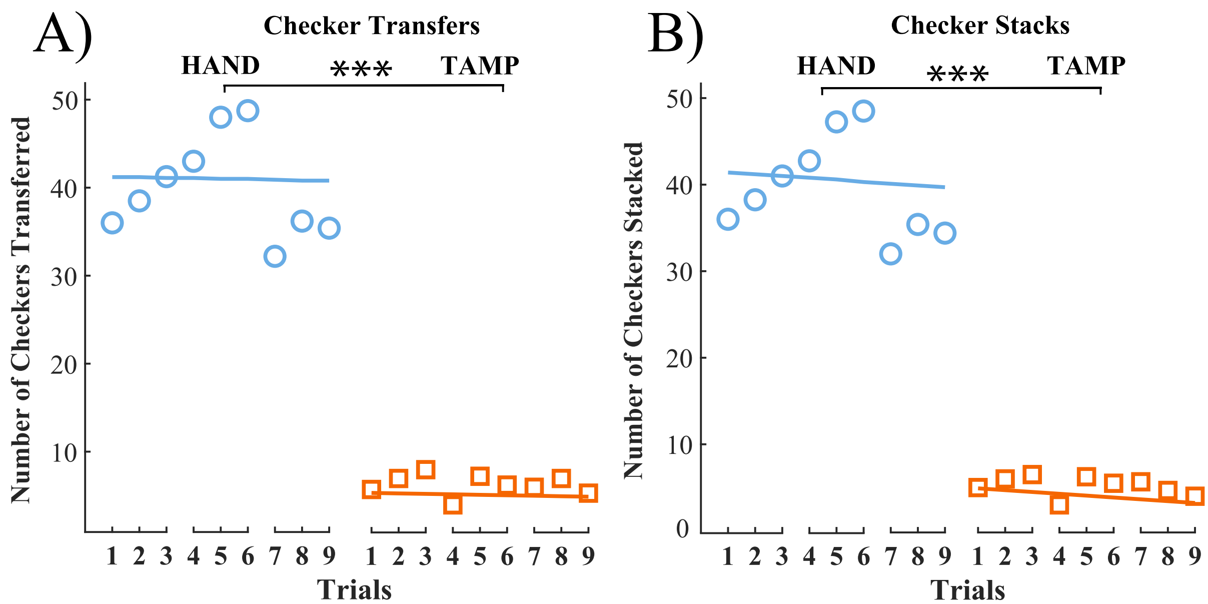

Figure S4: The average number of checker: (A) transfers and (B) stacks for each condition across trials, where the individual data points represent the average for each trial (for all participants in each condition), and the solid lines indicate the model's prediction. * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

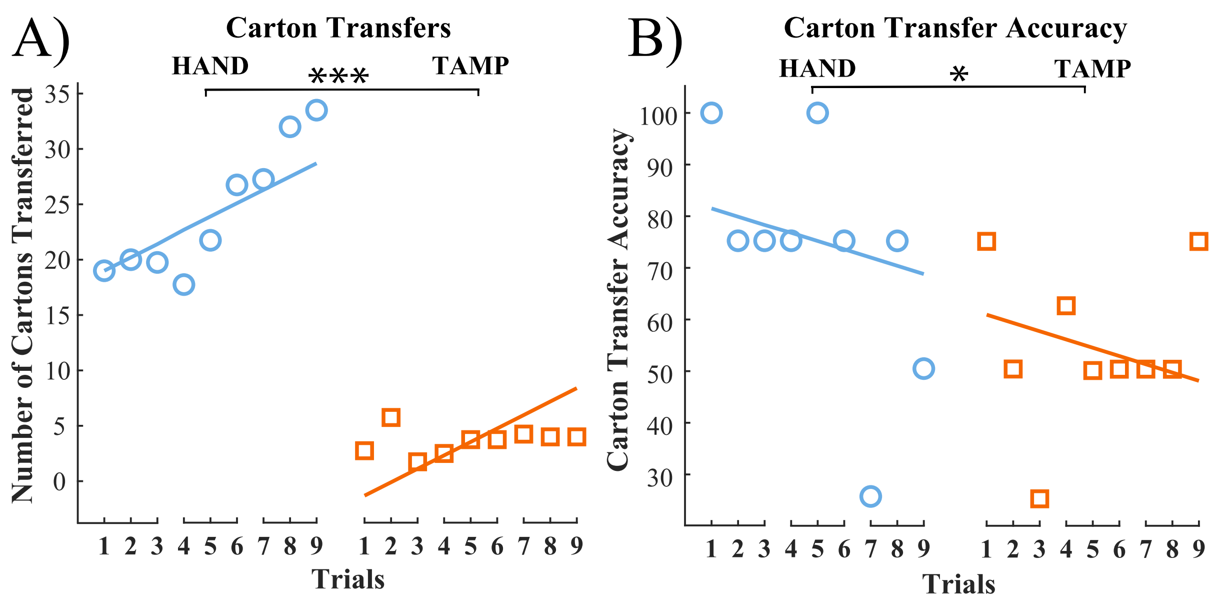

Figure S5: The average number of carton: (A) transfers and (B) transfer accuracy for each condition across trials, where the individual data points represent the average for each trial (for all participants in each condition), and the solid lines indicate the model's prediction. * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.