

# Effects of Shared Control on Cognitive Load and Trust in Teleoperated Trajectory Tracking

Jiahe Pan<sup>1</sup>, Jonathan Eden<sup>1</sup>, Denny Oetomo<sup>1</sup>, and Wafa Johal<sup>1</sup>

**Abstract**—Teleoperation is increasingly recognized as a viable solution for deploying robots in hazardous environments. Controlling a robot to perform a complex or demanding task may overload operators resulting in poor performance. To design a robot controller to assist the human in executing such challenging tasks, a comprehensive understanding of the interplay between the robot’s autonomous behavior and the operator’s internal state is essential. In this paper, we investigate the relationships between robot autonomy and both the human user’s cognitive load and trust levels, and the potential existence of three-way interactions in the robot-assisted execution of the task. Our user study (N=24) results indicate that while the autonomy level influences the teleoperator’s perceived cognitive load and trust, there is no clear interaction between these factors. Instead, these elements appear to operate independently, thus highlighting the need to consider both cognitive load and trust as distinct but interrelated factors in varying the robot autonomy level in shared-control settings. This insight is crucial for the development of more effective and adaptable assistive robotic systems.

**Index Terms**—Human-Robot Collaboration, Acceptability and Trust, Telerobotics and Teleoperation

## I. INTRODUCTION

FROM teleoperating a robot to inspect a disaster site to using a rehabilitation robot after stroke, human-robot interaction (HRI) aims to combine the respective strengths of humans and robots to achieve desired task outcomes. However, how to best configure robot autonomy during HRI such that it improves performance without being negatively perceived by its user remains an open problem. When a user and a robot work collaboratively to complete a task, problems such as unsafe operation or reduced user performance can arise if the user distrusts the robot and its behaviors [1], or if they have a sufficiently large cognitive load such that they cannot reliably command the robot [2].

The main research question of this study is: *How does the user’s cognitive load and trust vary with changes in the robot autonomy level during HRI?* To answer this question, we conducted a teleoperation study using a shared-control scheme that allowed for varying the robot autonomy level<sup>†</sup>. This used

a trajectory tracking task to minimize the effects of confounding variables while enabling a fundamental understanding of robot autonomy’s effect on human cognitive load and trust. Through metrics that captured trust and cognitive load, via both objective methods and questionnaires, we looked to identify the relationship between the autonomy level and these factors. Specifically, we formulate three main hypotheses:

- **H1** - The participants’ cognitive load will decrease with increasing robot autonomy
- **H2** - The participants’ trust in the robot will increase with increasing robot autonomy
- **H3** - There will be an autonomy dependent relationship between trust and cognitive load

## II. RELATED WORKS

Teleoperation refers to robot operation using human intelligence via a human-machine interface [3], and has shown robustness and real-world applicability in assisting humans to perform complex tasks in hazardous and uncertain environments [4]. As both robot and human intelligence are often required to successfully complete these tasks, teleoperation often makes use of shared-control [5], where command inputs from the human and the robot’s autonomous controller are integrated and arbitrated to determine the robot’s resulting actions. Here, existing shared-control methods have mostly focused on the task performance and/or the physical characteristics of the interaction such as the human and/or robot effort. These interaction metrics are also often task-specific, such as minimizing collisions with obstacles while navigating [6] or maximizing the percentage of successful object grasps [7].

It has however been shown that robot autonomy and behavior can also impact the human user’s internal state, such as their cognitive load [2], [8] and trust [1], [9]. Cognitive load is as an indicator of task complexity based on the number of conceptual elements that need to be held in the user’s mind at any one time to solve a specific task [10]. Here, subjective assessment schemes (e.g., Subjective Workload Assessment Technique (SWAT) [11], Rating Scale of Mental Effort (RSME) [12], Task Load Index (NASA-TLX) [13]) have been shown to reliably capture the perception of workload [14] but can only be conducted at low frequencies. In contrast, the dual-task paradigm of performing a secondary task (e.g., reproducing melodies [15], [16], performing simple math [17], [18], responding to stimuli [19], [20]) concurrently with a primary task and physiological measures (e.g., pupil dilation [21], [22], cardiac activity [23], [24], brain activity [25], [26]) allow the objective measure of cognitive load at higher frequencies

Manuscript received: December, 8, 2023; Revised March, 6, 2024; Accepted April, 14, 2024.

This paper was recommended for publication by Editor Angelika Peer upon evaluation of the Associate Editor and Reviewers’ comments. This work was partially supported by the Australian Research Council (Grant Nos. DE210100858 and DP240100938)

<sup>1</sup>The authors are with the Faculty of Engineering and Information Technology, The University of Melbourne, Australia. {jpan@student.unimelb.edu.au, {eden.j, doetomo, wafa.johal}@unimelb.edu.au

Digital Object Identifier (DOI): see top of this page.

<sup>†</sup>A demo video is available at <https://youtu.be/yizsBG1QJog>, and the code and data are found at <https://github.com/mpan31415/AutonomyCLTrust>.

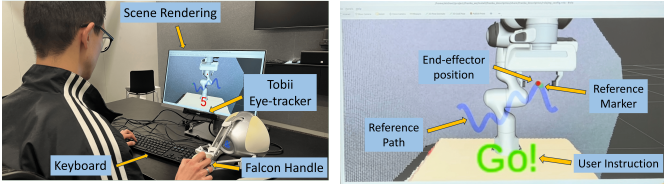


Fig. 1: Experiment setup. **(Left)** Task setup. The primary trajectory-tracking task was performed via a Novint Falcon controller and visual feedback on the screen. The secondary task involved using the “space key” to tap a given rhythm. **(Right)** Primary task visualisation. The trajectory to be followed was displayed in blue. At each time instant, the red dot showed the position of the center of the end-effector and the green dot showed the target.

[14] but suffer from the impact of confounding factors such as learning effects [27] and expertise in multitasking [28].

A human’s trust in another agent is defined as a multidimensional latent variable that mediates the relationship between events in the past and the former agent’s subsequent choice of relying on the latter in an uncertain environment [1]. Trust is therefore an internal measure experienced by humans, making it difficult to capture objectively [29]. Existing methods that have been explored include physiological information [30] and brain imaging [31] which do not necessarily capture trust independent of confounding factors. As a result, subjective questionnaires such as the Multi-Dimensional Measure of Trust (MDMT) [32] which captures trust across 2 categories -  $\{Capacity\ Trust, Moral\ Trust\}$  are most widely employed in HRI studies [33], [34].

Existing works have only studied the effect of autonomy on either the user’s cognitive load [35], [36] or trust [37] separately. However, there is evidence that cognitive load and trust may not be independent measures [38], [39]. Therefore, it is important to understand whether cognitive load and trust vary based on an underlying dependency between them, or if they are in fact independently influenced by autonomy.

To the best of our knowledge, there is currently no reported study on the three-way relationship between robot autonomy, cognitive load, and trust. Having a fundamental understanding of such a relationship is crucial, as with it roboticists can better design adaptive robots with varying autonomy levels that adjust their behavior in response to their user. Such adaptive schemes are gaining interest in areas like autonomous driving [37] and personal service robots [40].

### III. METHOD

Our study evaluated the cognitive load and trust of participants during their teleoperation of a (Franka Emika Research 3) robotic arm. The task entailed tracking a 3-dimensional trajectory with the robot’s end-effector. Trajectory tracking was selected as the primary task as it is a component within a variety of real-world applications, such as robotic welding [41], [42], cutting [43], [44] and painting [45]. Furthermore, it enabled the use of clearly defined robot autonomy levels within the blending control scheme (see Section III-B) and has a clear metric on task performance through the well-established root mean square error (RMSE) metric (see Section III-D). In addition to this primary task, we employed a dual-

task methodology to measure cognitive load, with a secondary rhythmic tapping task.

#### A. User Interface

The user interface was designed to replicate a real-world teleoperation scenario, where the participant operated from a distinct environment with no direct perception of the true physical robot, which was situated in an adjacent room. Instead, we presented a virtual rendering of the robot’s real-time state and a pre-captured Point Cloud of its surrounding environment on a computer screen (Figure 1 Left) using Rviz [46]. Teleoperation was performed via the handle of a Novint Falcon haptic controller, which was connected to the robot in the adjacent room via a local Ethernet connection. No haptic feedback was provided to participants in this study. Participants performed the secondary tapping task by pressing “space” on a computer keyboard (Figure 1 Left).

#### B. Primary Task

The primary task required the participant to teleoperate the robot to track a trajectory (see Figure 1 Right) of  $T = 10$  s duration using the robot’s end-effector. Participants were asked to “as accurately as possible” follow the 3-dimensional time-parameterized trajectory with  $t \in [0, T]$

$$\begin{aligned} x(t) &= x_0 \\ y(t) &= y_0 + d_y(t/T) \\ z(t) &= z_0 + d_z f(t), \end{aligned} \quad (1)$$

where

$$f(t) = \sin(a_1(t + \theta)) + \sin(a_2(t + \theta)) + \sin(a_3(t + \theta)) \quad (2)$$

is a multi-sine curve with parameters  $a_1, a_2, a_3 \in \mathbb{N}$  and  $\theta \in \mathbb{R}$ . Here,  $(x_0, y_0, z_0)$  represents a constant offset from the robot’s base, and  $d_y = 0.3$  m and  $d_z = 0.15$  m are the width and height parameters of the trajectory. Using (1) and parameters  $(a_1, a_2, a_3, \theta)$ , we generated a set of trajectories  $S = \{(2, 3, 4, \frac{4\pi}{3}), (1, 3, 4, \pi), (2, 2, 5, \pi), (2, 3, 5, \frac{8\pi}{5}), (2, 4, 5, \pi)\}$ . These different trajectories were used to make it unlikely for memorization during multiple rounds of repetition. The parameters were chosen with similar amplitudes and frequencies to ensure similar tracking complexity.

The robot was commanded through shared-control between the human and its own autonomous controller. This adopted a blending control (or convex combination) formulation [47], [48], where the input position was computed as

$$\mathbf{u} = \gamma \mathbf{u}_r + (1 - \gamma) \mathbf{u}_h, \quad \gamma \in [0, 1] \subset \mathbb{R}. \quad (3)$$

Here  $\mathbf{u}_r \in \mathbb{R}^3$  denotes the reference robot end-effector position input which corresponded to the trajectory value at that time, and  $\mathbf{u}_h \in \mathbb{R}^3$  denotes the reference end-effector position from the human’s input - determined by a constant 3 times scaling of the Falcon handle position, which was read at 500 Hz. The scalar  $\gamma$  represents the level of *robot autonomy*, where  $\gamma = 1$  corresponds to complete robot control (*full autonomy*) and  $\gamma = 0$  corresponds to complete human control (*no autonomy*). By

tap	tap	pause	pause	tap	tap	pause	pause	tap ...
	$\tau$		$3\tau$		$\tau$		$3\tau$	

Fig. 2: Rhythm Method Visualization (in four-four time). Participants followed the rhythm by tapping the “space” bar. The repeating pattern consisted of a short inter-tap interval followed by a long one (adapted from [15]).

altering  $\gamma$  the relative autonomy between the human and robot was varied. The resulting input  $\mathbf{u} \in \mathbb{R}^3$  was passed in ROS to the Kinematics and Dynamics Library (Orocos) inverse-kinematics solver to compute joint positions, which were then tracked by the robot’s joint controllers running at 500 Hz.

### C. Secondary Task

The secondary task consisted of the “Rhythm Method” [15]. This involved the participant tapping a pre-recorded rhythm from memory while concurrently performing the primary task. Our implementation followed that of [15], except that we generated multiple rhythms that used different tempos with the same pattern. Similar to our trajectory design, this was done to minimize the chance of participants learning the rhythm across multiple rounds. As shown in Figure 2, the pattern consisted of a short inter-tap interval of duration  $\tau$ , followed by a long inter-tap interval of duration  $3\tau$ . Hence, the speed of the rhythms were inversely related to and directly determined by the duration of the short inter-tap interval,  $\tau$ . Using this parametrization, we generated 5 rhythms of different speeds with  $\tau \in [400, 600]$  ms. For consistency with a dual-task paradigm, we explicitly instructed the participants that the primary task always *takes priority* over the secondary task.

Using this secondary tapping task alongside the primary teleoperation task, we aimed to capture the participant’s cognitive load and gain insights into how they can effectively manage their cognitive resources when faced with concurrent demands. This mirrored situations in which individuals need to multitask or divide their attention in real-world scenarios.

### D. Measures

We used objective and subjective measures to investigate the relationship between the robot’s *autonomy level* in the teleoperation task and the participant’s cognitive load and trust.

1) *Demographics Information*: For each participant, we collected basic demographic information via an initial questionnaire at the start of the session. Additionally, due to their potential relevance during teleoperation, we included the following three items which participants rated on a 10-point Likert scale from 1 (*strongly disagree*) to 10 (*strongly agree*): i) I trust new technology in general; ii) I play computer games regularly; and iii) I am proficient in a musical instrument.

2) *Performance and Physiological Measures*: During each trial, we collected data pertaining to the participant’s performance on the primary and secondary tasks, along with eye-tracking data. This was used to evaluate cognitive load.

**Trajectory Tracking Error** - The robot position information during each trial was recorded at 20 Hz. The trajectory tracking error was calculated as the RMSE between the reference and recorded trajectories. Feedback from initial pilot

studies showed that participants found it difficult to perceive and track depth components using the RViz display. Therefore, the RMSE calculation of the trajectory tracking error did not include the error in tracking the constant depth.

**Tapping Error** - Following [15], we recorded the time stamp of each tap executed by the participant. We then calculated the duration of each of the inter-tap intervals separately for both interval types (short and long inter-tap intervals). Finally, the tapping error was calculated by subtracting the average baseline error from the average recorded error for each trial and then normalizing with respect to the interval lengths.

**Pupil Diameter** - Larger pupil dilation has been associated with more intense cognitive processing [21], [22]. Therefore, we recorded pupil diameter data during each trial at 40 Hz using the Tobii Pro X2 screen-based eye-tracker. To derive the cognitive load estimate from this data, we applied the method outlined in [49], which involves performing a wavelet decomposition of the data sequence and computing a ratio of low/high frequencies of pupil oscillation.

3) *Self-Reported Measures*: After each round of five trials, the following measures were collected via questionnaires:

**Perceived Autonomy** - The “perceived autonomy questionnaire” [50] has been widely used in HRI [51], [52] to capture user perception of autonomy. Adapting the questionnaire, we included the item: “*How autonomous did you feel the robot was?*”, which participants reported on a 10-point discrete scale. This self-reported measure acted as a manipulation check and assessed if participants perceived the change of the robot’s autonomy level between rounds.

**Cognitive Load** - We administered the NASA-TLX questionnaire [13] with all six original sub-scales to provide a comprehensive assessment of cognitive workload.

**Trust** - The MDMT questionnaire [32] is composed of 8 sub-scales in each of the *Capacity Trust* and *Moral Trust* categories. Since our study design focused on the human’s trust relating to the robot’s behavior in the trajectory tracking task, we chose to only include the 8 sub-scales of *Capacity Trust*, and excluded all items in the *Moral Trust* category.

We complemented the MDMT with a single-scale trust question: “*How much do you trust the robot?*”, which participants also reported using a 10-point discrete scale. This was included to see if a single trust measure could be reliably used as a high-frequency self-reported measure instrument.

### E. Study Design

1) *Conditions*: To test our hypotheses, we used a counterbalanced, within-subject design to examine the effects of robot autonomy while controlling for order effects. The within-subject condition was the robot’s *autonomy level*, which we designed to be either *high* ( $\gamma = \{0.6, 0.8\}$ ) or *low* ( $\gamma = \{0.0, 0.2\}$ ). Two blocks of trials were conducted using the corresponding two values of  $\gamma$  for each *autonomy* condition. The balanced experimental design meant that half of the participants were exposed to these two autonomy levels in *increasing* ( $\gamma \in \text{low then high}$ ) order, and the other half in *decreasing* ( $\gamma \in \text{high then low}$ ) order.

Using the experimental conditions, we performed an a priori power analysis to determine our sample size using G\*Power

[53]. Considering a repeated-measure ANOVA with 0.8 power,  $\alpha = 0.05$  and a medium effect size of  $f = 0.25$ , the calculation resulted in a sample size of  $N = 24$  (split evenly into the two groups).

2) *Participants*: Ethical approval for the study was granted by The University of Melbourne’s Human Ethics Committee under project ID 27750. The 24 participants were aged from 18 to 32 years ( $M = 23.5, SD = 3.88$ ), 18 identified as female and 6 as male. All participants were asked to perform the primary task using their preferred dominant hand (22 right-handed). Participants received a \$20 gift voucher as compensation at the end of the approximately 60 minute study.

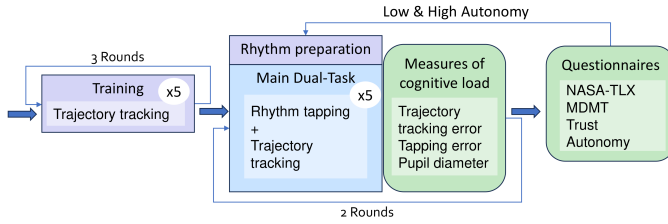


Fig. 3: Experimental procedure including the training phase and main phase for a given autonomy level  $\gamma$ .

#### F. Experimental Procedure

Each session had two phases (Figure 3): *training* and *main*. Within each phase, participants completed three rounds in training and two in main - where each round contained five trials -  $\{t_1, \dots, t_5\}$ . During each trial, the task performance and pupil diameter data was continuously recorded (see Section III-D2), while participants self-reported their cognitive load and trust after each round (see Section III-D3).

1) *Training Phase*: Participants were initially shown the shared-control setup and the visualization (see Figure 1), followed by two practice trials for them to familiarize with the system and the trajectory tracking task. They were then trained to perform the tracking task in a single *baseline* round. This round was used as a manipulation check to ensure that there was no effect of the secondary task on primary task performance and therefore required participants to perform the primary tracking task with *no autonomy* and without the secondary tapping task. After, participants performed two rounds of both the tracking and tapping tasks, at a *medium autonomy* level and a *no autonomy* level, to familiarize with the dual-task scheme and for all participants to have a common baseline for our comparisons. Each round also included training for the tapping task, following the procedure of [15]. In the training phase, participants were explicitly informed about the autonomy conditions of each round, which they could use as a reference for perceiving the robot’s autonomy levels.

2) *Main Phase*: The two rounds of the main phase required participants to perform both teleoperation and tapping. Although they were informed that the robot’s autonomy level might change from round to round, participants were not told the robot autonomy level, nor any information on its ability to perform the task. In each round, a different rhythm was used according to a pre-generated random-ordering. Participants

were trained for tapping the rhythm, including a recorded section used as their baseline tapping performance. They then performed the five trials of the round, after which they completed the questionnaire to report their own subjective assessments of cognitive load and trust during the round.

In each trial, there was an initial 5 s preparation window shown as a 5 s countdown on the screen (Figure 1), during which participants heard a short recording of the rhythm. One of the five pre-generated trajectories was also displayed in a random-ordering, which included a blue line indicating the *reference path* to track, and a green *reference marker* which followed the path from start to finish over the 10 s trajectory duration (see Section III-B). In addition, the robot’s end-effector position was shown as a red dot located between its parallel grippers. After the preparatory 5 s, the countdown text changed to the word “Go!”, indicating that the trial has started. Participants were instructed to only perform both tasks during the 10 s trajectory duration, after which the on-screen text changed to “Stop!”, indicating that the trial had finished. Each session took approximately 60 minutes, including the 10-minute introduction block and a 10-minute block for each autonomy condition across both the training and main phases.

#### IV. RESULTS

Preliminary analysis showed no order effects on any measure. Therefore all subsequent analysis combined the data from the *increasing* and *decreasing* autonomy groups. After all measures were verified to be normally distributed using the Shapiro-Wilk test, we performed repeated-measures ANOVA with *autonomy* as the within-subject variable. Results were considered significant at the threshold  $\alpha < .05$ . The ANOVA results are summarized in Table I.

Measure	(DFn, DFd)	F	p	$\eta^2$
Perceived Autonomy	(1, 22)	62.104	<0.001	0.262
NASA-TLX	(1, 22)	46.615	<0.001	0.262
Tapping Error (short)	(1, 22)	1.137	0.298	0.028
Tapping Error (long)	(1, 22)	0.873	0.36	0.008
Pupil Diameter Index	(1, 22)	0.088	0.769	0.0001
MDMT	(1, 22)	41.418	<0.001	0.234
Traj Tracking Error	(1, 22)	129.005	<0.001	0.748

TABLE I: Summary of ANOVA results for all measures against *autonomy*. DFn and DFd are the degrees of freedom in the numerator and denominator respectively, p specifies the p-value,  $\eta^2$  is the generalized effect size.

The demographics items listed in Section III-D1 were included as potential covariates. However, preliminary analysis showed no clear effect of these items on any of our dependent variables. Therefore, they were not included in the main analysis. Furthermore, potential learning effects across the five trials within each round were checked using Linear Mixed Models. While no learning effect was observed for any measure under both high and low *autonomy* conditions, only the second round of each autonomy condition ( $\gamma = 0.8$  for *high* autonomy and  $\gamma = 0.2$  for *low* autonomy) was used in the analysis to eliminate any adaptation to a new autonomy level. For the same reason, the first trial data of each round was also removed from the analysis.

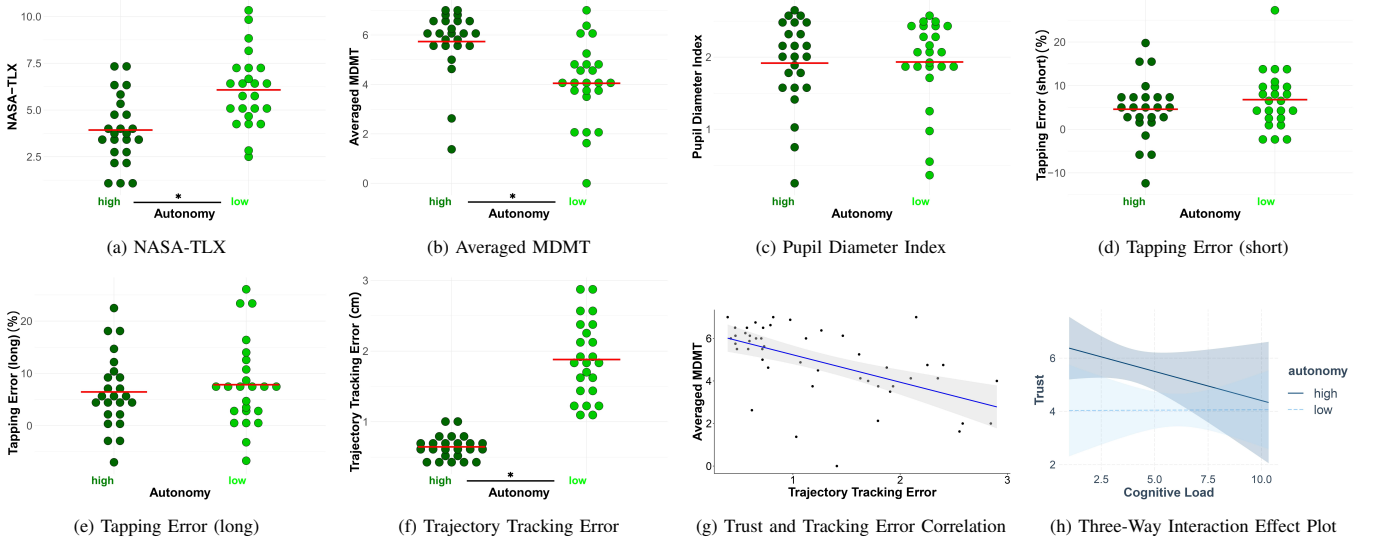


Fig. 4: (a)-(f) Plots of data distribution of each measure against *autonomy*. Red bars show the *mean* values of each distribution. \* denotes  $p < .05$ . (g) Correlation between averaged *MDMT* and *trajectory tracking error*. (h) Illustration of interaction effect between *cognitive load*, *trust*, and robot *autonomy*.

Measure	Low Autonomy		High Autonomy	
	$\mu$	$\sigma$	$\mu$	$\sigma$
Traj Tracking Error (cm)	2.163	0.402	0.834	0.069
Tapping Error (short) (%)	7.458	9.615	5.513	6.768
Tapping Error (long) (%)	8.129	10.402	6.849	10.402

TABLE II: Mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the *trajectory tracking* and *tapping errors* (short & long) under low and high *autonomy* conditions.

#### A. Perceived Autonomy

We evaluated whether the participants' perceived autonomy levels were consistent with the true autonomy levels. ANOVA indicated a clear *autonomy level* effect on the *perceived autonomy* ( $F(1, 22) = 62.104, p < .001, \eta^2 = .262$ ). This suggests that participants correctly perceived the applied autonomy levels and serves as an experiment manipulation check.

#### B. Cognitive Load

We first investigated the effect of *autonomy* on the three cognitive load measures: *NASA-TLX* (Fig. 4a), *tapping error* (Fig. 4d, 4e), and *pupil diameter index* (Fig. 4c). The *autonomy level* affected the raw *NASA-TLX* score ( $F(1, 22) = 46.615, p < .001, \eta^2 = .262$ ), where the 6 subscales showed internal consistency (Cronbach's  $\alpha = .762$ ). In contrast to this result, no clear effect of *autonomy* was observed on the *tapping error* for both the short ( $F(1, 22) = 1.137, p = .298, \eta^2 = .028$ ) and long ( $F(1, 22) = .873, p = .36, \eta^2 = .008$ ) rhythm intervals (tapping error mean and standard deviation is summarized in Table II). Similarly, no clear *autonomy* effect was observed on the *pupil diameter index* ( $F(1, 22) = .088, p = .769, \eta^2 = .0001$ ). Therefore while the subjective *NASA-TLX* load assessment indicated that users perceived lower cognitive load when the robot had higher autonomy, the dual-task and physiological objective measures did not show the same effect.

#### C. Trust

The *MDMT* (Fig. 4b) averaged across the *Reliable* and *Capable* sub-categories (Cronbach's  $\alpha = .968$ ) indicated a clear *autonomy* effect ( $F(1, 22) = 41.418, p < .001, \eta^2 = .248$ ). Similarly, the *autonomy level* was also observed to affect the *single-scale trust* measure ( $F(1, 22) = 33.169, p < .001, \eta^2 = .234$ ). Here, Pearson correlation analysis between averaged *MDMT* and data from the *single-scale trust* question found these two measures to be strongly positively correlated ( $r(46) = .873, p < .001$ ). Overall, these results indicate that users had higher trust under the *high autonomy* condition, despite having lower control authority in this condition.

#### D. Teleoperation Task Performance

We first compared the participants' tracking performance in the *baseline* and *no autonomy* rounds. A paired t-test showed no clear difference between the two means ( $t(23) = -.429, p = .672, d = -.083$ ), indicating that the secondary task did not impact the participants' primary task performance. The ANOVA showed an effect of *autonomy* on *trajectory tracking error* (Fig. 4f,  $F(1, 22) = 129.005, p < .001, \eta^2 = .748$ ), where participants performed trajectory tracking more accurately with a higher level of robot *autonomy*.

We further investigated the effect of tracking performance on the level of trust in the robot. A Pearson correlation analysis between the averaged *MDMT* measure and *trajectory tracking error* found the two measures to be negatively correlated ( $r(46) = -.563, p < .001$ ) as shown in Figure 4g. This suggests that participants adopted lower levels of trust towards the robot when they struggled to accurately track the trajectory.

#### E. Cognitive Load & Trust

We also explored the relationship between *autonomy*, *cognitive load* and *trust*. A multiple regression showed a collective effect between *cognitive load* (*NASA-TLX*), *autonomy* and *trust* (average *MDMT*), ( $F(3, 44) = 5.734, p = .002, R^2 =$



0.2811,  $R_{adj}^2 = 0.232$ ). However, there was no clear interaction between *cognitive load* and *autonomy* on *trust* ( $t = -0.96, p = .342$ ). Figure 4h shows the relationship between *cognitive load* and *trust*, grouped by *autonomy*. With high robot *autonomy*, there was a potential trend of participants reporting lower *trust* with higher *cognitive load* and vice versa. Under low *autonomy*, *trust* levels seemed approximately constant across the spectrum of *cognitive load* levels.

## V. DISCUSSION

### A. Impact of Autonomy on Subjective Cognitive Load Measure

Our results show that the subjective *cognitive load* is lower with higher robot *autonomy*. This is consistent with **H1** (and previous results in other applications [35], [36]) and suggests that users perceived the robot's intervention as reducing their own workload. However, in contrast to the effects observed on the subjective measure, neither of the objective measures indicated a clear effect of *autonomy* on *cognitive load*.

While all three methods have been shown to be effective at quantifying cognitive load, they each record at different time-scales - self-assessment is averaged over the entire round, tapping is recorded at the pattern frequency and pupil-diameter is recorded at 40Hz. Ideally for applications like teleoperation, cognitive load would be measured in real-time, enabling the robot's actions to be adapted in response to sudden load changes. Due to the discrepancy across measures, it is uncertain if our metrics are capturing different aspects of workload and of which is most appropriate for future real-time measurement. Here, a unifying framework across measures taken at multiple time-scales [54] might enable clearer evaluation.

All of the tested methods also suffer either from individual bias or other confounding factors which are difficult to control for in a realistic teleoperation setting [27], [28]. For instance, participants repeatedly reported that high *autonomy* made accurate trajectory tracking "more manageable and interesting". This suggests that the design of dual-task setups may need to consider the user's cognitive resource allocation as a function of not only the task's cognitive demand, but also its engagement. Therefore, given the observed inconsistency in our results, determining which measure provides the "ground-truth" of cognitive load remains an open problem.

### B. Autonomy Affects Trust Levels

The *MDMT* and the *single-scale trust* questionnaires support **H2**, where participants' trust towards the robot increased with higher *autonomy*. As the *MDMT* assesses both the reliability and capability aspects of trust, this finding suggests that participants could recognize and understand the robot's behavior and identify its superior performance in the trajectory tracking task. These results are however likely to be feedback-reliant [55], where the visual feedback in our teleoperation setup, which showed both the path and time information of the reference trajectory, could have enabled better understanding of the robot's actions. Such communication of the robot's behavior is likely critical in establishing trust in autonomy, and therefore should be further studied. Furthermore, while we have considered participants' individual expertise through the

demographics items and evaluated potential learning effects through preliminary analysis, we believe that rising expertise in operating the robot and increasing task familiarity may still affect the user's trust and cognitive load, and therefore need to be considered in the design of effective shared control systems.

The strong correlation between the averaged *MDMT* and the *single-scale trust* measure suggests that a *single-scale* measure may capture a subjective evaluation of trust as well as the *MDMT* for this class of tasks. Here, the use of a *single-scale* trust measure may also assist to develop higher frequency trust measures, as users can self-report more frequently. Indeed, using a similar single-scale trust question, [56] collected binary trust information (trust / distrust) from participants in 25 s time intervals in a simulated driving scenario, which was then used to adapt the assistance of an autonomous driving system. However, as trust is not a binary measure, this motivates further research into understanding its nuanced relationship with *autonomy* to enable the design of finer-grained robot shared-control systems.

### C. Does Cognitive Load Affect Trust?

Our results do not support **H3**, as we did not find an interaction between *cognitive load* and *trust* based on *autonomy*. However, visual inspection of the three-way interaction plot (Figure 4h) appears to suggest a potential trend in *trust* across the *cognitive load* values, and this *trust* - *cognitive load* relationship is different under the high and low *autonomy* conditions. The downwards trend in the *high autonomy* condition indicates that when the robot has high control in the task, participants either have more *trust* in the robot and experience less *cognitive load*, or have less *trust* and perceive the task to be more *cognitively demanding*. In contrast with *low autonomy*, where the human user exerted more effort into the task, their *trust* in the robot remains approximately constant for all perceived levels of *cognitive load*.

The absence of a three-way interaction effect may suggest that both cognitive load and trust are indeed measures which are independently influenced by the robot's autonomy. On the other hand, the potential correlation shown in Figure 4h motivates further investigation using finer autonomy variations, which may produce alternative results that unveil the existence of more complex underlying relationships between robot autonomy and the human's cognitive load and trust.

### D. Generalization to Other Factors

In this study, we designed the robot's autonomous controller to be highly accurate in its trajectory tracking. Therefore, as participants reported trust on the *Performance* category of the *MDMT*, it is possible that the observed effect of autonomy on trust might reflect participants being more comfortable with better robot performance, rather than them having more trust in the robot. It is also possible that the more accurate performance may lead to a feeling of less cognitive load since less consideration is given to improving accuracy, which resonates with similar findings in [57] which indicate that automated systems could lead to lower levels of effort exerted by the operator. Here, previous works have demonstrated that

factors such as robot error [58], [59] and system transparency [60] can indeed impact the operator's state during the interaction. Therefore, future studies should consider imperfect robot behaviour.

It is also worth noting that the shared control implementation in this study used the blending formulation typical to robotics literature. However, differing results may be observed for other methods of sharing control for which the allocation of autonomy may be different, such as through inferring human intent [61] or solving for the robot's control using optimization techniques [62]. To better generalize our findings, other controller designs should also be investigated.

While the chosen task of trajectory tracking allowed us to obtain a fundamental understanding of robot autonomy's effect on human cognitive load and trust while minimizing the effects of confounding factors, it is possible that the results of this study may be task-specific, and different results may be obtained from other tasks and alternative setups. Furthermore, as the observed results are likely to be feedback-reliant, and specifically for the visual feedback used in this study, a limitation arises from participants finding it difficult to perceive the depth dimension of the RViz display. It is therefore worth investigating how the visual feedback used in this study or alternative visual information may impact the user's state and task performances.

Additionally, previous works have shown that providing haptic feedback to the operator in shared control settings can result in better task performance [63]–[66], especially in situations with added workload such as from a concurrent second task [67]. While [66] also found that haptic feedback led to lower mental effort exerted by the user, from our results it is also possible that visual feedback alone may be sufficient to create the same effect. Therefore, future works should explore how the integration of visual and various types of haptic feedback under different levels of robot autonomy affect the task performance, the operator's state, and their perception of the autonomy level.

Finally while our setup was designed to replicate teleoperation, we minimized the delay between the operator's command and the robot's motion in our setup through high-frequency communication and tuned PID joint controllers. Here, delay may also have an effect on user perception and the task performance, especially in tasks such as trajectory tracking which require high precision under time constraints. Therefore, teleoperation setups with more realistic delay should be considered in future work.

## VI. CONCLUSION

This paper explored the relationship between robot autonomy and human cognitive load and trust during a collaborative HRI task. While our hypothesis of observing lower cognitive load under higher robot autonomy was confirmed via questionnaire data, the objective physiological and dual-task performance measures produced results that did not align with those from the subjective assessment. Additionally, our analysis revealed no clear interaction between trust and cognitive load, thus motivating future work to use a more

continuous spectrum of autonomy levels instead of binary (high, low) classification to further investigate the relationship between robot autonomy and the user's cognitive load and trust. The effects of varying the robot's reliability or the control interface properties on the user's cognitive load and trust levels are also promising directions worth investigating.

## REFERENCES

- [1] B. C. Kok and H. Soh, "Trust in robots: Challenges and opportunities," *Curr. Robot. Reports*, vol. 1, pp. 297–309, 2020.
- [2] S. Baltrusch, F. Krause, A. de Vries, W. van Dijk, and M. de Looze, "What about the human in human robot collaboration? a literature review on hrc's effects on aspects of job quality," *Ergonomics*, vol. 65, no. 5, pp. 719–740, 2022.
- [3] J. Cui, S. Tosunoglu, R. Roberts, C. Moore, and D. W. Repperger, "A review of teleoperation system control," in *Proc. Florida Conf. Recent Adv. Robot.*, pp. 1–12, CiteSeer, 2003.
- [4] B. Álvarez, A. Iborra, A. Alonso, and J. A. De la Puente, "Reference architecture for robot teleoperation: Development details and practical use," *Cont. Eng. Practice*, vol. 9, no. 4, pp. 395–402, 2001.
- [5] D. Kent, C. Saldanha, and S. Chernova, "A comparison of remote robot teleoperation interfaces for general object manipulation," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, p. 371–379, 2017.
- [6] X. Deng, Z. L. Yu, C. Lin, Z. Gu, and Y. Li, "A bayesian shared control approach for wheelchair robot with brain machine interface," *IEEE Trans. Neural Sys. Rehabil. Eng.*, vol. 28, no. 1, pp. 328–338, 2019.
- [7] K. Z. Zhuang, N. Sommer, V. Mendez, S. Aryan, E. Formento, E. D'Anna, F. Artoni, F. Petrini, G. Granata, G. Cannaviello, *et al.*, "Shared human-robot proportional control of a dexterous myoelectric prosthesis," *Nature Machine Intell.*, vol. 1, no. 9, pp. 400–411, 2019.
- [8] A. Martinetti, P. K. Chemweno, K. Nizamis, and E. Fosch-Villaronga, "Redefining safety in light of human-robot interaction: A critical review of current standards and regulations," *Front. Chem. Eng.*, vol. 3, p. 32, 2021.
- [9] A. Langer, R. Feingold-Polak, O. Mueller, P. Kellmeyer, and S. Levy-Tzedek, "Trust in socially assistive robots: Considerations for use in rehabilitation," *Neurosci. & Biobehavioral Reviews*, vol. 104, pp. 231–239, 2019.
- [10] J. L. Plass, R. Moreno, and R. Brünken, "Cognitive load theory," 2010.
- [11] G. B. Reid and T. E. Nygren, "The subjective workload assessment technique: A scaling procedure for measuring mental workload," in *Human Mental Workload*, vol. 52 of *Advances in Psych.*, pp. 185–218, 1988.
- [12] F. Zijlstra and L. Van Doorn, *The Construction of a Scale to Measure Perceived Effort*. University of Technology, 1985.
- [13] S. G. Hart, "Development of NASA TLX: Result of empirical and theoretical research," *California: San Jose State University*, 2006.
- [14] C. Marchand, J. B. De Graaf, and N. Jarrassé, "Measuring mental workload in assistive wearable devices: a review," *J. NeuroEng. and Rehabil.*, vol. 18, no. 1, pp. 1–15, 2021.
- [15] B. Park and R. Brünken, "The rhythm method: A new method for measuring cognitive load — an experimental dual-task study," *Appl. Cogn. Psychol.*, vol. 29, no. 2, pp. 232–243, 2015.
- [16] R. Sun and J. B. Shea, "Probing attention prioritization during dual-task step initiation: a novel method," *Exp. Brain Res.*, vol. 234, pp. 1047–1056, 2016.
- [17] H.-I. Lee, S. Park, J. Lim, S. H. Chang, J.-H. Ji, S. Lee, J. Lee, *et al.*, "Influence of driver's career and secondary cognitive task on visual search behavior in driving: a dual-task paradigm," *Adv. in Phys. Educ.*, vol. 5, no. 04, p. 245, 2015.
- [18] P.-F. Tang, H.-J. Yang, Y.-C. Peng, and H.-Y. Chen, "Motor dual-task timed up & go test better identifies prefrailty individuals than single-task timed up & go test," *Geriatrics & Gerontology Int.*, vol. 15, no. 2, pp. 204–210, 2015.
- [19] M. W. Scerbo, R. C. Britt, and D. Stefanidis, "Differences in mental workload between traditional and single-incision laparoscopic procedures measured with a secondary task," *The American J. Surgery*, vol. 213, no. 2, pp. 244–248, 2017.
- [20] M. Wirzberger, R. Herms, S. Esmaeili Bijarsari, M. Eibl, and G. D. Rey, "Schema-related cognitive load influences performance, speech, and physiology in a dual-task setting: A continuous multi-measure approach," *Cogn. Res.: Principles and Implications*, vol. 3, pp. 1–16, 2018.

- [21] M. K. Eckstein, B. Guerra-Carrillo, A. T. M. Singley, and S. A. Bunge, "Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?," *Dev. Cogn. Neurosci.*, vol. 25, pp. 69–91, 2017.
- [22] M. M. White, W. Zhang, A. T. Winslow, M. Zahabi, F. Zhang, H. Huang, and D. B. Kaber, "Usability comparison of conventional direct control versus pattern recognition control of transradial prostheses," *IEEE Trans. on Human-Machine Syst.*, vol. 47, no. 6, pp. 1146–1157, 2017.
- [23] C. Collet, P. Averty, and A. Dittmar, "Autonomic nervous system and subjective ratings of strain in air-traffic control," *Appl. Ergon.*, vol. 40, no. 1, pp. 23–32, 2009.
- [24] B. Mehler, B. Reimer, J. F. Coughlin, and J. A. Dusek, "Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers," *Transp. Res. Rec.*, vol. 2138, no. 1, pp. 6–12, 2009.
- [25] M. W. Miller, J. C. Rietschel, C. G. McDonald, and B. D. Hatfield, "A novel approach to the physiological measurement of mental workload," *Int. J. Psychophysiol.*, vol. 80, no. 1, pp. 75–78, 2011.
- [26] W. Klimesch, "Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Res. Reviews*, vol. 29, no. 2–3, pp. 169–195, 1999.
- [27] S. Esmaeili Bijarsari, "A current view on dual-task paradigms and their limitations to capture cognitive load," *Front. Psychol.*, vol. 12, p. 648586, 2021.
- [28] T. Strobach, M. Becker, T. Schubert, and S. Kühn, "Better dual-task processing in simultaneous interpreters," *Front. Psychol.*, vol. 6, p. 1590, 2015.
- [29] M. Chita-Tegmark, T. Law, N. Rabb, and M. Scheutz, "Can you trust your trust measure?," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, p. 92–100, 2021.
- [30] S. C. Kohn, E. J. de Visser, E. Wiese, Y.-C. Lee, and T. H. Shaw, "Measurement of trust in automation: A narrative review and reference guide," *Front. Psychol.*, vol. 12, p. 604977, 2021.
- [31] M. Wang, A. Hussein, R. F. Rojas, K. Shafi, and H. A. Abbass, "Eeg-based neural correlates of trust in human-autonomy interaction," in *IEEE Symp. Series on Comput. Intell.*, pp. 350–357, 2018.
- [32] D. Ullman and B. F. Malle, "Mdmt: Multi-dimensional measure of trust," 2019.
- [33] M. I. Ahmad and R. Refik, "'no chit chat!' a warning from a physical versus virtual robot invigilator: Which matters most?," *Front. Robot. AI*, vol. 9, p. 908013, 2022.
- [34] D. Ullman, S. Aladia, and B. F. Malle, "Challenges and opportunities for replication science in hri: A case study in human-robot trust," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, p. 110–118, 2021.
- [35] T.-C. Lin, A. Unni Krishnan, and Z. Li, "Shared autonomous interface for reducing physical effort in robot teleoperation via human motion mapping," in *IEEE Int. Conf. Robot. Autom.*, pp. 9157–9163, 2020.
- [36] M. Young, C. Miller, Y. Bi, W. Chen, and B. D. Argall, "Formalized task characterization for human-robot autonomy allocation," in *Int. Conf. Robot. Autom.*, p. 6044–6050, 2019.
- [37] R. Luo, Y. Weng, Y. Wang, P. Jayakumar, M. J. Brudnak, V. Paul, V. R. Desaraju, J. L. Stein, T. Ersal, and X. J. Yang, "A workload adaptive haptic shared control scheme for semi-autonomous driving," *Accident Analysis & Prevention*, vol. 152, p. 105968, 2021.
- [38] F. Chen, J. Zhou, Y. Wang, K. Yu, S. Z. Arshad, A. Khawaji, D. Conway, F. Chen, J. Zhou, Y. Wang, *et al.*, "Trust and cognitive load," *Robust Multimodal Cogn. Load Measure.*, pp. 195–214, 2016.
- [39] M. I. Ahmad, J. Bernotat, K. Lohan, and F. Eyssel, "Trust and cognitive load during human-robot interaction," *arXiv preprint arXiv:1909.05160*, 2019.
- [40] H. R. Preusse, G. Igwe, T. Sletten, D. K. Langlois, M. R. Fraune, and T. Fong, "'i'm not in the mode to help' interface design for robots operating at varying levels of autonomy," in *Comp. 2020 ACM/IEEE Int. Conf. Human-Robot Interaction*, pp. 594–596, 2020.
- [41] J. Ogbemhe and K. Mpofu, "Towards achieving a fully intelligent robotic arc welding: A review," *Industrial Robot: An Int. Journal*, vol. 42, no. 5, pp. 475–484, 2015.
- [42] T. Lei, Y. Rong, H. Wang, Y. Huang, and M. Li, "A review of vision-aided robotic welding," *Computers in Industry*, vol. 123, p. 103326, 2020.
- [43] Y. Chen and F. Dong, "Robot machining: Recent development and future research issues," *Int. J. Adv. Mfg. Tech.*, vol. 66, pp. 1489–1497, 2013.
- [44] S. Yuwen, J. Jinjie, X. Jintong, C. Mansen, and N. Jinbo, "Path, feedrate and trajectory planning for free-form surface machining: A state-of-the-art review," *Chinese Journal of Aeronautics*, vol. 35, no. 8, pp. 12–29, 2022.
- [45] Y. Chen, W. Chen, B. Li, G. Zhang, and W. Zhang, "Paint thickness simulation for painting robot trajectory planning: A review," *Ind. Robot: An Int. Journal*, vol. 44, no. 5, pp. 629–638, 2017.
- [46] H. R. Kam, S.-H. Lee, T. Park, and C.-H. Kim, "Rviz: A toolkit for real domain data visualization," *Telecom. Sys.*, vol. 60, pp. 337–345, 2015.
- [47] A. D. Dragan and S. S. Srinivasa, "A policy-blending formalism for shared control," *Int. J. Robot. Res.*, vol. 32, no. 7, pp. 790–805, 2013.
- [48] M. Marciano, S. Díaz, J. Pérez, and E. Irigoyen, "A review of shared control for automated vehicles: Theory and applications," *IEEE Trans. Human-Machine Syst.*, vol. 50, no. 6, pp. 475–491, 2020.
- [49] A. T. Duchowski, K. Krejtz, N. A. Gehrer, T. Bafna, and P. Bækgaard, "The low/high index of pupillary activity," in *Proc. Conf. Human Factors in Comp. Sys.*, p. 1–12, 2020.
- [50] M. Harbers, M. M. M. Peeters, and M. A. Neerincx, *Perceived Autonomy of Robots: Effects of Appearance and Context*, pp. 19–33. Cham: Springer International Publishing, 2017.
- [51] E. Roesler, S. C. Steinhäusser, B. Lugin, and L. Onnasch, "The influence of visible cables and story content on perceived autonomy in social human-robot interaction," *Robotics*, vol. 12, no. 1, p. 3, 2022.
- [52] P. Balatti, D. Kanoulas, N. Tsarakakis, and A. Ajoudani, "A method for autonomous robotic manipulation through exploratory interactions with uncertain environments," *Auton. Robot.*, vol. 44, pp. 1395–1410, 2020.
- [53] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G\*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior Res. Methods*, vol. 39, no. 2, pp. 175–191, 2007.
- [54] T. Asselborn, K. Sharma, W. Johal, and P. Dillenbourg, "Bridging multilevel time scales in hri: an analysis framework," *ACM Trans. Human-Robot Interaction*, vol. 8, no. 3, pp. 1–24, 2019.
- [55] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco, "Impact of robot failures and feedback on real-time trust," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, p. 251–258, 2013.
- [56] J. Ayoub, L. Avetisyan, X. J. Yang, and F. Zhou, "Real-time trust prediction in conditionally automated driving using physiological measures," *IEEE Trans. Intell. Transport. Sys.*, pp. 1–9, 2023.
- [57] M. R. Endsley and E. O. Kiris, "The out-of-the-loop performance problem and level of control in automation," *Human Factors*, vol. 37, no. 2, pp. 381–394, 1995.
- [58] M. Desai, M. Medvedev, M. Vázquez, S. McSheehy, S. Gadea-Omelchenko, C. Bruggeman, A. Steinfeld, and H. Yanco, "Effects of changing reliability on trust of robot systems," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, p. 73–80, 2012.
- [59] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, p. 141–148, 2015.
- [60] V. Alonso and P. De La Puente, "System transparency in shared autonomy: A mini review," *Front. Neurobot.*, vol. 12, p. 83, 2018.
- [61] L. Wang, Q. Li, J. Lam, Z. Wang, and Z. Zhang, "Intent inference in shared-control teleoperation system in consideration of user behavior," *Complex & Intell. Sys.*, pp. 1–11, 2021.
- [62] M. Selvaggio, J. Cacace, C. Pacchierotti, F. Ruggiero, and P. R. Giordano, "A shared-control teleoperation architecture for nonprehensile object transportation," *IEEE Trans. Robot.*, vol. 38, no. 1, pp. 569–583, 2021.
- [63] R. Balachandran, H. Mishra, M. Cappelli, B. Weber, C. Secchi, C. Ott, and A. Albu-Schaeffer, "Adaptive authority allocation in shared control of robots using bayesian filters," in *2020 IEEE Int. Conf. Robot. Autom. (ICRA)*, pp. 11298–11304, IEEE, 2020.
- [64] D. A. Abbink, M. Mulder, and E. R. Boer, "Haptic shared control: Smoothly shifting control authority?," *Cogn., Tech. & Work*, vol. 14, pp. 19–28, 2012.
- [65] D. Zhang, R. Tron, and R. P. Khurshid, "Haptic feedback improves human-robot agreement and user satisfaction in shared-autonomy teleoperation," in *2021 IEEE Int. Conf. Robot. Autom. (ICRA)*, p. 3306–3312, IEEE Press, 2021.
- [66] N. Thomas, A. J. Miller, H. Ayaz, and J. D. Brown, "Haptic shared control improves neural efficiency during myoelectric prosthesis use," *Scientific Reports*, vol. 13, no. 1, p. 484, 2023.
- [67] L. Becker, B. Weber, and N. Bechtel, "Haptic guidance for teleoperation: Optimizing performance and user experience," in *Int. Conf. Human Haptic Sensing and Touch Enabled Computer App.*, pp. 129–137, Springer, 2022.