ORIGINAL ARTICLE

The virtualization of human–robot interactions: a user-centric workload assessment

Federica Nenna[1,2] · Valeria Orso[1,2] · Davide Zanardi[1] · Luciano Gamberini[1,2]

Abstract

Interest in the virtualization of human–robot interactions is increasing, yet the impact that collaborating with either virtual or physical robots has on the human operator's mental state is still insufficiently studied. In the present work, we aimed to fill this gap by conducting a systematic assessment of a human–robot collaborative framework from a user-centric perspective. Mental workload was measured in participants working in synergistic co-operation with a physical and a virtual collaborative robot (cobot) under different levels of task demands. Performance and implicit and explicit workload were assessed as a function of pupil size variation and self-reporting questionnaires. In the face of a similar self-reported mental demand when maneuvering the virtual or physical cobot, operators showed shorter operation times and lower implicit workload when interacting with the virtual cobot compared to its physical counterpart. Furthermore, the benefits of collaborating with a virtual cobot most vividly manifested when the user had to position the robotic arm with higher precision. These results shed light on the feasibility and importance of relying on multidimensional assessments in real-life work settings, including implicit workload predictors such as pupillometric measures. From a broader perspective, our findings suggest that virtual simulations have the potential to bring significant advantages for both the user's mental well-being and industrial production, particularly for highly complex and demanding tasks.

1 Introduction

One of the core features of Industry 4.0 is the optimization of manufacturing processes without neglecting the worker's well-being (Kagermann 2015). In this context, many different technologies have been proposed and introduced, including collaborative robots (cobots). Unlike traditional robots, cobots not only support operators in carrying out a given task, but they literally share the task with them. In other terms, although both traditional robots and cobots can support the operator's work, only cobots allow close

fenceless cooperation. This distinctive feature of cobots has opened new possibilities for assembly tasks in particular, where human and machine can perform joint actions and work synergistically on a common task in a shared space, resulting in a powerful human–robot collaboration (HRC) (e.g., Cherubini et al. 2016). Indeed, the benefits of HRCs encompass time and space savings and support with repetitive tasks or handling and positioning of heavy materials. On the other hand, interacting with physical robotic arms can be expensive (Peters et al. 2018) and even dangerous in some cases (Oyekan et al. 2019). By tackling these issues, virtual simulations provide a cost-effective and safe environment for industrial operators who are called to collaborate with robots. The literature offers numerous examples of usage scenarios of VR within the industrial domain, including virtual prototyping (Berni and Borgianni 2020), training (Abidi et al. 2019; Pratticò and Lamberti 2021; Roldan et al. 2019), designing (Berg and Vance 2017; Fratczak et al. 2019; Krenn et al. 2021; Nee and Ong 2013), and teleoperation (Linn et al. 2017; Wang et al. 2019).

✉ Luciano Gamberini
   luciano.gamberini@unipd.it

   Federica Nenna
   federica.nenna@phd.unipd.it

1 Dipartimento Di Psicologia Generale, Università Degli Studi Di Padova, Padua, Italy

2 HIT Human Inspired Technology Centre, University of Padova, Padua, Italy

Springer

To support the actual implementation of physical and virtual human–cobot systems into industrial environments, in-depth user-centric investigations are needed. Billings (2018) introduced the concept of human-centered automation to indicate a system design in which humans and machines collaborate cooperatively in order to reach stated objectives. In this interplay, both humans and machines have strengths to be valued and weaknesses to be bridged. With the aim of improving overall system performance and maximizing the benefit of the HRC, each synergistic HRC should meet technical and ergonomic standards to optimize the operator's physical and mental workload (Nachreiner et al. 2006). However, to date, human psychophysical aspects have been regrettably marginalized. The majority of the papers dealing with smart manufacturing have focused on the feasibility of HRC systems and the efficacy of the related framework (e.g., Liu and Wang 2020), often neglecting the factors affecting the final user (for a review, see Damiani et al. 2018). Indeed, it is still unknown how HRC through a virtual simulation impacts the user's performance and cognitive workload and how it differs from collaboration with a physical robotic arm.

To address this research gap, we explored how interaction with a cobot in the physical or virtual environment affects the user. More specifically, we investigated operators' behavioral performance and cognitive state in a pick-and-place task executed in collaboration with a physical versus a virtual cobot. Additionally, we compared high and low task demand conditions. Data collected include operation times and task error as performance measures, changes in pupil size as a function of implicit workload, and self-reported explicit workload. We have organized the remainder of this paper as follows: Sect. 2 reviews the state of the art related to the use of virtual cobots in Industry 4.0 and various measures previously used for assessing HRCs. Section 3 addresses the reasoning and hypotheses related to the present investigation. Section 4 describes the chosen methods of our experiment and explains the experimental sample, technical setup, experimental procedure, measurements, and statistical analysis. Section 5 presents the results of our investigation, and Sect. 6 contextualizes the findings into the research panorama by clarifying their implications for the industrial domain. Finally, we present the main conclusions of this study, limitations, and potential future work in Sect. 7.

2 State of the art

2.1 Virtual robotics in Industry 4.0

This paragraph is meant to provide the reader with an understanding of various use cases related to immersive virtual HRC systems. When addressing VR applications for industrial robotics, we identified three main areas, namely (i) Education and training, (ii) Design of user interactions, and (iii) Telerobotics.

i. *Education and training:* Given that collaborative robots are relatively uncommon, virtual simulations for training staff can be helpful for familiarizing operators with robotic arms before they have to interact with them during a work shift (Horma et al. 2019). Pratticò and Lamberti 2021) developed a VR-based training system for a robotic arm. They showed that participants trained with the simulated cobot reached similar learning outcomes compared to those who underwent standard training. Matsas and Vosniakos (2017) presented a VR training system that simulates HRC activities for the manufacturing area. Their findings showed that most of the participants managed to complete the HRC tasks without entering the workspace (which was a potentially hazardous area) and reported favorable scores on the friendliness and understandability of the virtual interface. Participants who took part in Abidi et al.'s (2019) study were monitored in a product assembly task after having received either a VR training or a traditional training session. Those who had received the VR training made fewer errors and took less time in actual product assembly compared to those in the conventional training group.

ii. *Design of user interactions:* In the domain of HRC systems, where the characteristics of flexibility, adaptability, and safety are equally essential, reiterative design and testing are crucial for promoting human-centered solutions. Virtual simulations are a safe and economical space for testing and validating collaborative systems (Malik et al. 2020). Evidence of this is presented by Krenn et al. (2021), who investigated the understandability of different light- and motion-based signals delivered by a cobot in VR to indicate to the user when to take on the task. Virtual simulations of robotic arms have also been employed to determine the speed at which users felt safe when interacting with a cobot (Fratczak et al. 2019; Hansen et al. 2018). Additionally, human factor investigations have availed of VR to understand which factors contribute to a positive HRC (Mara et al. 2021) and how users react to more or less predictable robot movements (Oyekan et al. 2019). Kaufeld and Nickel (2019) conducted a virtual simulation study where human–robot interactions (HRIs) were evaluated at different levels of robot autonomy and in multi-modal signaling conditions.

iii. *Teleroboti*cs: A special mention is required for the field of telerobotics, in which human agents remotely guide robots through virtual interfaces. These human–robot systems are particularly useful for operating in

physical locations that are inaccessible or that involve physical risks of the operator (e.g., space repair, Xiao et al. 2020). Notably, with the advent of the COVID pandemic, interest in remote control and VR has turned into an urgent necessity even in the manufacturing field (Melluso et al. 2020). The spread of the pandemic has emphasized the weaknesses of current industrial systems and shed light on the importance of teleoperation and remote control of production sites. In this context, immersive VR devices are valuable mediums as they can supply information in a natural, interactive, and effective way, potentially promoting a high sense of presence, which in this context is better known as telepresence (Martín-Barrio et al. 2020; Zhang 2018). Wang et al. (2019) demonstrated the feasibility of remote HRCs via VR by designing a virtual teleoperation system for controlling an industrial cobot with satisfactory tracking accuracy. On an applied level, the adoption of VR has been used to enable remote maintenance service (Linn et al. 2017) or to support shop-floor operators by speeding up the programming and reconfiguration of a product line (Damiani et al. 2018).

2.2 Comparative literature on virtual and physical HRCs

Only a few studies have directly compared virtual and physical robotic systems, and even fewer have enabled direct physical manipulation of the robot. For instance, Inoue et al. (2005) assessed the psychological state of 13 users in co-presence with physical and virtual robots in a VR CAVE system. The results demonstrated that users self-reported similar feelings regarding the reliability, pleasantness, and friendliness of the robot's motion for real and virtual robots. In Li et al.'s (2019) study, preferences regarding distance from robots and the effects of appearance familiarity were compared between a virtual and physical robot in 80 participants. In both conditions, the robot gradually approached the participant and stopped when the participant perceived it as too close. Results suggested higher discomfort in the virtual condition, where greater distances were maintained between the user and the virtual robot, and no familiarity effects were observed. Weistroffer et al. (2014) assessed performance, physiological measures (heart rate and skin conductance) and self-reported co-presence comfort in 6 participants assembling a car door in a virtual and physical environment. They showed an increase in skin conductance after working close to the robot only in the physical situation, and no self-reported differences between the environments emerged. Hsieh and Lu (2018) asked 6 participants to pick some plastic balls delivered by a robotic arm and to sort them into specific boxes, both in the real and in an immersive virtual environment. In this context, where an actual HRI occurred, but the human movements were independent of the robot, different motion strategies were observed, but task completion time was similar between scenarios. Lipton et al. (2017) asked 8 participants to execute a pick-and-place with assembly task using a virtual interface, and another 7 participants performed the same task by teleoperating the physical robot. The virtual interface led to a higher number of correctly performed tasks and enabled faster operations in the pick task. In one of the few studies that allowed physical manipulation of the robot, Whitney et al. (2018) measured 2 participants' ability to complete 24 different tasks with both a virtual and a physical robot. Even though the physical robot enabled faster completion for most of the tasks, the virtual robot enabled more accurate results during the most complex robot positioning thanks to the lack of weight and inertia forces.

2.3 Metrics for the assessment of human factors in HRCs

When assessing an HRC, concepts like usability (Bevana et al. 1991), acceptance (Marangunìc and Granić 2015), and mental workload (Meijman and Mulder 2013) are of primary importance. In studies that addressed the effect of collaborating with a robotic arm on the user, these concepts have often been assessed through self-reporting and performance metrics. For instance, to investigate user experience with a robotic arm, Chowdhury et al. (2020) employed both qualitative (observations and semi-structured interviews) and quantitative (User Experience Questionnaire; Schrepp 2015) methods. Rossato et al. (2021) studied the subjective experience of younger and older adults teaming up with a cobot. More specifically, they used questionnaires to assess technology acceptance (TAM; Davis 1989), usability (SUS; Brooke 1996), user experience (Shirzad and Van der Loos 2016), and workload (NASA-TLX; Hart and Staveland 1988). Additionally, they measured the user's performance by coding the task execution time from the video recordings of the tasks. Chacon et al. (2021) assessed the usability of an HRC workspace. In particular, they measured efficiency as the time to complete the task and effectiveness as the percentage of task fulfillment. An adapted version of the System Usability Scale (Brooke 1996) was also employed to collect self-reported measures. Hsieh and Lu (2018) measured the task completion time of operators collaborating with either a physical interface or one of three different virtual interfaces. Similarly, Weistroffer et al. (2014) assessed the performance of users assembling a car door by recording the task duration and the number of completed operations and collision alarms activated. Kaufeld and Nickel (2019) evaluated human mental workload related to HRIs in VR. They collected performance data such as response times, error

rates, and the number of missed trials and the perceived mental workload through the NASA-TLX.

Besides performance and explicit workload measures, a number of studies found empirical evidence in favor of utilizing increased pupil diameter as an indicator of a higher mental workload (Beatty 1982; Iqbal et al. 2004; Van Orden et al. 2001). Pupil variation is linked to the sympathetic nervous system, which is involved in arousal and wakefulness (Mathot et al. 2018). Therefore, how pupil size variations are modulated by operations with one or the other interface can provide information about the level of implicit workload within a joint operation (Mingardi et al. 2020). It is worth noting that many variables other than the user's cognitive and emotional state (e.g., ambient lighting) can affect this metric (Kramer 2020). However, preprocessing methods and proper statistical analysis have been shown to overcome possible confounding variables relatively well (Mathot et al. 2018). Despite the stable relationship between pupil size and workload, research works applying pupillometry to the industrial and work field are scarce. Eye movements were employed as indicators of mental workload in a desktop-based version of a combat management workstation aboard naval vessels (de Greef et al. 2009); the authors proposed eye parameters as a potential trigger for adaptive automation of the system. Savur et al. (2019) presented two case studies involving a robotic arm in a collaborative pick-and-place task. Although the authors proposed a valuable framework for collecting and synchronizing multiple physiological outputs with the user's and the robot's behavior (pupil dilatation, EEG, GSR, PPG), the data collected about the user's cognitive state were not analyzed as a function of the task. Van Acker et al. (2020) tested the feasibility of deploying pupillometry in a work setting demanding operator mobility. They tested participants performing two manual assembly tasks with a different degree of complexity and found no significant differences in the implicit workload as suggested by pupil size variations, in contrast with substantial differences in the subjective workload. Therefore, they advocated further testing of pupillometry measures in real-life work settings to better understand their actual feasibility.

3 Our study

In this study, we report on a user-centric assessment of HRC in which we systematically discuss users' performance and implicit and explicit workload measures. Participants were asked to jointly perform a pick-and-place task with both a physical cobot and its virtual equivalent. The task was executed in conditions of either low or high cognitive load—that is, the user was concurrently busy with an arithmetic task (dual task)—and both in a physical and virtual industrial environment. Therefore, the experiment followed a 2×2 repeated measures design over interface (physical and virtual) and task load (single task, dual task). Performance and self-reported data as well as eye-tracking data were collected and analyzed. Furthermore, the level of participants' expertise with VR and users' preferences for potentially working with virtual or physical cobots were assessed. The hypotheses and research questions are listed below.

3.1 Hypotheses and research questions

3.1.1 Task load manipulation

Engaging the user in a dual task is a well-known condition that causes an increase in the load on cognitive resources (Navon and Miller 1987). Therefore, as a methodological control, we expect the task load manipulation to generate a higher explicit and implicit workload in the dual—compared to the single-task condition. More specifically, we predict a greater pupil size increase and a higher perceived workload from the NASA-TLX in the dual-task compared to the single-task condition. Additionally, we predict longer operation times for the pick-and-place task and more errors for the arithmetic task in the dual-task condition, which would indicate behavioral interference of the secondary task with the primary pick-and-place operation.

3.1.2 Operator's behavioral performance

The current literature on the impact of virtual vs. physical cobot manipulation on a user's performance is limited. However, the utility of virtual interfaces in industrial contexts can be corroborated only to the extent that the performance of users collaborating with a robot in VR does not decrease compared to that of users working in the physical space. In this respect, performance with the virtual interface is expected to be comparable to performance with the physical interface under both high and low workloads.

3.1.3 Operator's cognitive state

Previous studies did not systematically address how direct interactions with a physical or virtual cobot affect the user's cognitive state. Therefore, we intend to fill this gap by exploring whether collaborating with a physical or virtual cobot affects the user's workload under either the single- or dual-task condition. With this aim, we analyzed pupil size variations and responses to the NASA-TLX questionnaire as a function of implicit and explicit workload.

Fig. 1 **a** A participant performing the task with the physical cobot (e-Series UR5e); **b** a participant executing the task with the virtual robotic arm
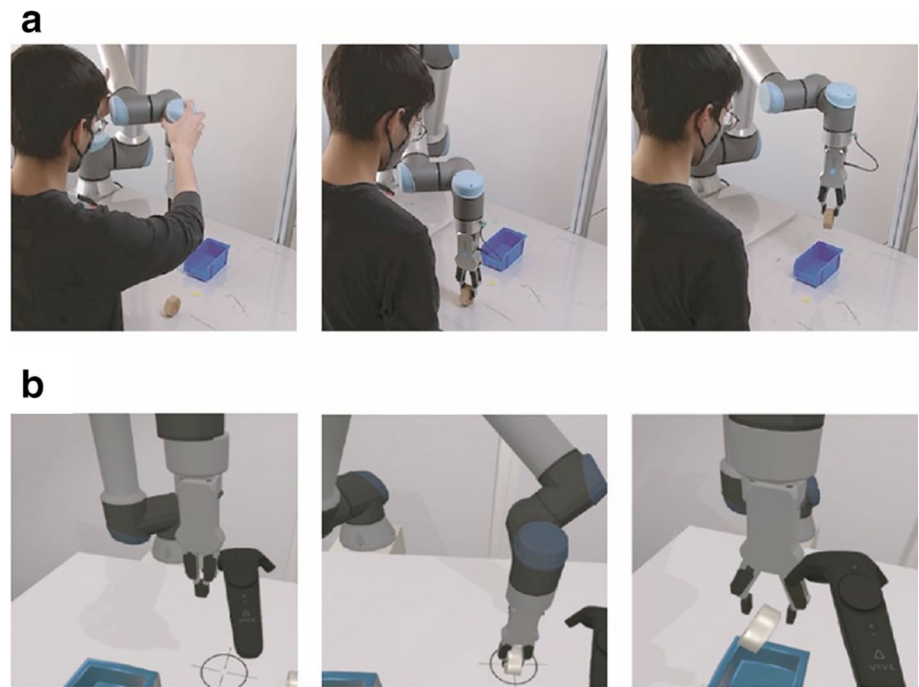
4 Methods

4.1 Sample

The experimental sample consisted of 26 participants, 8 women and 18 men ($M_{age} = 26.65$; $SD_{age} = 5.29$), who volunteered to take part in the study and signed the informed consent. None of the participants had current or past neurological or psychiatric problems. They all had normal or corrected-to-normal visual acuity and reported having normal color vision. The experimental protocol was approved by the local ethics committee and the study was conducted according to the principles of the Declaration of Helsinki. Three participants were excluded for technical issues related to the eye tracker device. Moreover, one participant was excluded for having an error rate of more than 50% at the arithmetic task and another for missing data in the arithmetic task. The final sample comprised 21 participants, 5 women and 16 men ($M_{age} = 26.95$; $SD_{age} = 2.52$).

4.2 Technical setup

In the physical condition, participants were provided with a pair of binocular eye-tracking glasses (Pupil Labs GmbH ©, Berlin, DE; weight 22.75 g) connected to an MSI laptop (model GT63 Titan 8RF, processor Intel Core i7-6700HQ, screen resolution 1920×1080, RAM 16 Gb). The software Pupil Capture (Pupil Labs GmbH ©, Berlin, DE) enabled system calibration and data recording (sampling frequency: 120 Hz; calibration: 5-point). The software Pupil Player (Pupil Labs GmbH ©, Berlin, DE) was utilized to export the eye-tracking data. Besides the eye data, the eye-tracker device also enabled first-person video recording through the embedded scene camera (480p, field of view: 100°×74°; sampling frequency: 120 Hz). The video recordings were then used for conducting a video analysis of the participants' arm behavior. The arithmetic task was managed through a program written and compiled in Visual Studio 2019 running on the same MSI laptop handling the eye-tracking recording. The pick-and-place task was performed jointly with an e-Series UR5e cobot (Fig. 1), which was installed on a height-adjustable worktable and was programmed in Polyscope (version 5.11) through its teach pendant. All data were recorded and processed by the same laptop and were thus synchronized based on the same internal clock.

In the virtual condition, participants were provided with an HTC Vive Pro Eye headset (resolution: 1440×1600 pixels per eye; refresh rate: 90 Hz; Field of view: 110°) and its controllers. The same headset also comprises an eye-tracking system (sampling frequency: 120 Hz; calibration: 5-point) which enables recording of eye parameters throughout the tasks. The virtual environment (Fig. 1) was programmed in Unity (version 2019.4.18f1) and faithfully reproduced not only the cobot and its workstation but also the surrounding environment (that is, windows, furniture, door). Participants interacted with the virtual cobot by means of physical action and by responding through the HTC Vive controller. At the end of each experimental session, all data (behavioral and eye data) were automatically saved on an MSI laptop (Intel Core i7-6700HQ, screen resolution 1920×1080).

Fig. 2 Pick-and-place task is depicted in all its steps constituting the pick phase (**a**) and place phase (**b**)
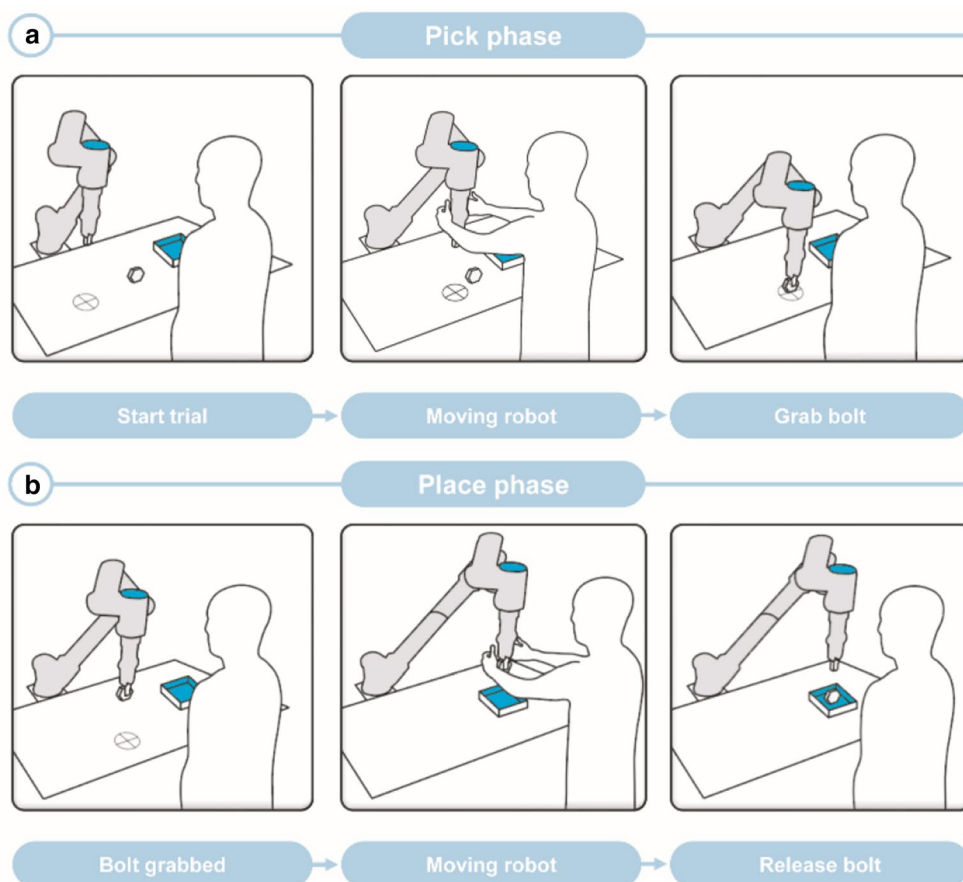
4.3 Measurements

4.3.1 Behavioral performance

In the pick-and-place task, the behavioral performance was measured as operation time, that is to say, the time required for the user to move the robotic arm to the desired location to either grab or release the bolt. More specifically, the operation time was computed from the time the user first touched the robotic arm (start) until the moment when the user released it (end) for both the pick (Fig. 2a) and the place phases (Fig. 2b). In the physical condition, the operation times were computed by coding the video recordings of the experimental trials[1] with the software BORIS (version 7.10.5, Friard and Gamba 2016). More specifically, the first frame showing the user's hands touching the physical cobot was coded as the beginning of the operation time, and the frame showing the user's hands releasing the cobot was coded as the end of the operation time. The obtained "start" and "end" timestamps were imported into the pupillometry data stream. In the virtual condition, the first movement of the virtual robotic arm was automatically logged by the Unity software as the timestamp of the button press (grip button, Fig. 3) that co-occurred with the contact between the controller and the virtual cobot. Likewise, the software logged the end of the operation time as the timestamp of the button press (pad button, Fig. 3). Overall, for both the physical and virtual conditions, the operation times for picking up and placing the bolt were considered independently because of different levels of difficulty; the pick phase required higher precision for positioning the cobot's joints in a suitable and accurate way, but in the place phase less accuracy was required. In the arithmetic task, we computed the percentage of wrong answers. This performance index provided information on the degree of cognitive interference that occurred in the dual—compared to the single-task condition both in the virtual and physical conditions.

4.3.2 Implicit workload (pupil size variation)

The pupil size variation was computed only during the moving robot phase (Fig. 2) and was considered as a proxy of the experienced workload (Beatty 1982; Iqbal et al. 2004; Van Acker et al. 2020). In this study, we followed the

---

[1] A first-person view of the experimental trials was video recorded with the PupilLabs world camera.

Fig. 3 Pick-and-place task is depicted in all its steps constituting the pick phase (**a**) and place phase (**b**)
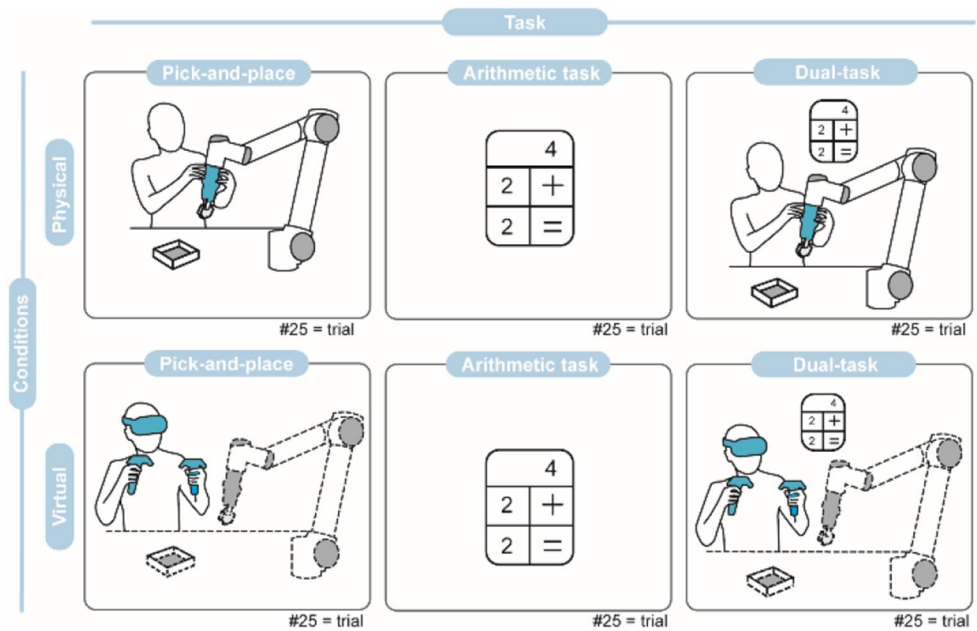
preprocessing methodology of Kret and Sjak-Shie (2019) and accommodated the precautions of Mathôt et al. (2018) for the baseline correction. First, we selected time windows within the moving robot parts of both the pick and the place phases (Fig. 2) and handled them independently from each other. Considering that different operation times resulted in different lengths of the selected time series, we used dynamic time warping (Berndt and Clifford 1994; Keogh and Pazzani 2001) to standardize the length of each time series. Thus, all pupil samples were constrained to fall within a warping window of 30 data points. The average length of the selected windows was about 1.5 s; therefore, each data point of the warped window corresponded to 50 ms on average. After averaging over the left and right eye, we computed the percentage of missing data in each trial and participant and removed those for which more than 35% of the data were missing (1 trial and 0 participants were removed). Data were then filtered through a median filter, and the first 4 data points of each trial—which correspond to 200 ms on average—were used to apply a subtractive baseline correction (Mathôt et al. 2018). By addressing the difference in pupil size compared to a baseline period, we marginalized absolute differences caused by external variables other than those due to changes in the cognitive state. Unlike for the processing of pupil response during the pick-and-place task, in the arithmetic task we selected four time windows corresponding to each number presentation. Because their duration varied between 2.3 and 2.7 s, dynamic time warping was applied in each of the windows to standardize their length (Berndt and Clifford 1994; Keogh and Pazzani 2001). Then the same procedure was followed for the processing of pupil data in the arithmetic task.

4.3.3 Explicit workload (NASA-TLX)

After each task, participants were asked to fill in the NASA-TLX questionnaire as a measure of perceived workload. This scale has been used extensively in many areas, with the industrial context being just one of them (e.g., Roldán et al. 2019).

4.3.4 Individual factors

Participants were asked to self-report their level of previous experience with VR technology by rating the frequency with which they had used VR devices on a 5-point scale. The aim was to control for the level of VR experience within the sample. Additionally, we explored participants' expectations of working with the cobot compared to their experience with it by asking their preferences before and after the experiment. More specifically, before the experiment, we asked: "If you had to collaboratively work with a cobot, which of the following interfaces would you prefer?" and, after the experiment, we asked, "With reference to the experience you have just concluded, which of the following interfaces did you prefer?" The possible answers were "Virtual cobot" and "Physical cobot."

4.4 Task and procedure

After signing informed consent, all participants filled out a demographic questionnaire and answered questions about their VR expertise and individual preference for virtual vs. physical cobots. Then, they undertook six tasks composed of 25 trials each. In particular, as shown in Fig. 3, a

pick-and-place task, an arithmetic task and a dual task were performed both in the virtual and physical environments (counterbalanced order). Half of the participants started with the virtual condition and the other half with the physical condition. At the beginning of each task condition (virtual and physical), all participants underwent a training session and performed a few trials of the same tasks administered in the subsequent experimental session. The task instructions were presented in paper format in the physical condition, and they were virtually delivered in text format in the virtual environment. The experiment started only when the participant understood all the task rules. In both contexts, a 5-point calibration of the eye-tracking systems was conducted before starting the experiment. After each task, participants filled in the NASA-TLX questionnaire and, only at the end of the virtual experimental session, the MEC-SPQ was also administered. Additionally, between each task, it was possible to take a break both in the virtual and physical environments, after which the eye-tracking system was re-calibrated before starting the next task. At the end of both the virtual and physical experimental sessions, the final questionnaire on the individual preference for virtual vs. physical cobots was administered.

4.4.1 Pick-and-place task

For each trial of the pick-and-place task, a bolt and a box were always placed in random positions on the worktable, still keeping 50 cm of distance between them. Participants were instructed to pick the bolt up from the worktable and place it into the box by physically moving the robotic arm. The activity was designed to distinguish clearly between the pick and place phases. The pick phase required precise maneuvering of the robotic arm to align its effector with the bolt to be picked up. For the place phase, on the other hand, less precision was needed because the box in which to place the bolt was relatively large.

In the physical condition, participants first had to grasp the robotic arm with their hands and physically move it close to the bolt (moving robot part of the pick phase, Fig. 2). Once the robot's effector was in line with the bolt, participants initiated the grab bolt (Fig. 2) automation by gently hitting the worktable with their hand and the robotic arm automatically picked up the bolt (bolt grabbed, Fig. 2). Afterward, they grasped the robotic arm, positioned it over the box (moving robot part of the place phase, Fig. 2), and hit the worktable again to enable the cobot to automatically release the bolt in the chosen position (release bolt, Fig. 2). We used the Wizard of Oz method (Hsieh and Lu 2018; for a review, see Riek 2012) for initiating the grab bolt and release bolt automations: when participants touched the worktable, an experimenter standing behind the participants initiated the grab/release bolt command from the teach pendant. This

mechanism was hidden from the participants, who were led to perceive the feature as related to their action of touching the table.

In the virtual condition, on the other hand, participants used the HTC VIVE controllers to perform the same VR task. Specifically, they were instructed to approach the cobot with their hand. When the controller physically collided with the virtual cobot, participants could grasp it by keeping the grip button pressed and move it to the desired position as in the physical condition. To initiate the grab bolt and release bolt automations, they pressed the pad button on the right controller.

4.4.2 Arithmetic task

A series of numbers randomly ranging between 1 and 10 were aurally presented to the participants, who were asked to mentally sum them and then report the result of the arithmetic operations. Between each number, a time interval of $2.5\,\text{s} \pm 0.3\,\text{s}$ of jitter elapsed, and each series comprised 4 or 5 numbers to avoid possible learning effects. In the virtual condition, participants reported the result of each mental operation by interacting with a virtual numeric keyboard via controller. In the physical condition, they were asked to report the sum's result verbally. The response was systematically collected via the Visual Studio application described in subparagraph 4.2.

4.4.3 Dual task

In the dual task, participants were instructed to perform the pick-and-place task and the arithmetic task concurrently. In each trial, the numbers of the arithmetic task were presented for the whole pick-and-place task, and the result was then reported only after the release bolt action (Fig. 2).

4.5 Statistical analysis

4.5.1 Behavioral performance

Performance data were analyzed using generalized linear models (GLMs from the *lme4* package, Bates et al. 2014) in RStudio (Team 2021). To analyze performance at the arithmetic task and at the pick-and-place task, we computed a GLM that included the factors task load (single task, dual task) and interface (virtual, physical) with participant as a random effect. Specifically, for the operation times at the pick-and-place task, the pick and place phases were analyzed independently. The Bonferroni correction was always applied when interpreting the post hoc contrasts within the significant interactions.

4.5.2 Implicit workload (pupil size variation)

To analyze pupil size variations, we used generalized additive mixed models (GAMMs; Hastie and Tibshirani 2017; Wood 2017) and linear mixed-effects models (LMERs; Bates et al. 2014). However, for greater conciseness of the results, further specification on the GAMMs' chosen parameters and relative results is found in Appendix, and we hereafter report only on results obtained through the LMERs. GAMMs are a good fit for our pupil data, as they are able to model nonlinear patterns by using penalized regression splines and estimate the shape of the regression line based on the data (for an introduction, see van Rij et al. 2019; Wieling 2018). However, this methodology is still poorly explored. Furthermore, the GAMMs' summary statistics do not tell where the difference curves are different from zero, nor the amplitude of the difference. Therefore, for higher robustness of our results, we ran a chunk analysis over six windows (each corresponding to 250 ms on average) to determine significant differences in the time course through the LMERs. The latter models involved task load (single task, dual task), interface (virtual, physical), window (1, 2, 3, 4, 5, 6), and their interactions with the participant as a random effect. As for the operation time, the pick and the place phases were analyzed independently.

In the single arithmetic task, we analyzed whether there were significant differences in pupil size between the beginning of the arithmetic task (start) and the following arithmetic sums (first, second, and third arithmetic operations). As three or four arithmetic operations occurred randomly, only the first three arithmetic operations were considered to prevent the learning effect. We computed one LMER for each interface condition (virtual and physical) with arithmetic operation as a fixed factor (start, first sum, second sum, third sum) and participant as a random effect. The Bonferroni method was consistently applied in the post hoc contrasts analysis (Bonferroni 1936).

4.5.3 Explicit workload (NASA-TLX)

The analysis of the NASA-TLX questionnaire score was conducted through a GLM over task load (single task, dual task), interface (virtual, physical), and items (mental demand, physical demand, temporal demand, performance, effort, frustration), with participant as random effect. Post hoc contrasts were performed on each of the significant interactions with the application of the Bonferroni correction for multiple comparisons (Bonferroni 1936).

4.5.4 Individual factors

For individual VR experience, we first standardized the participants' responses and then created two levels of VR experience: participants with a scaled score below 0.5 were assigned to the low VR experience level, and those with a scaled score higher than 0.5 were assigned to the high VR experience level. With regard to the individual preference for a virtual or physical cobot as expressed before and after the experiment, we reported the percentage of answers in favor of the virtual or physical cobot.

5 Results

5.1 Performance measures

5.1.1 Operation time

Results yielded significant main effects only for interface, both in the pick phase (X^2 (1, $N = 21$) = 1057.5, $p < 0.0001$) and in the place phase (X^2 (1, $N = 21$) = 1252.4, $p < 0.0001$), with a faster operation time for the virtual interface compared to the physical interface, both in the pick and in the place phases (Fig. 4). The task load manipulation, however, did not yield any significant differences in operation times in the pick or in the place phase. Descriptive statistics relative to the operation time is found in Table 1.

5.1.2 Arithmetic task error

When analyzing the effects of task load and interface on the arithmetic task error, none of the factors reached the significance threshold.

### 5.2 Implicit workload (pupil size variation)

5.2.1 Pick-and-place task

When running the chunk analysis over the six time windows, the results obtained by the LMERs were in line with those obtained through the GAMMs (which are found in Appendix) (Fig. 5). To analyze whether the effects changed in the time course, we specifically focused on interactions involving the window factor. Significant interactions were observed between: task load and window only in the pick phase (X^2 (5, $N = 21$) = 148.38, $p < 0.0001$), interface and window (pick: X^2 (5, $N = 21$) = 442.4, $p < 0.0001$; place: X^2 (5, $N = 21$) = 23.72, $p < 0.001$), and task load, interface and window (pick: X^2 (5, $N = 21$) = 80.51, $p < 0.0001$; place: X^2 (5, $N = 21$) = 34.88, $p < 0.0001$). Post hoc contrasts that are of interest for the present study are shown in Figs. 6 and 7.

5.2.2 Arithmetic task

The results of pupil size variations in the virtual condition highlighted a significant effect of the arithmetic task (X^2 (3,

Fig. 4 Averaged operation time with standard error at the pick and place phases independently according to the interface (virtual, physical)
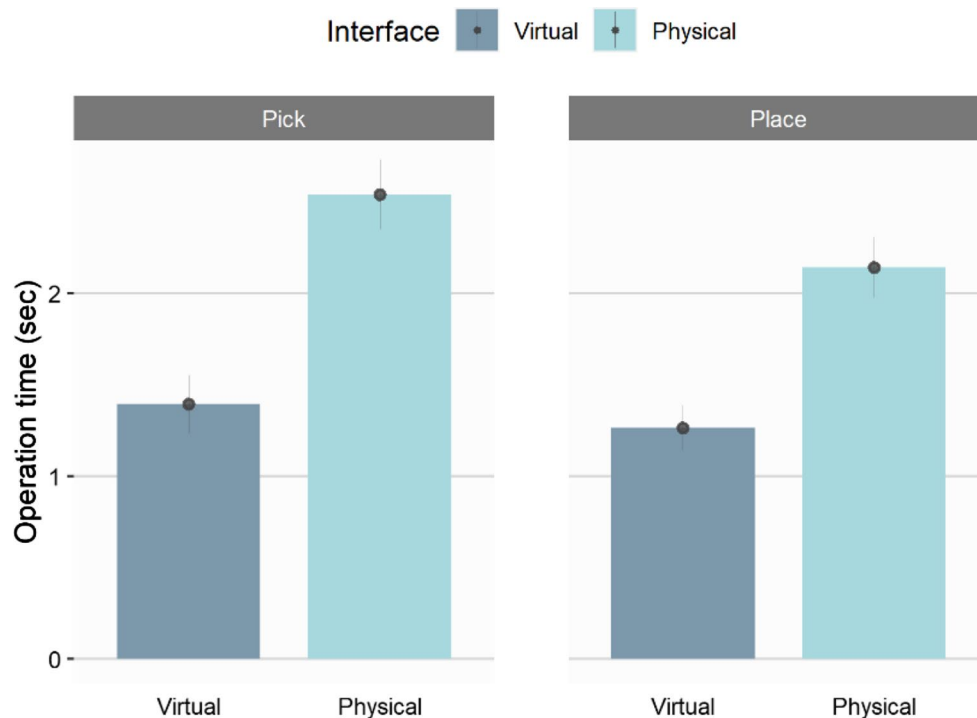

Table 1 Descriptive statistics of operation time at the pick-and-place task

		Pick mean (SD)	Place mean (SD)
Task load	Single task	1.89 (0.96)	1.67 (0.77)
	Dual task	1.91 (1.00)	1.69 (0.81)
Interface	Virtual	1.40 (0.73)	1.26 (0.56)
	Physical	2.54 (0.87)	2.14 (0.75)

Fig. 5 Averaged operation time with standard error at the pick and place phases independently according to the interface (virtual, physical)

$N = 21) = 893.96$, $p < 0.0001$). Similar results were observed in the physical condition, where there was a significant effect of the arithmetic task (X^2 (3, $N = 21) = 97.05$, $p < 0.0001$). Post hoc contrasts run with Bonferroni correction are shown in Fig. 8, and descriptive statistics are shown in Table 2.

5.3 Explicit workload (NASA-TLX questionnaire)

The results of the linear mixed model (LMM) demonstrated significant effects of task load (X^2 (1, $N = 21) = 45.6$, $p < 0.0001$) and item (X^2 (5, $N = 21) = 311.79$, $p < 0.0001$) and interactions between task load and item (X^2 (5, $N = 21) = 42.04$, $p < 0.0001$) and between interface and item (X^2 (5, $N = 21) = 32.3$, $p < 0.0001$). Specifically, a higher NASA-TLX score was reported in the dual-task condition ($M = 10.9$; SD $= 5.09$) than in the single-task condition ($M = 8.79$; SD $= 5.82$). The post hoc contrasts on the
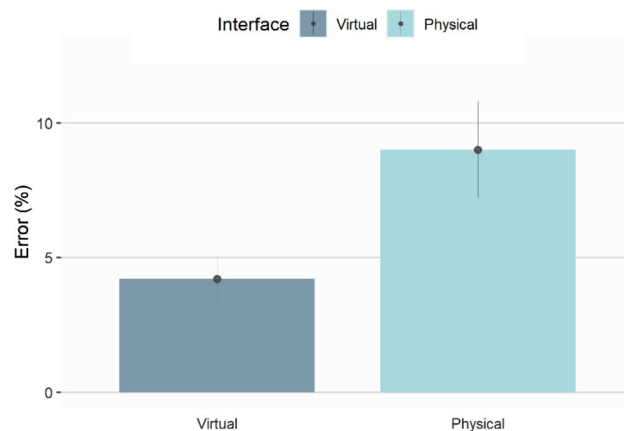
interaction between task load and item revealed a higher NASA-TLX score in the dual-task condition than in the single-task condition in the following items: mental demand ($p = < 0.0001$; ST: $M = 7.85$, SD $= 5.44$; DT: $M = 13.00$, SD $= 4.80$), physical demand ($p < 0.05$; ST: $M = 5.85$, SD $= 4.54$; DT: $M = 7.67$, SD $= 4.38$), and effort ($p < 0.001$; ST: $M = 9.57$, SD $= 5.51$; DT: $M = 13.6$, SD $= 4.32$). Moreover, post hoc contrasts over the interaction between interface and items yield a higher NASA-TLX score in the physical
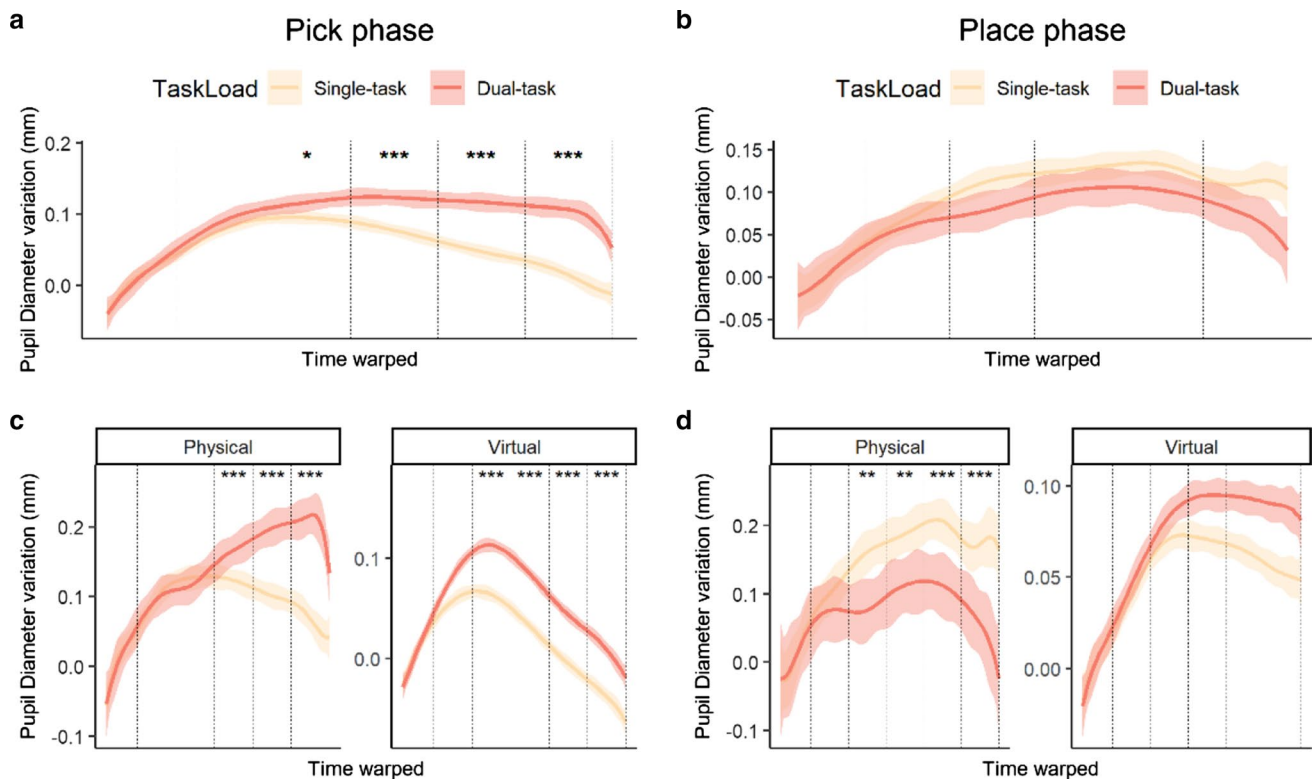
Fig. 6 Pupil size variations relative to the task load conditions in the pick (**a**, **c**) and place (**b**, **d**) phases. Plots a and b depict the main effect of task load. Plots c and d display the effects of task load by interface. All the plots are complemented by stars indicating the significance level of the statistical test (*$p \leq .05$; **$p \leq .01$; ***$p \leq .0001$)

condition than in the virtual condition for the item performance ($p < 0.01$; virtual: $M = 14.5$, SD $= 4.3$; physical: $M = 15.8$, SD $= 4.15$) and a higher score in the virtual condition than in the physical condition for the item frustration ($p < 0.05$; virtual: $M = 6.70$, SD $= 4.24$; physical: $M = 5.65$, SD $= 4.35$). NASA-TLX score differences for both task load and interface are depicted in Fig. 9

5.4 Individual factors

5.4.1 VR experience

On a scale from 1 to 5, the median VR experience was 2, with a standard deviation of 0.84. In our sample, 3 participants were considered to have high VR experience, as their scaled rating was higher than 0.5, and 18 participants were considered to have low VR experience, as their rating was below 0.5.

5.4.2 Individual preferences for a virtual or physical cobot

Finally, individual preferences for virtual or physical cobots expressed before and after the experiment are shown in Fig. 10. Before the experiment, 19.05% of participants expressed a preference for virtual cobots, but after the experiment, the percentage increased to 61.9%.

6 Discussion

Virtual simulations of industrial cobots have shown to be promising tools for training (Prattiçò and Lamberti 2021), teleoperation (Wang et al. 2019), and design and prototyping (Kaufeld and Nickel 2019). The increasing interest in similar virtual interfaces derives from their potential to reduce experiential differences between virtual simulations and real operations and exclude risks related to physical interactions with robots and materials (Li et al. 2019). Despite its relevance for both the operator's psychophysical well-being and industrial productivity, the assessment of the cognitive states of users working with either virtual or physical cobots has received little emphasis.
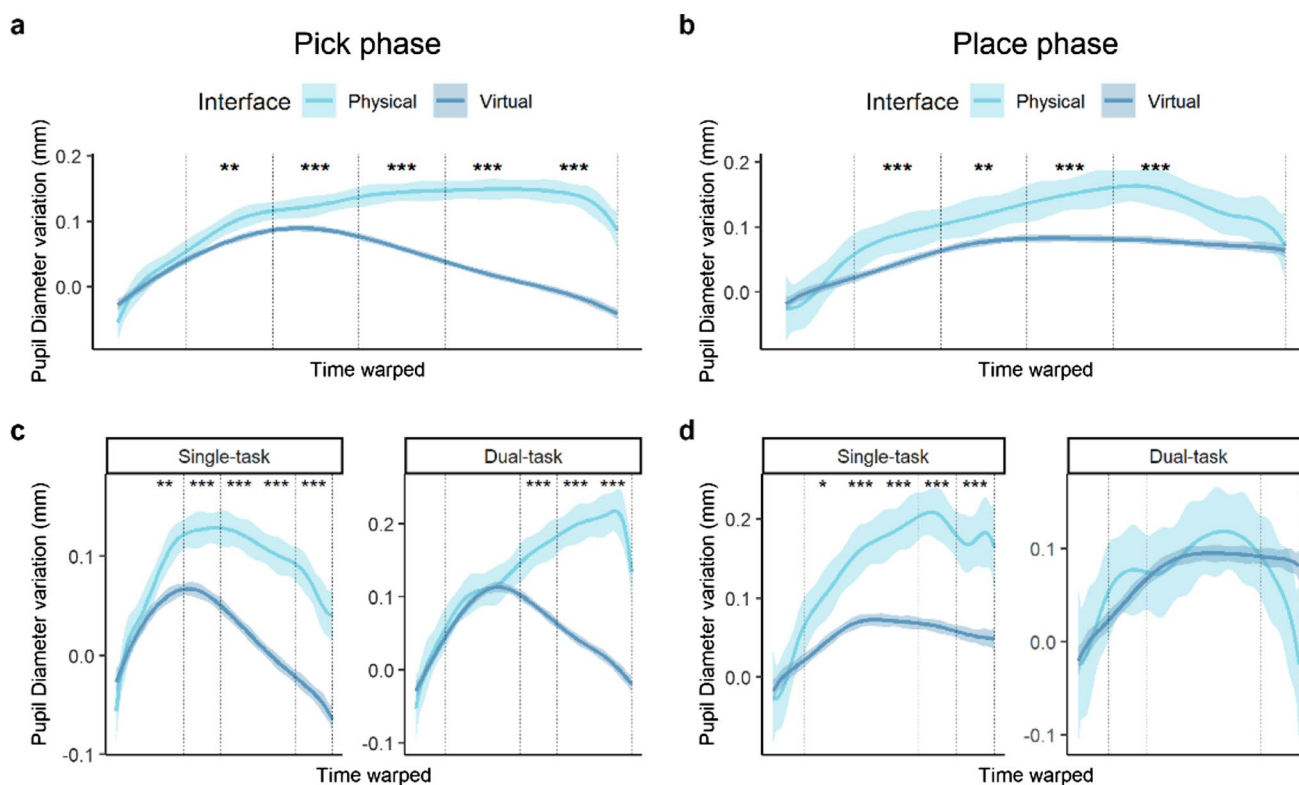
Fig. 7 Pupil size variations relative to the interface conditions in the pick (plots **a** and **c**) and place (plot **b** and **d**) phases. Plots a and b depict the main effect of interface. Plots c and d display the effects of interface by task load in the pick phase and place phase. All the plots are complemented by stars indicating the significance level of the statistical test (*$p \leq .05$; **$p \leq .01$; ***$p \leq .0001$)

In this study, we thus faithfully reproduced a cobot in VR and tested the impact of such a virtual simulation on the user's cognitive state compared to its physical counterpart. As the core of this work, in addition to users' performance, we explored implicit and explicit workload measures when operating with both interfaces and under high and low demand. The purpose underlying this user-centric investigation was to systematically define which of the two interfaces allows users to perform actions with the lowest cognitive effort in both slightly (single-task) and highly (dual-task) demanding working conditions. Ultimately, we also looked at whether participants' preferences for the physical or virtual interface changed after using them for the duration of the experimental session.

6.1 Task load manipulation

6.1.1 The arithmetic task load is reflected in pupil size variations

Our task load manipulation relied on the dual-tasking methodology. In addition to the primary pick-and-place task, we introduced a secondary task where users were asked to compute a series of arithmetic sums. The analysis of pupil

size variations in the arithmetic task revealed a significant increase in pupil size as participants were moving from the start through the subsequent mental operations (first sum, second sum, third sum, Fig. 8). In both the physical and virtual conditions, the pupil diameter was stable from the start to the first sum and then increased considerably in the last two mental operations. This gradual increase in pupil size was clearly observable with visual inspection of Fig. 8 in each experimental setting, suggesting that our arithmetic task induced a gradual increase in the implicit workload in both the virtual and physical conditions.

### 6.1.2 The dual task affected user's implicit and explicit workload

Regarding the explicit workload as self-reported through the NASA-TLX questionnaire, participants indicated higher mental demands, physical demands, and efforts when executing the pick-and-place task along with the arithmetic task compared to the single-task condition (Fig. 9). This result supports our hypothesis that the task load manipulation would affect the explicit workload. Additionally, implicit workload and behavioral performance measures were collected and analyzed independently in the pick and the place

Table 2 Descriptive statistics of the absolute pupil size (mm) in the arithmetic task

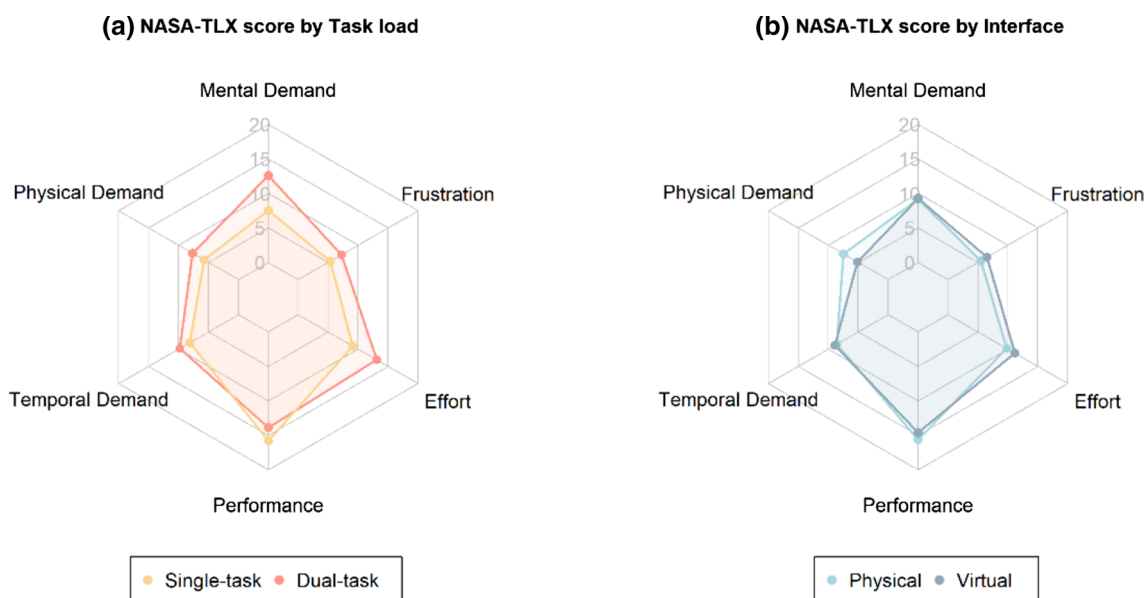|	Start mean (SD)	First sum mean (SD)	Second sum mean (SD)	Third sum mean (SD)
Virtual	3.06 (0.37)	3.06 (0.37)	3.10 (0.38)	3.12 (0.38)
Physical	2.82 (0.94)	2.83 (0.92)	2.86 (0.93)	2.88 (0.93)

phases. In this respect, if the dual tasking demonstrated an effect on the implicit workload, the users' performances did not differ between the single- and dual-task conditions, failing to meet our expectations. Specifically, the operation times were not significantly different between the single- and dual-task conditions. Similarly, even though participants committed 2.03% more errors on average in the dual arithmetic task compared to the single one, this difference did not reach the significance threshold. However, an increase in implicit workload was shown by the pupil size variation within the pick action. In fact, a significantly higher pupil size variation was evident in the dual-task condition compared to the single-task condition beginning 620 ms on average after the pick task began and continuing until the end of the action (Fig. 6a). This task load effect was evident when participants were working jointly with both the physical and virtual cobots: in both cases, the pupil size associated with the dual-task condition demonstrated a continuous increase throughout the pick action in contrast to the pupil size variation captured within the single task, which was characterized by lower variation and a faster decrease (Fig. 6c). Differently, no stable differences in pupil diameter variation between the single- and dual-task conditions were observed within the place phase (Fig. 6b). In the latter case, it is possible that the extreme simplicity of the place action prevented pupil size variations between the single- and

dual-task conditions. Moreover, unexpectedly, regarding the pupil size variation within the place action performed with the physical and virtual cobots (Fig. 6d), a higher pupil size variation was observed in the single-task condition compared to the dual-task condition only in the physical condition. This effect might be related to different levels of precision required by the two maneuvering actions (pick and place) and/or to the temporality of the same actions. Indeed, it is possible that in the dual-task condition, users employed higher cognitive resources at the beginning of the task for concurrently handling the arithmetic task and initiating the pick action (Fig. 6c), and they relieved their mental efforts during the subsequent and more rough place action (Fig. 6d). When participants were performing the same pick-and-place task as a single task, their pupil sizes instead just gradually increased throughout the task. Still, it is interesting to notice how this reverse effect was visible only in the physical condition but not in the virtual condition, where the dual tasking affected the pupil size variation without any influence of the temporality of the actions. Overall, in line with previous studies on pupil size variations and workloads (Beatty 1982; Iqbal et al. 2004; Van Orden et al. 2001), the trend observed when the task required precise maneuvering of the robotic arm stands for a higher implicit workload in the dual-task condition compared to the single-task condition.
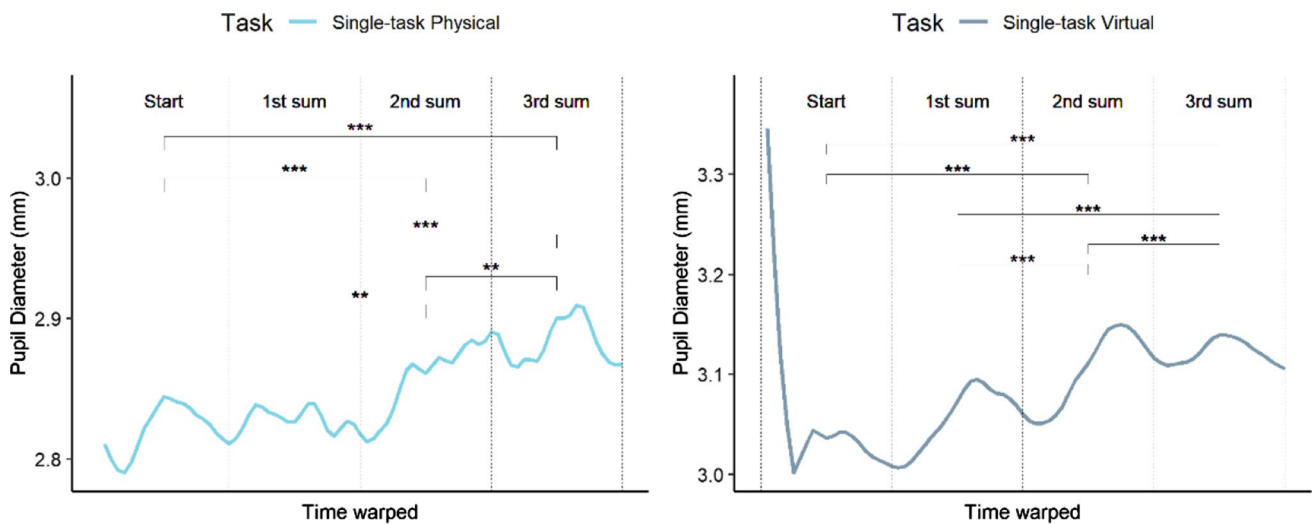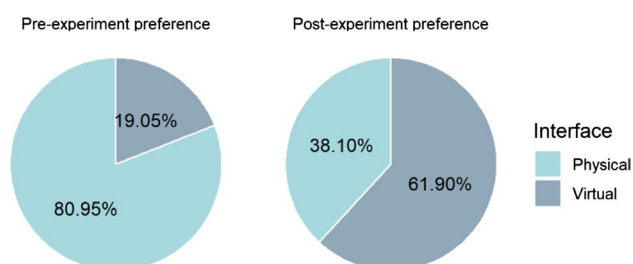

Fig. 8 Pupil diameter in the physical and virtual conditions. The vertical dashed lines divide the plots by windows (start, first sum, second sum, third sum). Plots are complemented by stars indicating the significance level of the statistical test between windows (*$p \leq .05$; **$p \leq .01$; ***$p \leq .0001$)

(a) NASA-TLX score by Task load

Mental Demand

Physical Demand

Frustration

Temporal Demand

Effort

Performance

● Single-task ● Dual-task

(b) NASA-TLX score by Interface

Mental Demand

Physical Demand

Frustration

Temporal Demand

Effort

Performance

● Physical ● Virtual

Fig. 9 Averaged NASA-TLX score in each NASA-TLX item for task load (**a**) and interface (**b**)

Pre-experiment preference

19.05%

80.95%

Post-experiment preference

38.10%

61.90%

Interface

Physical

Virtual

Fig. 10 Pre- and post-experiment preferences for working in collaboration with a physical or virtual robot as expressed by participants before and after the experimental session

6.2 Operator's behavioral performance

6.2.1 The virtual cobot enables faster operations than its physical counterpart

The operators' behavioral performances were evaluated in terms of operation times, which we expected to be comparable between the virtual and physical cobots. The literature on users' performances in HRIs referred mainly to situations in which a direct manipulation of the robot by the hand of the user were missing (Hsieh and Lu 2018; Lipton et al. 2017). Differently, when direct HRIs were enabled, operators using physical robots have been observed to be faster than virtual interaction. However, this advantage did not hold for tasks requiring complex robot positioning (Whitney et al. 2018). Interestingly, we found a clear reduction in operation times when participants were working with the virtual cobot rather than the physical equivalent, regardless of the degree of

complexity of the robotic arm positioning. Specifically, users saved about 1 s on average in each of the pick-and-place phases when cooperating with the virtual cobot (Fig. 4). This advantage could be crucial for optimizing manufacturing processes involving human–cobot operations that can be performed remotely. Indeed, some considerations in this regard are needed. The physical structure of the robotic arm required participants to use their strength to drive the cobot through the desired positions on their workstations. Differently, the virtual robotic system had no inertia forces, which led users to perform the operations as freely as if they were unbounded from the robotic arm. Therefore, performing the same physical actions with and without the resistance of the physical cobot largely affected operation times. This peculiarity of VR technology, which allowed faster and safer operations with the cobot, could be particularly beneficial in the field of teleoperation but precludes scenarios where humans and robots are actually collaborating physically. Among the virtual systems applications that would allow significant time savings in industry, the literature offered valuable examples of VR integrations into the design and testing of HRC systems (e.g., Fratczak et al. 2019; Hansen et al. 2018; Krenn et al. 2021). In these contexts, the advantages of reduced operation times in VR could significantly speed up the HRC design and validation processes of those aspects that do not involve the physicality of the robotic system. Moreover, Wang et al. (2014) highlighted how reducing robot programming time and providing collision-free solutions are central issues in distributed manufacturing. For this purpose, a virtual simulation such as the one we developed in this work could optimize the manual configuration of a

cobot before the preset tasks are autonomously run, enhancing the efficiency and safety of the programming operations. Finally, VR could also be a valuable tool for training novices before they approach physical cobots (Abidi et al. 2019; Matsas and Vosniakos 2017; Prattico and Lamberti 2021; Roldan et al. 2019). Indeed, the agility and freedom of movement within the virtual simulation could help less-experienced operators become familiar with the task and cobot management modalities through trial and error in a safe, economical, and fast fashion. In this sense, operators can learn the procedural aspects of a task in VR as a first training step, making the learning process faster and safer, before they perform the actual task with a physical robot.

6.3 Operator's cognitive state in the virtual and physical environments

6.3.1 Interacting with the virtual cobot reduces the implicit workload

In the virtual environment, a significantly lower pupil variation was observed, suggesting a lower implicit workload. This effect was observed throughout the whole pick-and-place task, but it was particularly evident for highly accurate movements (namely, the pick phase; Fig. 7a). A further differentiation between the single- and dual-task conditions was made on the low and high mental demands, respectively, imposed on the operator. When participants were executing the pick task as a single task, the pupil variation relative to the virtual and physical cobots' operations followed almost the same pattern, with a slightly higher pupil size variation in the physical condition compared to the virtual one. Differently, under the dual-task condition, the fluctuations in pupil size elicited by the pick operation with the physical cobot underwent a much greater increase than those elicited by the virtual simulation (Fig. 7c). Regarding more gross maneuvering of the cobot as in the place phase, the pupil size variation still suggests the virtual cobot had an advantage over the physical one in terms of implicit workload (Fig. 7b). However, because the task load manipulation did not impact the pupil size variations in the place phase (Fig. 6b, d), the effects of the virtual and physical interfaces on pupil size variation were not interpretable in their interactions with the task load (Fig. 7d). Our data thus revealed that a lower implicit workload was experienced in the virtual environment than in the physical one, particularly when precise maneuvering of the robotic arm was required and when the operator was imposed with a high mental demand. The same trend applied for tasks requiring rough maneuvering of the robotic arm, but the effect was less marked. Thus, the higher the task complexity, the more the virtual simulation revealed to be preferable, because it allowed the user to save mental resources. Notably, the usefulness of virtual

simulations emerged when the tasks became more complex in terms of both mental demand (dual vs. single task) and the precision of the maneuvering (pick vs. place), whereby a highly controlled visuomotor coordination was implied. This finding may foster the introduction of virtual HRC systems, particularly in highly complex or demanding work environments where a higher risk of accidents is involved, and the usefulness of virtual simulations thus touches its peak. In each HRC framework, whether dangerous or not, it is essential to ensure the user is able to maintain high vigilance and awareness with the minimum workload for avoiding mental and physical safety issues (Matsas et al. 2018). Therefore, the lower implicit workload related to virtualization is a relevant advantage in terms of users' safety and well-being, which especially applies to complex and potentially hazardous HRCs.

6.3.2 Advantages of virtualization were not reflected in the explicit workload

Interestingly, the trend for the implicit workload only partially matched the trend for the explicit workload. Specifically, when operating the physical robot rather than the virtual one, participants reported a generally higher NASA-TLX score, which suggests higher perceived workloads. However, regarding the single questionnaire's dimensions, users self-rated their own performances as better when using the physical robot compared to the virtual one, whereas they also reported higher frustrations when working with the virtual robot compared to the physical one. This might be due to the participants' limited VR expertise: performing operations in VR for the first time might generate uncertainties, likely leading participants to question the quality of their performances and possibly feel frustrated. Future studies might better examine this aspect and systematically measure whether the repeated use of such a virtual interface has a positive influence on the perception of one's performance and on the feeling of frustration. However, it was interesting to notice that the objective performance did not match the subjective ratings. Indeed, both performance and pupil size suggest an advantage of the virtual simulation over the physical cobot. Similar results were observed in a mixed-reality framework by Kaufeld and Nickel (2019), who found that the level of autonomy and information aids in their robotic system affected users' performances but not their mental workload ratings on the NASA-TLX. The authors interpreted this effect according to the compensatory control model and assumed that users adjusted their task performance strategies by shifting to simpler or less precise procedures (Hockey 1997). Similar dynamics likely occurred in our scenario, where the performance worsening observed when participants were working with the physical cobot might have reduced the human mental load and thus mitigated the level

of perceived mental demand. One important implication of such a finding is that questionnaire-based cognitive evaluations might not be sufficient because what the human being consciously perceives does not always accurately reflect actual activation. It follows that when designing and testing HRCs, creators should employ multidimensional evaluations involving subjective ratings and performance and implicit indexes related to mental workload or stress for a global understanding of the cognitive states of users engaged in HRCs.

6.4 Individual factors

Before the experiment, only 19.05% of participants expressed a preference for working with a virtual cobot rather than with a physical one. Notably, preferences for the virtual cobot increased to 61.9% after participants used both cobot interfaces for the duration of the experimental session. This is promising data, suggesting a likely positive acceptance of virtual HRCs in the perspective of large-scale implementations of virtual simulations in industry.

7 Conclusions, limitations, and future works

With this study, we contributed to the state of smart manufacturing by conducting a systematic user-centric assessment of human performance and mental workload in virtual versus physical cobot operations and under different levels of mental load. Our findings suggest that virtual simulations of HRCs have the potential to create significant advantages for both users' mental well-being and industrial production, particularly for highly complex or demanding tasks. Specifically, even though the tested users perceived similar workloads when maneuvering the virtual and physical cobots, our virtual simulation entailed shorter operation times and lower implicit workloads compared to the physical task. These results apply to VR users with relatively low experience who did not have any knowledge of or experience with robots. This thus suggests that even nonexperts can benefit from the advantages of employing virtual simulations in HRC frameworks.

Nonetheless, we acknowledge the following limitations. First, two different eye-tracking systems were employed in the two conditions. Although the PupilLabs device was deployed in the physical condition, the Tobii eye-tracking system integrated into the HTC Vive Pro Eye was deployed in the virtual condition. Therefore, even though we applied a proper baseline correction (Mathôt et al. 2018), there might still be slight differences in the proprietary algorithms used by different systems to acquire eye-tracking data. Second, on the applicability of our findings in the field, it is important to mention that the effectiveness of a

virtual simulation depends not only on the realism of the virtual environment but also on the quality of the computerized tools employed. For instance, technical features of the head mounted display (HMD) such as the visual field of view, resolution, and latency of the graphical interface might influence the user's performance with the virtual cobot. Therefore, in view of a large-scale implementation of virtual devices into the industrial domain, highly immersive and advanced virtual device adoption is fundamental. Third, even though we tried to leave out any differences between the virtual and the physical pick and place actions, some procedural differences between the two task flows were still present. For instance, the arithmetic sums were reported verbally in the physical condition, whereas they were reported on a virtual keyboard in the virtual condition. Moreover, the pick and place actions were initiated via controller buttons in VR, whereas they were initiated by the user physically touching the worktable in the physical condition. Finally, our conclusions were gleaned from a pool of young users with relatively little experience. Although this choice was motivated by the desire to maintain a homogeneous sample, it comes at the cost of possible low generalizability.

Considering that active industrial operators usually fall within a much greater age range, further investigations might increase the experimental sample and include both young and senior users. In this way, it would be possible to understand better whether the advantages of the virtual simulation also extend to older people. Additionally, based on the finding that even a scant knowledge of VR devices is sufficient for revealing the advantages of virtual simulations, it would be interesting to test whether these advantages increase as the VR experience also grows. Another crucial point arises when addressing the task complexity. The choice of such an easy task, such as the pick-and-place one, was intentional to allow a highly naturalistic investigation without constraining the users' actions and at the same time ensure good experimental control. Future works might gradually increase the task complexity and evaluate to what extent the virtual simulation is still preferable over a physical cobot. A systematic assessment of increasingly demanding tasks would also provide relevant knowledge on the applicability of pupil size variation as an implicit workload index in different levels of task complexity. If future research proves pupillometry to be a reliable and flexible index of implicit workload—in either virtual or physical environments or both—it would become feasible for systems to auto-adjust the cobot's behavior based on human pupil responses. Therefore, a better understanding of pupil changes according to HRC difficulty would have relevant implications on the robotic automation domain.

Overall, this research has just started to shed light on the potential of virtual simulations within HRC frameworks. With the introduction of VR devices in the industry, the

design, validation, training, and even active operations on cobots can definitely take a turn for the better, with the humans' mental and physical health being the cradle of faster and safer interactions between humans and robots.

Appendix

Analysis of pupil size variation through GAMMs

About the pick-and-place's GAMMs, we used binary difference smooths and tested whether they significantly differed from zero (Wieling 2018; Rij et al. 2019). We thus included a smooth modeling the difference between the single-task and dual-task original smooths (Task load), as well as for the Virtual and Physical conditions' smooths (Interface). As it was demonstrated to significantly improve the model fit (van Rij et al. 2019), we included a random smooth of time for each individual time series which we coded as *Event* (unique combination of Participant and Trial) and we additionally introduced random factors smooths for Participant and Trial in order to lighten the computational demand (van Rij et al. 2019). We then used the fREML smoothing parameter estimation method (fast restricted maximum likelihood) and discretized covariates. The statistical analysis was performed in RStudio (Team 2021) using the packages *mgcv* (Wood 2017) and *itsadug* (van Rij et al., 2017) for evaluation and visualization of the GAMM models.

Results of pupil size variation through GAMMs

The summary statistics of the GAMMs fitting the pupil data indicate that the following difference curves are significantly different from zero in the Pick phase: Task load ($F(3.52, 4.18) = 30.30$, $p < 0.0001$), Interface ($F(7.45, 8.53) = 125.72$, $p < 0.0001$), and Task load * Interface ($F(5.11, 6.04) = 16.24$, $p < 0.0001$). Similarly, in the Place phase, the following difference curves were find to differ from zero: Task load ($F(2, 2) = 11.10$, $p < 0.0001$), Interface ($F(5.04, 5.98) = 5.9$, $p < 0.0001$) and Task load * Interface ($F(2, 2) = 87.87$, $p < 0.0001$). When running the chunk analysis over the 6 time windows, all results obtained through the LMERs were in line with those obtained through the GAMMs, except for the effect of Task load on the pupil size variation registered during the Place phase, which was only captured by the GAMM model but not by the LMER.

Author contributions Federica Nenna and Luciano Gamberini took part in conceptualization; Federica Nenna and Luciano Gamberini involved in methodology; Federica Nenna and Davide Zanardi involved in formal analysis and investigation; Federica Nenna and Valeria Orso took part in writing—original draft preparation; Federica Nenna, Valeria Orso, Luciano Gamberini involved in writing—review and editing; Luciano Gamberini took part in resources; Luciano Gamberini took part in supervision.

Data availability Data will be made available to interested researchers under reasonable request.

Declarations

Conflict of interests The authors have no relevant financial or non-financial interests to disclose.

References

Abidi MH, Al-Ahmari A, Ahmad A, Ameen W, Alkhalefah H (2019) Assessment of virtual reality-based manufacturing assembly training system. Int J Adv Manuf Technol 105(9):3743–3759

Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Beatty J (1982) Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychol Bull 91(2):276

Berg LP, Vance JM (2017) Industry use of virtual reality in product design and manufacturing: a survey. Virtual Reality 21(1):1–17

Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. In: KDD workshop 10(16): 359–370

Berni A, Borgianni Y (2020) Applications of virtual reality in engineering and product design: why, what, how, when and where. Electronics 9(7):1064

Bevana N, Kirakowskib J, Maissela J (1991) What is usability. In:Proceedings of the 4th International Conference on HCI.

Billings CE (2018) Aviation automation: the search for a human-centered approach. CRC Press.

Bonferroni C (1936) Teoria statistica delle classi e calcolo delle probabilita. Pubbl Del R Ist Super Sci Econ Commer Firenze 8:3–62

Brooke J (1996) SUS-A quick and dirty usability scale. Usability Eval Ind 189(194):4–7

Chacón A, Ponsa P, Angulo C (2021) Usability study through a human-robot collaborative workspace experience. Designs 5(2):35

Cherubini, A., Passama, R., Crosnier, A., Lasnier, A., Fraisse, P. (2016). Collaborative manufacturing with physical human–robot interaction. Robotics and Computer-Integrated Manufacturing, 40, 1-13.

Chowdhury A, Ahtinen A, Pieters R, Vaananen K (2020) User experience goals for designing industrial human-cobot collaboration: a case study of franka panda robot. In: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society, pp 1–13

Damiani L, Demartini M, Guizzi G, Revetria R, Tonelli F (2018) Augmented and virtual reality applications in industrial systems: a qualitative review towards the industry 4.0 era. IFAC-PapersOnLine 51(11):624–630

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 13:319–340

Fratczak P, Goh YM, Kinnell P, Soltoggio A, Justham L (2019) Understanding human behaviour in industrial human-robot interaction by means of virtual reality. In: Proceedings of the Halfway to the Future Symposium 2019. pp 1–7

Friard O, Gamba M (2016) BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods Ecol Evol 7(11):1325–1330

de Greef T, Lafeber H, van Oostendorp H, Lindenberg, J (2009) Eye movement as indicators of mental workload to trigger adaptive automation. In: International Conference on Foundations of Augmented Cognition, Springer, Berlin, Heidelberg, pp 219–228

Hansen LIN, Vinther N, Stranovsky L, Philipsen MP, Wu H, Moeslund TB (2018) Collaborative meat processing in virtual reality: evaluating perceived safety and predictability of robot approach. In: International Conference on Human Robot Interaction (HRI 2018). VAM-HRI

Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Advances in psychology, vol 52. Elsevier, Amsterdam, The Netherlands, pp 139–183

Hastie TJ, Tibshirani RJ (2017) Generalized additive models. Routledge

Hockey GRJ (1997) Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol Psychol 45(1–3):73–93

Hormaza LA, Mohammed WM, Ferrer BR, Bejarano R, Lastra JLM (2019) On-line training and monitoring of robot tasks through virtual reality. In: 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), 1: 841–846, IEEE

Hsieh SY, Lu JM (2018) Feasibility evaluation for immersive virtual reality simulation of human-machine collaboration: a case study of hand-over tasks. In: Congress of the International Ergonomics Association, Springer, Cham, pp 364–369,

Inoue K, Nonaka S, Ujiie Y, Takubo T, Arai T (2005) Comparison of human psychology for real and virtual mobile manipulators. In: ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005. pp 73–78, IEEE.

Iqbal ST, Zheng XS, Bailey BP (2004) Task-evoked pupillary response to mental workload in human-computer interaction. In: CHI'04 extended abstracts on Human factors in computing systems, pp 1477–1480

Kagermann H (2015) Change through digitization—Value creation in the age of Industry 4.0. In: Management of permanent change, Springer Gabler, Wiesbaden, pp 23–45

Kaufeld M, Nickel P (2019) Level of robot autonomy and information aids in human-robot interaction affect human mental workload—an investigation in virtual reality. In: International Conference on Human-Computer Interaction, Springer, Cham, pp 278–291

Keogh EJ, Pazzani MJ (2001) Derivative dynamic time warping. In: Proceedings of the 2001 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, pp 1–11

Kramer AF (2020) Physiological metrics of mental workload: a review of recent progress. In: Multiple-task performance, pp 279–328

Krenn B, Reinboth T, Gross S, Busch C, Mara M, Meyer K, Layer-Wagner T (2021) It's your turn!--a collaborative human-robot pick-and-place scenario in a virtual industrial setting. arXiv preprint arXiv:2105.13838

Kret ME, Sjak-Shie EE (2019) Preprocessing pupil size data: guidelines and code. Behav Res Methods 51(3):1336–1342

Li R, van Almkerk M, van Waveren S, Carter E, Leite I (2019) Comparing human-robot proxemics between virtual reality and the real world. In: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, pp 431–439

Linn C, Bender S, Prosser J, Schmitt K, Werth D (2017) Virtual remote inspection—a new concept for virtual reality enhanced real-time maintenance. In: 2017 23rd International Conference on Virtual System & Multimedia (VSMM), IEEE, pp 1–6

Lipton JI, Fay AJ, Rus D (2017) Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing. IEEE Robot Autom Lett 3(1):179–186

Liu H, Wang L (2020) Remote human–robot collaboration: a cyber–physical system application for hazard manufacturing environment. J Manuf Syst 54:24–34

Malik AA, Masood T, Bilberg A (2020) Virtual reality in manufacturing: immersive and collaborative artificial-reality in design of human-robot workspace. Int J Comput Integr Manuf 33(1):22–37

Mara M, Meyer K, Heiml M, Pichler H, Haring R, Krenn B, Layer-Wagner T (2021) CoBot studio VR: a virtual reality game environment for transdisciplinary research on interpretability and trust in human-robot collaboration.

Marangunić N, Granić A (2015) Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 14(1):81–95

Martín-Barrio A, Roldán JJ, Terrile S, del Cerro J, Barrientos A (2020) Application of immersive technologies and natural language to hyper-redundant robot teleoperation. Virtual Reality 24(3):541–555

Mathôt S (2018) Pupillometry: psychology, physiology, and function. J Cogn. https://doi.org/10.5334/joc.18

Mathôt S, Fabius J, Van Heusden E, Van der Stigchel S (2018) Safe and sensible preprocessing and baseline correction of pupil-size data. Behav Res Methods 50(1):94–106

Matsas E, Vosniakos GC (2017) Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. Int J Interact Des Manuf (IJIDeM) 11(2):139–153

Matsas E, Vosniakos GC, Batras D (2018) Prototyping proactive and adaptive techniques for human-robot collaboration in manufacturing using virtual reality. Robot Comput Integr Manuf 50:168–180

Meijman TF, Mulder G (2013) Psychological aspects of workload. In: A handbook of work and organizational psychology, Psychology Press, pp 15–44

Melluso N, Fareri S, Fantoni G, Bonaccorsi A, Chiarello F, Coli E, Manafi S (2020) Lights and shadows of COVID-19, Technology and Industry 4.0. arXiv preprint arXiv:2004.13457

Mingardi M, Pluchino P, Bacchin D, Rossato C, Gamberini L (2020) Assessment of implicit and explicit measures of mental workload in working situations: implications for industry 40. Appl Sci 10(18):6416

Nachreiner F, Nickel P, Meyer I (2006) Human factors in process control systems: the design of human–machine interfaces. Saf Sci 44(1):5–26

Navon D, Miller J (1987) Role of outcome conflict in dual-task interference. J Exp Psychol Hum Percept Perform 13(3):435

Nee AY, Ong SK (2013) Virtual and augmented reality applications in manufacturing. IFAC Proc Vol 46(9):15–26

Oyekan JO, Hutabarat W, Tiwari A, Grech R, Aung MH, Mariani MP, Dupuis C (2019) The effectiveness of virtual environments in developing collaborative strategies between industrial robots and humans. Robot Comput Integr Manuf 55:41–54

Peters C, Yang F, Saikia H, Li C, Skantze G (2018) Towards the use of mixed reality for hri design via virtual robots. In: 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI), Cambridge, UK, March 23, 2020

Prattico FG, Lamberti F (2021) Towards the adoption of virtual reality training systems for the self-tuition of industrial robot operators: a case study at KUKA. Comput Ind 129:103446

Riek LD (2012) Wizard of oz studies in hri: a systematic review and new reporting guidelines. J Hum Robot Interact 1(1):119–136

Roldán JJ, Crespo E, Martín-Barrio A, Peña-Tapia E, Barrientos A (2019) A training system for Industry 4.0 operators in complex assemblies based on virtual reality and process mining. Robot Comput Integr Manuf 59:305–316

Rossato C, Pluchino P, Cellini N, Jacucci G, Spagnolli A, Gamberini L (2021) Facing with collaborative robots: the subjective experience in senior and younger workers. Cyberpsychol Behav Soc Netw 24(5):349–356

Savur C, Kumar S, Sahin F (2019) A framework for monitoring human physiological response during human robot collaborative task. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), IEEE, pp 385–390

Schrepp M (2015) User experience questionnaire handbook: All you need to know to apply the ueq successfully in your projects. URL http://www.ueq.online.org

Shirzad N, Van der Loos HM (2016) Evaluating the user experience of exercising reaching motions with a robot that predicts desired movement difficulty. J Mot Behav 48(1):31–46

Team (2021) RStudio: integrated development environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Van Acker BB, Bombeke K, Durnez W, Parmentier DD, Mateus JC, Biondi A, Vlerick P (2020) Mobile pupillometry in manual assembly: a pilot study exploring the wearability and external validity of a renowned mental workload lab measure. Int J Ind Ergon 75:102891

Van Orden KF, Limbert W, Makeig S, Jung TP (2001) Eye activity correlates of workload during a visuospatial memory task. Hum Factors 43(1):111–121

van Rij J, Hendriks P, van Rijn H, Baayen RH, Wood SN (2019) Analyzing the time course of pupillometric data. Trends in Hear 23:2331216519832483

Wang L, Mohammed A, Onori M (2014) Remote robotic assembly guided by 3D models linking to a real robot. CIRP Ann 63(1):1–4

Wang H, Zhang B, Zhang T, Jakacky A (2019) Tele-operating a collaborative robot for space repairs with virtual reality. In: 2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), IEEE, pp 175–180

Weistroffer V, Paljic A, Fuchs P, Hugues O, Chodacki JP, Ligot P, Morais A (2014) Assessing the acceptability of human-robot co-presence on assembly lines: a comparison between actual situations and their virtual reality counterparts. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, IEEE, pp 377–384

Whitney D, Rosen E, Ullman D, Phillips E, Tellex S (2018) Ros reality: a virtual reality framework using consumer-grade hardware for ros-enabled robots. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) . IEEE, pp 1–9

Wieling M (2018) Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. J Phon 70:86–116

Wood SN (2017) Generalized additive models: an introduction with R. CRC Press

Xiao J, Wang P, Lu H, Zhang H (2020) A three-dimensional mapping and virtual reality-based human–robot interaction for collaborative space exploration. Int J Adv Rob Syst 17(3):1729881420925293

Zhang J (2018) Natural human-robot interaction in virtual reality telepresence systems. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE, pp 812–813