TU Delft

Delft University of Technology

Fitts' List Revisited: An Empirical Study on Function Allocation in a Two-Agent Physical Human-Robot Collaborative Position/Force Task

Mol, N.; Prendergast, J.M.; Abbink, David; Peternel, L.

DOI
10.1109/LRA.2025.3632607

Publication date
2025

Document Version
Final published version

Published in
IEEE Robotics and Automation Letters

Citation (APA)
Mol, N., Prendergast, J. M., Abbink, D., & Peternel, L. (2025). Fitts' List Revisited: An Empirical Study on Function Allocation in a Two-Agent Physical Human-Robot Collaborative Position/Force Task. *IEEE Robotics and Automation Letters*, *11*(1), 202-209. https://doi.org/10.1109/LRA.2025.3632607

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Fitts' List Revisited: An Empirical Study on Function Allocation in a Two-Agent Physical Human-Robot Collaborative Position/Force Task

Nicky Mol, *Graduate Student Member, IEEE*, J. Micah Prendergast, David A. Abbink, *Senior Member, IEEE*, and Luka Peternel, *Member, IEEE*

***Abstract*—In this letter, we investigate whether classical function allocation—the principle of assigning tasks to either a human or a machine—holds for physical Human-Robot Collaboration, which is important for providing insights for Industry 5.0 to guide how to best augment rather than replace workers. This study empirically tests the applicability of Fitts' List within physical Human-Robot Collaboration, by conducting a user study (N=26, within-subject design) to evaluate four distinct allocations of position/force control between human and robot in an abstract blending task. We hypothesize that the function in which humans control the position achieves better performance and receives higher user ratings. When allocating position control to the human and force control to the robot, compared to the opposite case, we observed a significant improvement in preventing overblending. This was also perceived better in terms of physical demand and overall system acceptance, while participants experienced greater autonomy, more engagement and less frustration. An interesting insight was that the supervisory role (when the robot controls both position and force) was rated second best in terms of subjective acceptance. Another surprising insight was that if position control was delegated to the robot, the participants perceived much lower autonomy than when the force control was delegated to the robot. These findings empirically support applying Fitts' principles to static function allocation for physical collaboration, while also revealing important nuanced user experience trade-offs, particularly regarding perceived autonomy when delegating position control.**

***Index Terms*—Physical human-robot interaction, human factors and human-in-the-loop, human-centered robotics, human-robot collaboration.**

Received 7 May 2025; accepted 25 October 2025. Date of publication 14 November 2025; date of current version 19 November 2025. This article was recommended for publication by Associate Editor Arash Arami and Editor Haoyong Yu upon evaluation of the reviewers' comments. This work was supported by BrightSky Project, funded by the R&D Mobiliteitsfonds from the Netherlands Enterprise Agency (RVO) and commissioned by the Ministry of Economic Affairs and Climate Policy. *(Corresponding author: Nicky Mol.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the TU Delft Human Research Ethics Committee (HREC) under Application No. 2721, and performed in line with the Declaration of Helsinki.

Nicky Mol, J. Micah Prendergast, and Luka Peternel are with the Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: nicky.mol@tudelft.nl; j.m.prendergast@tudelft.nl; l.peternel@tudelft.nl).

David A. Abbink is with the Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands, and also with the Department of Sustainable Design Engineering, Faculty of Industrial Design Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: d.a.abbink@tudelft.nl).

This article has supplementary downloadable material available at https://doi.org/10.1109/LRA.2025.3632607, provided by the authors.

Digital Object Identifier 10.1109/LRA.2025.3632607

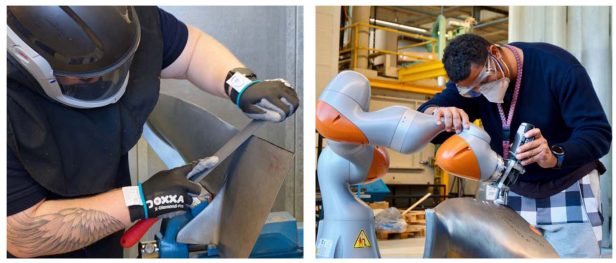

Fig. 1. Illustrative comparison of a manual blending task (left) versus a collaborative robot assisting a human worker (right) by potentially handling force or position control, based on function allocation.

I. Introduction

THE integration of robots in industrial settings has been growing rapidly in recent years, driven by the need to address complex societal challenges such as labor shortages, aging populations, and ever-increasing production demands. While traditional industrial robots have proven their worth in highly controlled and predictable, high-volume settings like automotive assembly lines, they lack the flexibility required for dynamic environments prevalent in high-mix, low-volume manufacturing or maintenance and repair. As a response, collaborative robots (cobots) have emerged, designed specifically for safe operation near human workers. This unlocks the potential for direct physical Human-Robot Collaboration (pHRC) that promises to merge the complementary skills of both agents, creating more effective, flexible, and meaningful collaborations [1]. Such vision aligns with the human-centric principles of Industry 5.0, which emphasizes technology in service of worker well-being and resilient production systems. However, despite the potential of collaborative robots, achieving truly effective, fluent, and satisfying collaboration remains a significant challenge [2].

One element of this challenge lies in *function allocation:* deciding which agent–human or robot–should perform which aspects of a shared task [3], [4], [5], [6], [7], [8]. We can model physical interactions with impedance, which deals with two key variables: motion and forces [9]. Consider a task like manual blending or polishing of a complex surface, such as an aircraft fan blade (see Fig. 1 left). This requires both precise positioning of the tool and careful application of force. Classic frameworks, notably Fitts' "MABA-MABA" (Men Are Better At / Machines Are Better At) list stemming from early human factors research [10], provide initial guidance. Despite its age and extensive criticism, this framework has shown remarkable persistence throughout

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

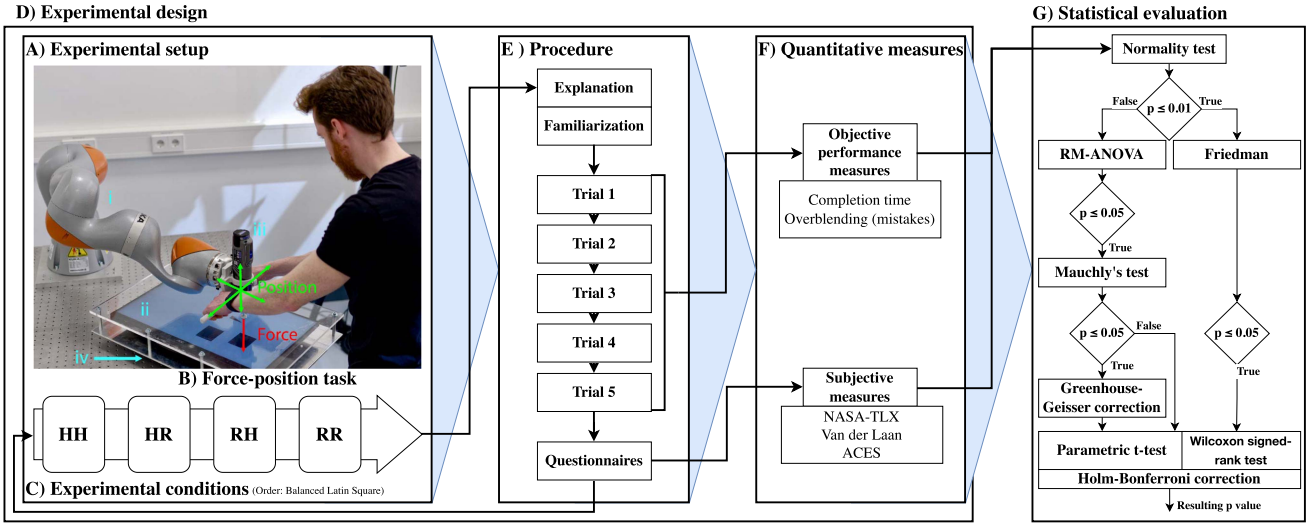

Fig. 2. Illustration of pipeline for the method. The experimental setup on the left involves a blending task with force (red) and position (green) sub-tasks. The combinations of assigning these sub-tasks between human (H) and robot (R) result in four experimental conditions. In the labeling of conditions, the first letter is for the position sub-task, while the second letter is for the force sub-task (e.g., the human doing positioning and the robot producing force would be HR). The experiment procedure is followed by the calculation of measures, which are finally used in the statistical evaluation.

the history of function allocation [11]. The framework suggests humans generally excel at tasks requiring perception, judgment, and adaptability (e.g., identifying and deciding where to polish), while machines excel at speed, consistent force application, and repetitive precision movements (e.g., producing a constant force). Modern cobots offer the technical capability to split control of force and position between the human and robot (see Fig. 1 right).

Some existing work in pHRC has exploited this ability of cobots to take human inputs during physical interaction, and allocate positions and force sub-tasks. The approach in [12] made the distribution of functions during an assembly task, where the robot performed most of the large movements, while the human stepped in for precise movements. Humans can also step in with physical interaction to give the robot occasional hints on how to re-plan longer movement trajectories in real-time during assembly, where there are multiple possible goals [13]. The method in [14] was in line with the classical framework, where the assembly of bulky objects was handled by allocating load carrying to the robot, while the precise positioning of the object during installation was allocated to the human. In [15], a collaborative polishing task was solved by allocating force control to the robot and position control to the human, consistent with Fitts' principle. In industrial tasks like sanding, letting a robot perform repetitive high-force motions can reduce human physical strain, but fully autonomous solutions often struggle with quality and adaptability [16]. Nevertheless, the existing solutions built on the general Fitts' principle without having a clear insight into whether it is valid in the context of pHRC, prompting a re-evaluation of these classic principles in the context of direct physical collaboration.

To address this gap and provide the missing insight, this study investigates a fundamental question: *who should control what in a two-agent physical human-robot collaborative position/force task?* To answer this question, we conducted human factors experiments involving a blending task requiring both force application and tool positioning. Participants experienced four distinct conditions representing the extremes of static task allocation on the shared control spectrum: HH (fully manual control), HR (human controls position, robot controls force), RH (robot controls position, human controls force), and RR (robot controls both, human supervises). By collecting and statistically analyzing both objective and subjective data across these conditions, we aimed to gain new insights into function allocation in the context of pHRC. We hypothesize that:

- H1 Van der Laan usefulness and satisfaction score higher in **HR** than in RH.
- H2 Task performance, in terms of (a) completion time and (b) overblending, score higher in **HR** than in RH.
- H3 Perceived workload is lower in **HR** than in RH.
- H4 Perceived autonomy, engagement, competence, and usefulness score higher in **HR** than in RH.

The contributions of this paper are: 1) validating whether Fitts' "MABA-MABA" principle holds for pHRC involving hybrid position/force tasks, and 2) providing new user experience insights about autonomy, competence, engagement, and usefulness in different position/force distributions.

II. Methods

The method is summarized by a schematic in Fig. 2. The experimental setup included a sensorized work surface with an interactive screen (A) for participants to be able to perform a blending task (B). There were four experimental conditions based on human-robot allocation of position and force sub-tasks (C). The procedure involved five trials per condition, and each condition included a brief explanation and familiarization before and questionnaires after the trials (E). Based on the gathered data, both objective and subjective measures were extracted (F) and fed to the statistical evaluation process (G). Each component (A-G) is described separately in detail in the remainder of this section.

A. Experimental Setup

The experimental setup used in this study is illustrated in the leftmost block in Fig. 2. It consists of a i) 7-DOF KUKA LBR iiwa 14 R820 (KUKA AG, Augsburg, Germany) robotic manipulator, equipped with a ii) custom mount with handlebars, housing the rotary tool at its end-effector for interacting with the environment. To present the task, a iii) 23.8-inch horizontally aligned Full HD Dell monitor (Dell Technologies Inc., Round Rock, Texas, US) is used. It is encased in a box with a transparent plexiglass surface, which is mechanically connected to a iv) 6-DOF SCHUNK FTN-Delta SI-330-30 Force/Torque sensor (SCHUNK GmbH & Co. KG, Lauffen/Neckar, Germany), enabling measurement of interaction forces applied to the surface of the box.

The robot is controlled at the joint-torque level using a Cartesian Impedance Controller operating at 200 Hz. The interaction force/torque in Cartesian space was defined as:

$$\boldsymbol{f} = \begin{bmatrix} \boldsymbol{K}_t & 0 \\ 0 & \boldsymbol{K}_r \end{bmatrix} (\boldsymbol{x}_a - \boldsymbol{x}_d) + \boldsymbol{D}(\dot{\boldsymbol{x}}_a - \dot{\boldsymbol{x}}_d) + \boldsymbol{S}_f \boldsymbol{f}_d, \quad (1)$$

where $\boldsymbol{K}_t \in \mathbb{R}^{3\times3}$ and $\boldsymbol{K}_r \in \mathbb{R}^{3\times3}$ are the translational and rotational parts, respectively, of the stiffness matrix $\boldsymbol{K} \in \mathbb{R}^{6\times6}$. $\boldsymbol{D} \in \mathbb{R}^{6\times6}$ represents the damping matrix, $\boldsymbol{x}_a \in \mathbb{R}^6$ and $\boldsymbol{x}_d \in \mathbb{R}^6$ being the actual and desired robot end-effector pose in Cartesian space, $\boldsymbol{S}_f \in \mathbb{R}^{6\times6}$ a diagonal matrix used for selecting the axis in which the desired force is applied and $\boldsymbol{f}_d \in \mathbb{R}^6$ the desired force, both in Cartesian space. We used a critically damped design $\boldsymbol{D} = 2\zeta\sqrt{\boldsymbol{K}}$, with damping ratio $\zeta = 0.7$. The robot and tool were gravity-compensated, while a nullspace controller enforced an "elbow up" configuration.

The trajectory generator for reference pose is defined as:

$$\boldsymbol{x}_d(t) = \text{Interp}\left(\boldsymbol{x}_{\text{start}}, \boldsymbol{x}_{\text{goal}}, \frac{t}{t_{\text{total}}}\right), \quad \text{for } t \in [0, t_{\text{total}}], \quad (2)$$

where $\boldsymbol{x}_d(t)$ is the desired end-effector pose at time t, Interp() is a time-based interpolation function (linear for position and spherical for orientation) taking the start pose ($\boldsymbol{x}_{\text{start}}$), goal pose ($\boldsymbol{x}_{\text{goal}}$) and the time t divided by the total execution time (t_{total}), which calculated as follows: $t_{\text{total}} = \max(\frac{\|\boldsymbol{x}_{\text{goal}} - \boldsymbol{x}_{\text{start}}\|}{\boldsymbol{v}_{\text{d}}})$, with $\boldsymbol{v}_{\text{d}}$ being the desired velocity.

B. Position/Force Task

Participants were tasked with removing damage from a surface, visualized as black rectangles of varying sizes, as shown in the leftmost block in Fig. 2. The task can roughly be divided into two sub-tasks: 1) positioning of the tooltip, being held by the robot end-effector, over the damage to be removed and 2) applying a downward force perpendicular to the surface of the monitor to remove the damage. The removal rate is directly related to the applied force (within thresholds) and can be expressed as a function of the applied force:

$$\frac{\partial h}{\partial t}(f_{\text{z}}) = \begin{cases} 0, & \text{if } f_{\text{z}} < f_{\min} \\ \frac{h_{\max}}{t_{\min}} \cdot \frac{f_{\text{z}} - f_{\min}}{f_{\max} - f_{\min}}, & \text{if } f_{\min} \le f_{\text{z}} \le f_{\max} \\ \frac{h_{\max}}{f_{\min}}, & \text{if } f_{\text{z}} > f_{\max} \end{cases}, \quad (3)$$

TABLE I
CONDITIONS AND THEIR DIFFERENCES IN TERMS OF CONTROL

Condition	Position	Force	$\boldsymbol{K}_t$	$f_{\text{d,z}}$
1: HH	Human	Human	$\boldsymbol{K}_t = \boldsymbol{0}$	$f_{\text{d,z}} = 0$
2: HR	Human	Robot	$\boldsymbol{K}_t = \boldsymbol{0}$	$f_{\text{d,z}} = f_{\max}$
3: RH	Robot	Human	$\boldsymbol{K}_t \neq \boldsymbol{0}$	$f_{\text{d,z}} = 0$
4: RR	Robot	Robot	$\boldsymbol{K}_t \neq \boldsymbol{0}$	$f_{\text{d,z}} = f_{\max}$

where $\frac{\text{d}h}{\text{d}t}$ is the removal rate, defined as the amount of health points removed per second. f_{z} is the force applied by the participant along the z-axis perpendicular to the surface as measured by the force/torque sensor, $f_{\min}$ and $f_{\max}$ are the minimum and maximum force thresholds. Forces below $f_{\min}$ do not contribute to removal, while forces above $f_{\max}$ saturate the removal rate. $h_{\max}$ is the maximum amount of health points per pixel, and $t_{\min}$ is defined to be the minimum time required to completely remove a pixel with $h_{\max}$ when applying the maximum effective force $f_{\max}$.

Each trial involved the removal of three rectangular damage blocks consisting of pixels with an initial amount of health points. As the task progressed, damage opacity decreased proportionally to the remaining health value, providing continuous visual feedback. A rectangular block was considered successfully removed when at least 95% of its initial health had been removed, at which point the block turns green and fades from view, indicating its completion. When participants applied force to an area where no damage (or no remaining damage) was present, they introduced new damage (called overblending) to which the same force-dependent removal model (3) was applied and was visually indicated by an increase in opacity and a color shift from black to red of the affected pixels.

C. Experimental Conditions

The study included four experimental conditions that represent all combinations of extremes on the shared control spectrum for both the position and force sub-tasks. The conditions are summarized in Table I, where also the differences in terms of control with respect to (1) are specified.

Condition 1 - HH: Participants control both the position of the tool as well as the application of force.

Condition 2 - HR: Participants control the position of the tool while the robot applies force.

Condition 3 - RH: Robot controls the position of the tool while participants apply force.

Condition 4 - RR: Robot controls both the position of the tool as well as the application of force, while participants supervise and intervene in case the robot makes mistakes.

D. Experimental Design

In a within-subject design, 26 participants (aged 22–42, 4 female, 22 male, physically healthy) completed the task under all four conditions. The study was approved by the TU Delft Human Research Ethics Committee. All participants gave written informed consent prior to their participation. To mitigate order effects, a Balanced Latin Square design was applied on the order in which the conditions were presented to the participants. Each trial involved removing three damage spots under consistent layout constraints to control task difficulty across conditions.

The four conditions were designed to isolate the effects of delegating position and/or force control: a manual baseline (HH), an automated supervisory condition (RR), and two "mixed-task" conditions (HR and RH). RR was tuned to operate near the performance limits of what is physically possible for a human to achieve. All conditions used the same Cartesian impedance controller in (1), with condition-specific translational stiffness settings. For HH and HR, translational elements of the stiffness matrix $\boldsymbol{K}$ were set to 0N/m in all directions, to allow for unhindered human positioning. For RH, it was tuned to 5000N/m in x and y and for RR in all translational directions. Rotational stiffness was fixed at 100 Nm/rad in all conditions. These values were determined through pilot testing to be high enough to rigidly guide the user's motion along the robot's path without causing instability. Force thresholds were set at $f_{\text{min}} = 5$ N and $f_{\text{max}} = 40$ N. Pixel damage removal time at f_{max} was $t_{\text{min}} = 2$ s, with $h_{\text{max}} = 100$ health points per pixel. Trajectory velocities (RH/RR) were tuned to 0.0083 m/s (blending), 0.0400 m/s ((un)latching), and 0.0800 m/s (travel).

In HR, the robot applied a force capped at f_{max} whenever the tooltip was within 1 mm of the surface and stopped when it moved above or below 3 mm, creating an intuitive latching effect consistent across participants. In RR, the robot autonomously controlled both position and force by following a predefined adaptive trajectory that systematically covered the damage spot, divided into sections of equal width. If insufficient removal was detected in a section, the robot would revisit it. To make the task engaging for the human in the supervisory role, at fixed points, automation errors were introduced by deviating the trajectory outside the damaged area, leading to overblending unless the participant intervened promptly by pressing a button. This condition modeled supervisory control, acknowledging that full autonomy is infeasible in most complex practical tasks [17].

E. Procedure

Participants first received a general explanation of the experiment and task, where they were instructed to remove three damage spots as quickly as possible while minimizing overblending. Before each condition, they were briefed on its specifics and completed a 2-minute familiarization period. They then performed five trials, where damage spots were always the same size and remained at the same pairwise distance from each other, but were rotated, translated, or flipped. After the trials, the participants were asked to fill out three questionnaires.

F. Quantitative Measures

The objective performance metrics that were considered are the time of completion, which is defined as the amount of time it takes the participant to remove the damage during a trial, and overblended health defined as the amount of cumulative health points resulting from applying force at locations where there is no damage (any longer). Participants' perceived acceptance was assessed using the Van der Laan usefulness and satisfying scales [18], while perceived workload was assessed using the NASA-TLX questionnaire [19].

To probe the psychological impact of pHRC beyond standard metrics [20], we developed the exploratory Agentic Collaboration and Engagement Scale (ACES) because existing scales from workplace psychology are ill-suited for a single-session, abstracted experiment and lack a human-robot role perspective. The scale's theoretical foundation draws from self-determination theory (SDT), which identifies **autonomy** and **competence** as basic psychological needs that foster intrinsic motivation and **engagement** [21]. These factors, together with a sense of **usefulness**, are drivers of meaningful work, an experience that is known to enhance job satisfaction and well-being [22]. To assess reliability, Cronbach's alpha (α) was calculated, which showed an internal consistency of $\alpha = 0.73$, 95% CI [0.63, 0.80] for the "Human" scale (measuring perceived meaningful involvement) and $\alpha = 0.79$, 95% CI [0.71, 0.85] for the "Robot" scale (measuring perceived quality of the robot teammate). ACES uses a dual-perspective format, rating each construct for the participant's experience ("I") and their perception of the robot ("the robot") on a Likert scale using four statements:

- I felt that XX had a high level of autonomy while completing the task.
- I felt that XX had the necessary skills and knowledge to complete the part of the task.
- I felt XX was useful in performing its part of the task.
- I felt XX was engaged in the role.

G. Statistical Evaluation

As shown in Fig. 2 G), data normality was assessed using the Shapiro–Wilk test. If normality was met ($p > 0.01$, including minor violations), a Repeated Measures ANOVA was used. Violations of sphericity (Mauchly's test) were corrected with Greenhouse–Geisser. Significant main effects were followed by paired t-tests with Holm–Bonferroni correction. Effect size was reported as η_p^2. If normality was majorly violated ($p \leq 0.01$), a Friedman test was applied. Post-hoc Wilcoxon signed-rank tests with Holm–Bonferroni correction were used for pairwise comparisons, with effect size reported using Kendall's W.

III. Results

Here, the experimental results are presented, highlighting the differences across the four conditions (HH, HR, RH, RR). Findings are structured by measure type, referencing Figures 3–6 and the detailed statistical evaluation in Table II.

A. Van Der Laan System Acceptance

Acceptance was evaluated using the Van der Laan questionnaire. Figure 3 shows a scatter plot of the results, displaying individual participant ratings along with the mean scores and standard deviation error bars for each condition, annotated with statistically significant pairwise comparisons.

Significant differences were found across conditions for perceived usefulness ($\text{F}(2.7, 66.7) = 12.39$, $p < 0.0001$) and satisfaction ($\text{F}(2.3, 56.9) = 13.55$, $p < 0.0001$). Pairwise comparisons revealed a strong preference for HR over RH, with HR rated significantly higher on both usefulness (t-test, $p = 0.0002$) and satisfaction (t-test, $p < 0.0001$). HR was also rated significantly more useful (t-test, $p < 0.0001$) and satisfying (t-test, $p = 0.0028$) than the baseline HH condition. Conversely, RH showed no usefulness benefit over HH (t-test, $p = 0.2810$) and was significantly less satisfying (t-test, $p = 0.0029$). Notably, the supervisory RR condition was rated significantly higher than RH on usefulness (t-test, $p = 0.0086$) and satisfaction (t-test, $p = 0.0009$), positioning it second in user acceptance, statistically comparable to HH but generally lower than HR.

TABLE II
REPORTED DESCRIPTIVE AND INFERENTIAL STATISTICS OF THE QUALITATIVE MEASURES

Measure		HH	HR	RH	RR	Main Effect	Effect Size	Piece-wise Comparison[b]
Van der Laan[a]								
Usefulness	M	0.18	1.04	-0.01	0.66	F(2.7,66.7)=12.39	η_p^2=0.331	1: **p<0.0001** 2: p=0.2810 3: p=0.1038
(-2, 2)↑	SD	0.76	0.69	0.90	0.84	**p<0.0001**†		4: **p=0.0002** 5: p=0.1038 6: **p=0.0086**
Satisfying score	M	0.20	0.94	-0.62	0.52	F(2.3,56.9)=13.55	η_p^2=0.351	1: **p=0.0028** 2: **p=0.0029** 3: p=0.3371
(-2, 2) ↑	SD	0.91	0.81	0.98	1.13	**p<0.0001**†		4: **p<0.0001** 5: p=0.2138 6: **p=0.0009**
NASA-TLX								
Mental demand	Mdn	60.00	29.00	37.50	28.00	F(3,75)=1.19	η_p^2=0.045	1: p=0.3017 2: p=1.0000 3: p=1.0000
(0, 100) ↓	IQR	22.00-74.00	20.00-56.75	15.50-65.75	15.00-73.00	p=0.3207		4: p=1.0000 5: p=1.0000 6: p=1.0000
Physical demand	Mdn	66.50	35.00	65.50	3.00	F(2.3,57.6)=63.35	η_p^2=0.717	1: **p<0.0001** 2: p=0.2603 3: **p<0.0001**
(0, 100) ↓	IQR	61.25-73.75	21.25-47.75	40.50-76.75	0.00-5.75	**p<0.0001**†		4: **p=0.0039** 5: **p<0.0001** 6: **p<0.0001**
Temporal demand	Mdn	65.00	51.50	28.50	12.00	F(2.0,49.5)=24.41	η_p^2=0.494	1: **p=0.0440** 2: **p=0.0019** 3: **p<0.0001**
(0, 100) ↓	IQR	43.50-73.75	27.25-65.75	14.50-54.50	5.00-25.00	**p<0.0001**†		4: p=0.0953 5: **p<0.0001** 6: **p=0.0019**
Performance	Mdn	18.00	18.50	25.50	11.00	χ^2(3)=13.04	W=0.167	1: p=0.7878 2: p=0.1031 3: p=0.1903
(0, 100) ↓	IQR	9.25-25.75	10.25-31.50	19.25-54.00	5.00-15.75	**p=0.0046**°°		4: p=0.1779 5: p=0.1031 6: **p=0.0018**
Effort	Mdn	68.50	50.50	61.00	13.00	F(3,75)=24.17	η_p^2=0.492	1: **p=0.0010** 2: **p=0.0136** 3: **p<0.0001**
(0, 100) ↓	IQR	62.00-78.25	26.50-65.75	37.00-73.50	10.00-33.25	**p<0.0001**°		4: p=0.3790 5: **p=0.0004** 6: **p<0.0001**
Frustration	Mdn	24.00	29.00	66.00	10.00	F(3,75)=12.41	η_p^2=0.332	1: p=0.8882 2: **p<0.0001** 3: p=0.8882
(0, 100) ↓	IQR	10.00-30.00	11.75-46.25	28.75-74.75	2.75-35.50	**p<0.0001**		4: **p=0.0040** 5: p=0.8882 6: **p=0.0002**
Performance metrics								
Completion time	Mdn	109.99	112.87	109.33	90.72	χ^2(3)=49.71	W=0.637	1: p=0.0528 2: p=0.4992 3: **p<0.0001**
(s) ↓	IQR	103.88-115.77	105.70-119.33	103.12-111.07	90.09-92.01	**p<0.0001**°°		4: p=0.1261 5: **p<0.0001** 6: **p<0.0001**
Overblended health	Mdn	0	21	12368	5269	χ^2(3)=16.34	W=0.209	1: p=1.0000 2: p=0.0705 3: p=1.0000
(health points) ↓	IQR	0-11866	0-9147	3467-36513	3993-8156	**p=0.0010**°°		4: **p=0.0397** 5: p=1.0000 6: **p<0.0001**
ACES - Human								
Autonomy	Mdn	20.50	18.50	6.50	4.50	χ^2(3)=64.78	W=0.831	1: **p=0.0034** 2: **p<0.0001** 3: **p<0.0001**
(1, 21) ↑	IQR	19.00-21.00	15.00-20.00	5.00-7.75	3.00-7.75	**p<0.0001**°°		4: **p<0.0001** 5: **p<0.0001** 6: p=0.6250
Competence	Mdn	18.50	18.00	16.00	19.00	χ^2(3)=17.83	W=0.229	1: p=0.8262 2: p=0.0704 3: p=0.7311
(1, 21) ↑	IQR	16.25-20.00	15.25-20.00	11.00-19.00	17.00-21.00	**p=0.0005**°°		4: p=0.1076 5: p=0.3049 6: **p=0.0012**
Usefulness	Mdn	19.00	18.00	9.00	9.50	F(2.1,52.8)=34.63	η_p^2=0.581	1: p=0.2095 2: **p<0.0001** 3: **p<0.0001**
(1, 21) ↑	IQR	18.00-20.00	16.00-20.00	6.25-15.00	7.00-14.00	**p<0.0001**†		4: **p<0.0001** 5: **p<0.0001** 6: p=0.7471
Engagement	Mdn	18.50	16.00	13.00	12.00	F(2.1,53.0)=15.65	η_p^2=0.385	1: **p=0.0044** 2: **p<0.0001** 3: **p<0.0001**
(1, 21) ↑	IQR	15.50-20.00	13.50-19.00	7.00-16.50	5.50-14.00	**p<0.0001**†		4: **p=0.0122** 5: **p=0.0010** 6: p=0.3431
ACES - Robot								
Autonomy	Mdn	3.00	8.00	15.50	20.00	χ^2(3)=60.75	W=0.779	1: **p<0.0001** 2: **p<0.0001** 3: **p<0.0001**
(1, 21) ↑	IQR	2.00-4.00	5.00-11.75	13.00-18.00	18.00-21.00	**p<0.0001**°°		4: **p=0.0035** 5: **p<0.0001** 6: **p=0.0027**
Competence	Mdn	18.50	17.50	16.00	18.00	F(2.0,51.2)=2.27	η_p^2=0.083	1: p=0.5993 2: p=0.7554 3: p=0.5074
(1, 21) ↑	IQR	11.00-20.00	14.00-19.00	13.25-18.00	16.00-20.00	p=0.1127°†		4: p=0.4529 5: p=0.5074 6: p=0.1117
Usefulness	Mdn	8.50	17.00	16.00	19.00	F(2.1,52.0)=15.98	η_p^2=0.390	1: **p=0.0010** 2: **p=0.0155** 3: **p<0.0001**
(1, 21) ↑	IQR	4.00-15.50	14.25-19.00	12.25-18.75	18.00-20.00	**p<0.0001**°†		4: p=0.1807 5: **p=0.0135** 6: **p=0.0078**
Engagement	Mdn	3.00	8.00	15.50	20.00	F(2.0,50.6)=25.10	η_p^2=0.501	1: **p=0.0001** 2: **p=0.0003** 3: **p<0.0001**
(1, 21) ↑	IQR	2.00-4.00	5.00-11.75	13.00-18.00	18.00-21.00	**p<0.0001**†		4: p=0.1031 5: **p=0.0002** 6: **p=0.0229**

↑: Higher values are better, ↓: lower values are better Color scale (normalized for each metric): Worst — Best
[a] Van der Laan items reported as Mean (M) ± SD. All other measures reported as Median (Mdn) and IQR.
[b] 1: HH-HR, 2: HH-RH, 3: HH-RR: 4: HR-RH, 5: HR-RR, 6: RH-RR
° Minor violation of normality (Shapiro-Wilk test). No action taken since RM-ANOVA is robust against minor violations due to the Central Limit Theorem.
°° Non-parametric test used (Friedman) following major violation of normality (Shapiro-Wilk test).
† Greenhouse-Geisser correction applied following violation of sphericity (Mauchly's test).

B. NASA-TLX Workload

Workload was evaluated using the NASA-TLX questionnaire. Box-plots showing median and interquartile range of the scores for each condition per item are visualized in Fig. 4, annotated with statistically significant pairwise comparisons.

Significant effects were found for all workload dimensions except mental demand ($F(3, 75) = 1.19$, $p = 0.3207$) (Fig. 4). Physical demand showed the largest differences ($F(2.3, 57.6) = 63.35$, $p < 0.0001$), with RR being significantly lowest (t-test, $p < 0.0001$) compared to all other conditions. HR significantly reduced physical demand compared to both HH (t-test, $p < 0.0001$) and RH (t-test, $p = 0.0039$). Effort followed similar patterns, being lowest in RR. Frustration was uniquely high in the RH condition, significantly higher than HH (t-test, $p < 0.0001$), HR (t-test, $p = 0.0040$), and RR (t-test, $p = 0.0002$). Although mental demand showed no significant main effect, variability within conditions was high (see SDs in Table II).

C. Performance Metrics

Performance was evaluated using two objective metrics: task completion time and amount of overblending as a measure of mistakes made. Box-plots showing median and interquartile range of the scores for each condition are visualized in Fig. 5, annotated with statistically significant pairwise comparisons.

Both completion time ($\chi^2(3) = 49.71$, $p < 0.0001$) and overblended health ($\chi^2(3) = 16.34$, $p = 0.0010$) showed significant differences. RR yielded significantly faster completion times than all other conditions (Wilcoxon signed-rank, $p < 0.0001$), while overblended health was significantly higher in RH compared to HR (Wilcoxon signed-rank, $p = 0.0397$) and RR (Wilcoxon signed-rank, $p < 0.0001$).

D. Agentic Collaboration and Engagement Scale (ACES)

The subjective quality of the collaboration in terms of autonomy, competence, usefulness, and engagement was evaluated

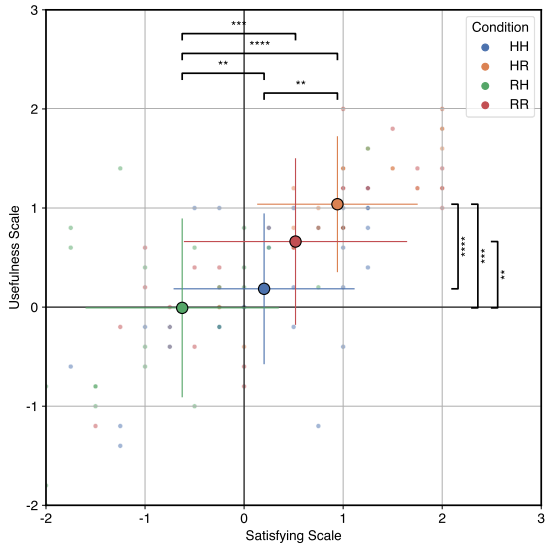

Fig. 3. Plot showing subjective scores on the Van der Laan acceptance scales: usefulness (yaxis) versus satisfying (xaxis). Large circles with error bars show the mean ± SD for each condition, while small semitransparent dots are the corresponding singleparticipant scores. Points that lie higher and further right indicate greater perceived acceptance. Statistically significant pairwise differences are annotated on each axis, where: **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

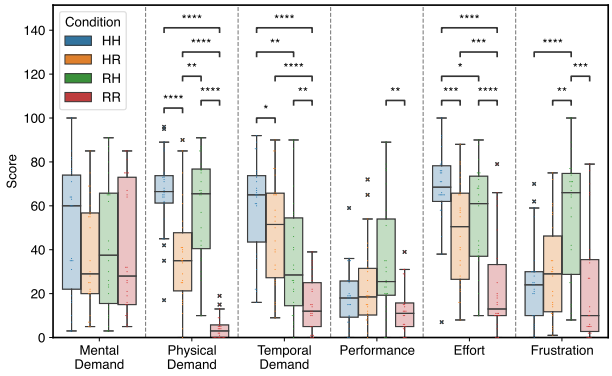

Fig. 4. Box plots showing NASATLX scores. Each swimlane shows the results of one workload subscale for each of the conditions. Boxes show the median and the interquartile range, where whiskers extend to the 5th and 95th percentiles, while small dots are the individual participant scores. Higher scores correspond to greater perceived workload. Statistically significant pairwise differences are annotated, where: *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

using ACES based on a Likert scale and was rated on how the participants perceived themselves and how they perceived the robot. Median and interquartile ranges per condition are shown in Fig. 6.

Significant differences were observed for participants' ratings of their own autonomy ($\chi^2(3)$=64.78, $p < 0.0001$), competence ($\chi^2(3)$=17.83, $p = 0.0005$, usefulness ($F(2.1, 52.8) = 34.63, p < 0.0001$), and engagement ($F(2.1, 53.0) = 15.65, p < 0.0001$). Ratings for autonomy, usefulness, and engagement were highest when participants controlled position (HH and HR). Specifically, HR yielded significantly higher ratings compared HR on autonomy (Wilcoxon signed-rank, $p < 0.0001$), usefulness (t-test,

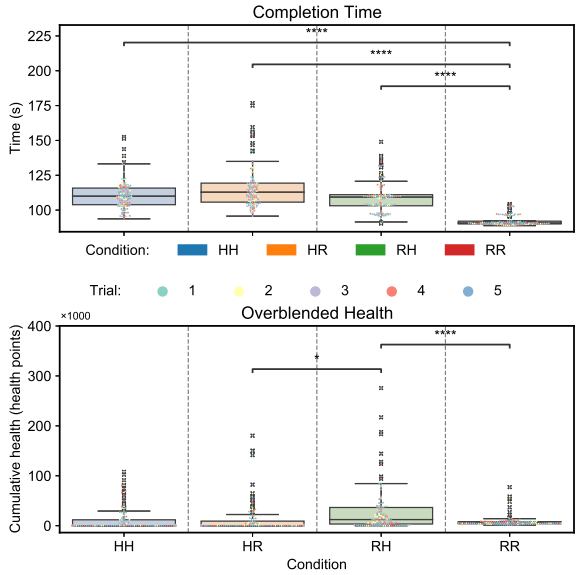

Fig. 5. Boxplots showing objective performance metrics in terms of completion time in seconds (top) and cumulative overblended health in health points (bottom). Boxes show the median and the interquartile range, where whiskers extend to the 5th and 95th percentiles, while each small dot is one trial and is colored according to the trial number. Lower values correspond to better performance. Statistically significant pairwise differences are annotated, where: *: $0.01 < p \leq 0.05$; ****: $p \leq 0.0001$.

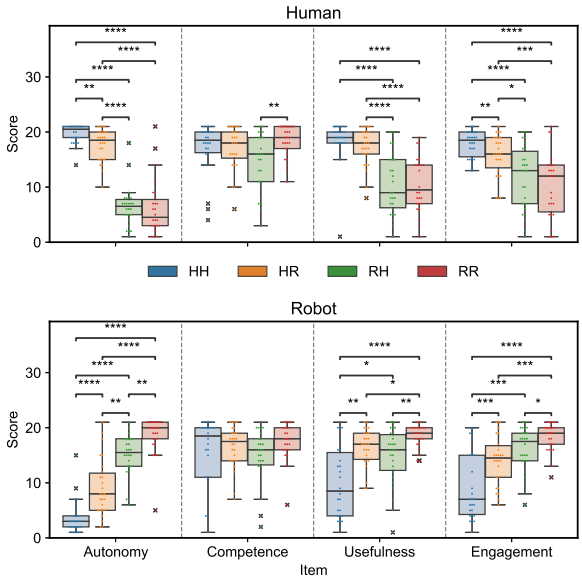

Fig. 6. Boxplots showing how participants rated themselves (top) and how they rated the robot (bottom) on four dimensions: autonomy, competence, engagement, and usefulness. For every condition, boxes show the median and the interquartile range, where whiskers mark the 5th–95th percentiles, while small dots show individual ratings. Statistically significant pairwise differences are annotated, where: *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

$p < 0.0001$), and engagement (t-test, $p = 0.0122$). Delegating position reduced perceived autonomy significantly more (HH vs RH: Wilcoxon signed-rank, $p < 0.0001$) than delegating force (HH vs HR: Wilcoxon signed-rank, $p = 0.0034$). Self-rated competence was lowest in RH, significantly lower than in RR (Wilcoxon signed-rank, $p = 0.0012$).

Significant differences were also observed for ratings of the robot's autonomy ($\chi^2(3)$=60.75, $p < 0.0001$), usefulness ($F(2.1, 52.0) = 15.98, p < 0.0001$), and engagement ($F(2.0, 50.6) = 25.10, p < 0.0001$), except for competence ($F(2.0, 51.2) = 2.27, p = 0.1127$). Robot autonomy was perceived highest when it controlled position (RH and RR). The robot in the RR condition was rated significantly most useful and engaging compared to all other conditions.

IV. Discussion

This study investigated the relevance of Fitts' classical "Men Are Better At / Machines Are Better At" (MABA-MABA) framework [10] in the context of two-agent pHRC. The central question aimed to identify the optimal static allocation of position and force control between a human and a robot during physical collaboration. The findings presented here provide compelling evidence suggesting that the fundamental principles articulated by Fitts over seven decades ago retain significant relevance for designing effective collaboration strategies in modern pHRC scenarios involving direct physical interaction and shared control. Beyond this validation, the results also yield additional insights into user experience.

Validating Fitts. HR outperforms RH: The direct comparison between the HR and RH conditions consistently favored HR across multiple dimensions. Participants reported higher acceptance (supporting H1, Fig. 3), and demonstrated significantly less overblending, while completion time was found to be similar (largely supporting H2, Fig. 5). This equivalence in completion time likely reflects a balance of two different inefficiencies: the time lost to the human's imperfect exploratory path in HR was matched by the time the robot in RH spent on corrective re-blending due to the human's imperfect force application for this specific task. Participants also experienced significantly less frustration and lower physical demand (largely supporting H3, Fig. 4), and experienced greater autonomy, felt more useful, and engaged (strongly supporting H4, Fig. 6) when they controlled positioning while the robot controlled the force. These findings align well with the "MABA-MABA" principle: leveraging human dexterity and adaptability for position control and delegating force to the machine leads to a more effective and preferred collaboration.

RH leads to increased overblending and frustration: The difficulties encountered in the RH condition, notably increased overblending and elevated frustration, suggest a misalignment. This could stem from the robot dictating the position trajectory, potentially introducing unpredictability, or a path that conflicts with the human's natural movement tendencies or intentions. This most likely forces the human operator to be reactive, attempting to synchronize their force output with a robot-imposed motion profile. A dip in self-rated competence (although not found to be statistically significant, see Fig. 6) could result from a loss in confidence due to this misalignment.

HR provides better experience compared to HH: Interestingly, the HR condition was not only preferred over the inverse task allocation (RH), but was also rated as significantly more useful and satisfying than the HH baseline (Fig. 3), suggesting the Fitts-aligned collaboration (HR) provides a genuine enhancement over unassisted work. This is likely due to the substantial reduction in physical demand (Fig. 4), without compromising the user's sense of agency over the task's positional execution or a large drop in perceived autonomy (see HH vs HR and RH in Fig. 6).

Conversely, when comparing RH to HH, significantly lower satisfaction ratings were possibly due to a significant increase in frustration. Usefulness was rated similarly low, perhaps due to perceived higher physical demand. The participants rated themselves significantly less useful in the collaboration and felt less engaged. The latter aspect is interesting since one would expect that a human having to be more reactive to a robot would increase the engagement.

It is important to note that neither HR nor RH improved objective task performance compared to HH (Fig. 5). The absence of a performance increase could imply that robot assistance is superfluous. Nevertheless, pHRC has many other benefits that do not directly affect the performance. For example, pHRC improves ergonomics in physically demanding tasks, such as blending, to prevent work-related musculoskeletal injuries [15].

Supervisory control and the ironies of automation: The RR condition presented a unique profile characterized by trade-offs. Objectively, RR showed the best completion times and the lowest cumulative overblended health, as it was designed to do (Fig. 5). It also resulted in the lowest perceived workload across almost all NASA-TLX dimensions (Fig. 4). However, these efficiency gains came at a cost: participants reported the lowest levels of perceived autonomy, usefulness, and engagement (Fig. 6). This resonates with Bainbridge's classic "ironies of automation" [23]: optimizing system performance can paradoxically diminish human experience by reducing meaningful involvement.

Nevertheless, our findings, specific to pHRC, also partially contradict that classical notion since RR achieved surprisingly favorable acceptance ratings, ranking second overall, statistically comparable to HH, though still lower than HR (Fig. 3). One explanation is the substantial reduction in physical and temporal demands was highly valued by participants. Moreover, the supervisory role, requiring vigilance and occasional intervention, likely provided more engagement (perceived similarly to RH, see Fig. 6) than purely passive oversight, mitigating the worst effects of automation. Participants also rated the robot in RR as more useful and engaging than in HH and RH (Fig. 6), suggesting they recognized its effectiveness even while acknowledging their own reduced role.

Finally, this study assessed short-term task performance. In industrial settings, prolonged supervision may shift the trade-off between efficiency and meaningful involvement, potentially reducing acceptance over time.

The primacy of position control for experienced autonomy: Results reveal a pronounced asymmetry in how delegation affected participants' sense of autonomy (Fig. 6). Delegating position control to the robot (HH vs RH) resulted in a substantially larger decrease in reported autonomy compared to relinquishing force control (HH vs HR), suggesting that position control is more intrinsically linked to the operator's sense of agency than force control.

Position control determines *where* the interaction occurs and directs the tool's action towards achieving the task goal, whereas force control *modulates* the intensity of the interaction at a given location. Participants may perceive the goal-directed aspect as the primary locus of control in executing the physical task. Relinquishing positional authority may thus feel akin to losing command over task progression, downgrading the human contribution to one of reactive effort modulation.

This drop in autonomy when delegating position control aligns well with Self-Determination Theory (SDT) in a pHRC context. SDT posits that autonomy is a fundamental psychological need, and its frustration thwarts motivation and well-being, which can lead to negative outcomes such as burnout. Our findings empirically mirror this principle: RH recorded the lowest levels of perceived autonomy (shared with RR) and also yielded the highest levels of frustration and the lowest satisfaction scores. This suggests that for participants, delegating position control was not just a change in task mechanics but the removal of a psychological need, which resulted in a negative experience.

Introduction of robot does not affect mental demand: The non-significant difference in mental demand suggests a comparable cognitive load, albeit of a different nature: HH required focused motor control for both sub-tasks, HR required coordination with the robot producing force, RH demanded a more reactive vigilance to synchronize with the robot's movement, and RR necessitated supervisory vigilance. This suggests that introducing a robot in any function does not make the task largely more mentally demanding, which can be a valuable insight for workplaces considering pHRC solutions.

Limitations: Several limitations warrant consideration. The predominantly male participant pool (4 female, 22 male) restricts generalisability, and the abstracted lab task may not fully transfer to complex industrial settings where optimal task allocation could differ. Furthermore, our evaluation was based on short-term interaction; long-term use, especially in the supervisory (RR) condition, could alter user experience due to factors like sustained vigilance. Finally, as the new ACES questionnaire lacks full psychometric validation, its findings should be interpreted with caution.

ACKNOWLEDGMENT

The authors would like to thank Alexis Derumigny for advice on the statistical evaluation.

REFERENCES

[1] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schäffer, K. Kosuge, and O. Khatib, "Progress and prospects of the human–robot collaboration," *Auton. Robots*, vol. 42, no. 5, pp. 957–975, Jun. 2018.

[2] J. E. Michaelis, A. Siebert-Evenstone, D. W. Shaffer, and B. Mutlu, "Collaborative or simply uncaged? Understanding human-cobot interactions in automation," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–12.

[3] A. Mörtl, M. Lawitzky, A. Kucukyilmaz, M. Sezgin, C. Basdogan, and S. Hirche, "The role of roles: Physical cooperation between humans and robots," *Int. J. Robot. Res.*, vol. 31, no. 13, pp. 1656–1674, Nov. 2012.

[4] N. Jarrassé, V. Sanguineti, and E. Burdet, "Slaves no longer: Review on role assignment for human–robot joint motor action," *Adaptive Behav.*, vol. 22, no. 1, pp. 70–82, Feb. 2014.

[5] I. El Makrini, M. Omidi, F. Fusaro, E. Lamon, A. Ajoudani, and B. Vandcrborght, "A hierarchical finite-state machine-based task allocation framework for human-robot collaborative assembly tasks," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2022, pp. 10238–10244.

[6] M. Selvaggio, M. Cognetti, S. Nikolaidis, S. Ivaldi, and B. Siciliano, "Autonomy in physical human-robot interaction: A brief survey," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 7989–7996, Oct. 2021.

[7] E. Merlo et al., "An ergonomic role allocation framework for dynamic human–robot collaborative tasks," *J. Manuf. Syst.*, vol. 67, pp. 111–121, Apr. 2023.

[8] L. Vianello, S. Ivaldi, A. Aubry, and L. Peternel, "The effects of role transitions and adaptation in human–cobot collaboration," *J. Intell. Manuf.*, vol. 35, no. 5, pp. 2005–2019, Jun. 2024.

[9] N. Hogan, "Impedance control: An approach to manipulation," in *Proc. Amer. Control Conf.*, Jun. 1984, pp. 304–313.

[10] P. M. Fitts, Ed., *Human Engineering for an Effective Air-Navigation and Traffic-Control System*, Washington, DC, USA: Nat. Res. Council, Mar. 1951.

[11] J. C. F. de Winter and D. Dodou, "Why the Fitts list has persisted throughout the history of function allocation," *Cogn. Technol. Work*, vol. 16, no. 1, pp. 1–11, Feb. 2014.

[12] A. Cherubini, R. Passama, A. Crosnier, A. Lasnier, and P. Fraisse, "Collaborative manufacturing with physical human–robot interaction," *Robot. Comput.-Integr. Manuf.*, vol. 40, pp. 1–13, Aug. 2016.

[13] K. Haninger, C. Hegeler, and L. Peternel, "Model predictive impedance control with Gaussian processes for human and environment interaction," *Robot. Auton. Syst.*, vol. 165, Jul. 2023, Art. no. 104431.

[14] L. Roveda et al., "Human-robot cooperative interaction control for the installation of heavy and bulky components," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, Oct. 2018, pp. 339–344.

[15] L. Peternel, N. Tsagarakis, D. Caldwell, and A. Ajoudani, "Robot adaptation to human physical fatigue in human–robot co-manipulation," *Auton. Robots*, vol. 42, no. 5, pp. 1011–1021, Jun. 2018.

[16] A. Konstant et al., "Human–robot collaboration with a corrective shared controlled robot in a sanding task," *Hum. Factors*, vol. 67, no. 3, pp. 246–263, Mar. 2025.

[17] J. M. Bradshaw, R. R. Hoffman, D. D. Woods, and M. Johnson, "The seven deadly myths of 'autonomous systems'," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 54–61, May-Jun. 2013.

[18] J. D. Van Der Laan, A. Heino, and D. De Waard, "A simple procedure for the assessment of acceptance of advanced transport telematics," *Transp. Res. C., Emerg. Technol.*, vol. 5, no. 1, pp. 1–10, Feb. 1997.

[19] S. G. Hart, "Nasa-task load index (NASA-TLX); 20 years later," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, Oct. 2006, vol. 50, no. 9, pp. 904–908.

[20] S. Fletcher, I. Eimontaite, P. Webb, and N. Lohse, "'We don't need ergonomics anymore, we need psychology!' – The human analysis needed for human-robot collaboration," in *Proc. Int. Conf. Appl. Hum. Factors Ergonom.*, 2023, vol. 80, pp. 189–199.

[21] M. Gagné et al., "Understanding and shaping the future of work with self-determination theory," *Nature Rev. Psychol.*, vol. 1, no. 7, pp. 378–392, Jul. 2022.

[22] J. Smids, S. Nyholm, and H. Berkers, "Robots in the workplace: A threat to—or opportunity for—meaningful work?," *Philosophy Technol.*, vol. 33, no. 3, pp. 503–522, Sep. 2020.

[23] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, Nov. 1983.