

An experimental focus on learning effect and interaction quality in human–robot collaboration

Original

An experimental focus on learning effect and interaction quality in human–robot collaboration / Gervasi, Riccardo; Mastrogiacomo, Luca; Franceschini, Fiorenzo. - In: PRODUCTION ENGINEERING. - ISSN 0944-6524. - STAMPA. - 17:3-4(2023), pp. 355-380. [10.1007/s11740-023-01188-5]

Availability:

This version is available at: 11583/2978535 since: 2023-05-16T12:25:41Z

Publisher:

Springer

Published

DOI:10.1007/s11740-023-01188-5

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)



An experimental focus on learning effect and interaction quality in human–robot collaboration

Riccardo Gervasi¹ · Luca Mastrogiacomo¹ · Fiorenzo Franceschini¹

Received: 5 December 2022 / Accepted: 23 January 2023 / Published online: 30 January 2023
© The Author(s) 2023

Abstract

In the landscape of the emerging Industry 5.0, human–robot collaboration (HRC) represents a solution to increase the flexibility and reconfigurability of production processes. Unlike classical industrial automation, in HRC it is possible to have direct interaction between humans and robots. Consequently, in order to effectively implement HRC it is necessary to consider not only technical aspects related to the robot but also human aspects. The focus of this paper is to expand on previous results investigating how the learning process (i.e., the experience gained through the interaction) affects the user experience in the HRC in conjunction with different configuration factors (i.e., robot speed, task execution control, and proximity to robot workspace). Participants performed an assembly task in 12 different configurations and provided feedback on their experience. In addition to perceived interaction quality, self-reported affective state and stress-related physiological indicators (i.e., average skin conductance response and heart rate variability) were collected. A deep quantitative analysis of the response variables revealed a significant influence of the learning process in the user experience. In addition, the perception of some configuration factors changed during the experiment. Finally, a significant influence of participant characteristics also emerged, auguring the necessity of promoting a human-centered HRC.

Keywords Industry 5.0 · Human–robot collaboration · Human-factors · Physiological signals · Learning process · User experience · Affective state

1 Introduction

The emerging Industry 5.0 focuses attention on human-centricity, promoting the development and implementation of technologies that can support humans in their activities. Human–robot collaboration represents one solution for achieving these goals. Human–robot collaboration is a form of direct interaction between human and robotic system apt to combine the abilities of the entities involved to achieve a common goal [1]. On the one hand, robots support with power, precision, and repeatability; on the other

hand, humans have flexibility, problem-solving capabilities, and intelligence. The robots involved in HRC are called collaborative robots (or cobots), which, unlike classical industrial robots, are special robots designed to allow physical interaction with human operators. This feature allows for the removal of barriers separating the workspace of robots from that of operators, making the industrial layout more flexible and enabling new forms of interaction [2]. In addition, this perspective allows to promote greater adaptability and reconfigurability of production processes [3, 4]. However, working closely with a robot can induce stressful situations for operators, impacting the performance and quality of production processes. Such situations can impact the performance and quality of production processes, as well as the well-being of the operator [5]. From the perspective of Industry 5.0, in order to promote a human-centred industry, it is therefore necessary also to analyze and take human factors into consideration to effectively implement HRC [6, 7].

Understanding the operator mental and physical state during HRC is first required in order to provide the optimal support for their wellbeing. In addition to traditional

✉ Fiorenzo Franceschini
fiorenzo.franceschini@polito.it
Riccardo Gervasi
riccardo.gervasi@polito.it
Luca Mastrogiacomo
luca.mastrogiacomo@polito.it

¹ DIGEP (Department of Management and Production Engineering), Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129 Turin, Italy

evaluation tools, such as questionnaires, the analysis of physiological response is a valuable resource for investigating user experience. Through physiological response, it is possible to monitor the operator's state as the task is being carried out in real-time, even discovering potential unconscious reactions. Only a relatively small number of publications have so far included physiological measurements to comprehend the operator's state in HRC [8–10]. In addition, a gap is present in the literature on how user experience and perception of HRC factors may evolve over time due to the experience gained from interacting with the cobot. The study of these aspects can be helpful for outlining guidelines for the introduction of a collaborative task to an operator.

In a previous work [11], an original experimental setting aimed at investigating how different configurations of a cobot impact the user experience was presented, involving 42 participants. The objective of this paper is to expand on previous findings by addressing the following research question: *How do the learning process influence operator's user experience and stress?*

Through a deep quantitative analysis, the joint effects of learning process (i.e., the experience gained by interacting with the cobot during the experiment) and different HRC configuration factors on self-reported affective state, perceived interaction quality and physiological response are investigated. In addition, the influence of operator characteristics (e.g., gender, age, previous experience with cobots, and attitude towards robots) on the different response variables is explored.

The main contributions of this paper can be summarized as follows:

- (i) A deep quantitative analysis on the learning process and its joint effect with different HRC configuration factors on response variables (i.e., perceived interaction quality, affective state, and physiological response).
- (ii) Exploration of the influence of operator characteristics on response variables.

The findings of this study can have an impact on the design of collaborative tasks as well as the way they are implemented.

The paper structure is as follows. A review on HRC and the human factors involved is provided in Sect. 2. Section 3 provides a summary of the methodology and experimental setting. Section 4 presents the quantitative analysis results, highlighting the relationships with the response variables of the joint effect of learning process and configuration factors. In Sect. 5, discussion of the study results and potential implications for the improvement of HRC are presented. Finally, Sect. 6 focuses on conclusions and future work.

2 Literature analysis

HRC paradigm is characterized by several aspects related to both the robotic system and humans [6]. For the interaction to effectively support humans in the most difficult operations, careful planning is necessary [12]. Inkulu et al. [13] offered an overview of HRC, outlining some of the critical challenges and prospects. Natural ways of communicating with robots, such as voice and gestures, enable natural engagement and may cut down on idle time, however these recognition techniques need to be strengthened to withstand any environmental disturbances. Power force limitation strategies are helpful for effectively collaborating with low-payload robots, but they might not be appropriate for high-speed and high-payload robots, necessitating the deployment of additional flexible safety methods. Collaborative robots are enabling technologies for reconfigurable production systems, however, to minimize potential production downtime, more research is required on robot adaptive systems.

Human factors related to HRC have received more attention in recent years [14]. Psychological and cognitive ergonomics are equally important to the successful implementation of HRC as physical ergonomics [15, 16]. Gualtieri et al. [17] presented an overview on human factors and cognitive ergonomics aspects involved the design of HRC assembly systems. Aspects such as cognitive workload, stress, usability, perceived enjoyment, acceptance, trust, and frustration were highlighted to be of particular interest in the enhancement of operator's work conditions.

The idea of symbiotic HRC was introduced by Wang et al. [18]. In tradition automation practice, humans are required to adhere to rigid work processes. Symbiotic HRC aims at promoting: (i) natural interaction with the robot; (ii) multi-modal, user-friendly programming environment that doesn't demand in-depth system expertise; (iii) an increased context dependency; (iv) an immersive collaboration that enables the operator to participate in the tasks through wearable technology (e.g., smart watches, augmented reality (AR) glasses).

Recent efforts have concentrated on presenting techniques for improving the adaptiveness of robot systems during HRC in order to enhance HRC potentials [19, 20]. Neves and Neto [21] proposed a reinforcement learning approach for assembly sequence planning that included user preferences. Mohammed et al. [22] introduced a method for efficient online collision avoidance in an augmented environment. Buerkle et al. [23] presented a sensor framework to model humans, which incorporated subjective, objective, and physiological metrics.

Literature has extensively covered concepts like stress, fatigue, and mental workload which are particularly

relevant in the context of manufacturing [5, 24, 25]. Self-reporting methods like the Subjective Workload Assessment Technique (SWAT) [26] and the NASA-TLX [27] are frequently used to assess these dimensions [28]. However, these methods are not well suited to gather real-time data and requires the subject to recall an event, which may introduce some bias [29, 30]. In order to work through these constraints, recent years have seen a growing emphasis on physiological measures for the study of the operator's condition [31, 32]. Various works on this topic have been presented, but only a limited fraction is specifically focused on industrial HRC. Koppenborg et al. [33] investigated the effects of movement speed and path predictability on the operator through virtual reality. In addition to subjective responses, heart rate was collected by a chest band. Arai et al. [8] used an industrial manipulator to assess how various speeds and distances from the operator affected mental strain, measured through electrodermal activity (EDA). Kühnlenz et al. [9] investigated how various robot trajectory profiles affected operators stress by analyzing EDA and heart rate variability (HRV).

3 Methodology

3.1 Experimental setup

In the “Mind 4 Lab” (Manufacturing Industry 4.0 Laboratory) of the “Politecnico di Torino”, a collaborative assembly task was implemented to emulate an HRC setting [11]. Using the assistance of the cobot UR3e, the task involved attaching two mechanical flanges to a base by tightening two pairs of screws (Fig. 1). The operations can be divided in the following phases:

- (1) The cobot picks the square flange and places it in the correct position on the base.
- (2) The operator takes the screws, inserts them into the holes and tightens them.
- (3) The cobot takes the oval flange and places it correctly on the base.
- (4) The operator takes the other two screws, inserts them into the holes and tightens them.
- (5) The cobot takes the assembled mechanical component and places it in another work area.

A within-subjects experimental design was implemented in this study to examine the effects of three fixed factors (i.e., robot's movement speed, proximity to robot workspace, and execution time control) and their interactions with the learning process on affective state, interaction quality, and physiological response [11].

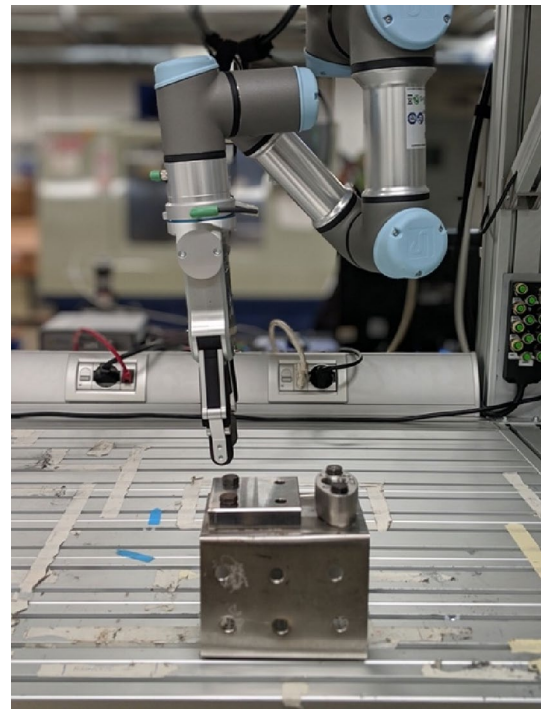


Fig. 1 Collaborative robot UR3e and the mechanical component

Three levels of the robot's joint speed (*Speed*) were implemented: 30°/s (*Low*), 90°/s (*Medium*), and 270°/s (*High*). These values represent the maximum speed that all the robot's joints could reach. Two levels of proximity of the robot workspace to the operator (*Distance*) were introduced: 30 cm (*Close*) and 40 cm (*Far*). The distances refer to the minimum distance between the operator's chest and the robot workspace. Lastly, two levels of control for the task execution time (*Control*) were implemented: one in which the operator had a push-button to command the robot to continue with the task (*Human*) and another in which the cobot proceeded automatically with the operations after waiting 25 s (*NoHuman*). Each participant performed the task in all 12 possible configurations in random order.

Forty-two participants, with an average age of 28.2 years (standard deviation = 8.1), were recruited from the “Politecnico di Torino” and the surroundings for the study (28.6% females and 71.4% males). In order to collect information about the participants and their user experience, a number of questionnaires were implemented. Before beginning the experiment, each participant was given an initial questionnaire to collect age, gender, and degree of prior experience with cobots (according to the following scale: L0—“I have never interacted with a cobot and I did not know them before now”; L1—“I have never interacted with a cobot but I know what they are”; L2—“I have interacted at least once with a cobot”; L3—“I have already programmed and interacted with a cobot”).

Afterwards, the Negative Attitude toward Robots Scale (NARS) [34] was administered to assess general attitude toward robots of the participant. The NARS items are divided in the following three sub-scales and, for each of them, a score can be obtained: *Negative Attitudes toward Situations and Interactions with Robots* (S1) (scoring between 6 and 30); *Negative Attitudes toward Social Influence of Robots* (S2) (scoring between 5 and 25), and *Negative Attitudes toward Emotions in Interaction with Robots* (S3) (scoring between 3 and 15) [34].

At the end of each trial, the Self-Assessment Manikin (SAM) [35, 36] was administrated to participants to gather their affective state in the different task configurations by evaluating on a 9-point scale three dimensions: valence, arousal, and dominance. Valence represents the pleasantness relative to a stimulus (for instance, happiness and relaxation are associated with a high valence, while anxiety or anger with a low valence). Arousal refers to the intensity of emotion provoked by a stimulus (e.g., fear and anger are usually associated with a high arousal, while relaxation and boredom with a low arousal). Dominance represents the degree of control felt relative to a stimulus (e.g., relaxation or anger are usually associated with a high dominance, while fear or anxiety with low dominance).

In addition, at the end of each trial, an interaction quality questionnaire (Table 1) based on Hoffman [37] and Baraglia et al. [38] and composed of seven items was used to collect participants' perception on different dimensions related to the interaction with the cobot [11]. The items were evaluated on a 7-point scale (from “strongly disagree” to “strongly agree”).

Finally, at the end of the experiment, participants were asked to provide overall unstructured feedback regarding the experience.

In addition to subjective evaluations, physiological signals were also collected to deepen the participant state during the experimental trials. EDA data and heart data through Photoplethysmogram (PPG) were obtained, respectively, at 4 Hz and 64 Hz using the non-invasive biosensor Empatica E4 wristband [39]. From EDA and PPG signals stress and arousal indicators were derived for each HRC configuration, as explained in the following sub-section.

3.2 Data processing and analysis

Table 2 provides a summary of all the dependant and independent variables included in the analysis. From physiological signals, potential artifacts were identified and removed. By using the MATLAB-software “Ledalab”, EDA signal were decomposed in tonic component and phasic component through Continuous Decomposition Analysis (CDA) [40]. The tonic component is characterized by the Skin Conductance Level (SCL), which represents the long-term fluctuations in EDA that are not directly derived by external stimuli. Short-term EDA fluctuations elicited by an external stimulus represents phasic component. From the phasic component, Skin Conductance Responses (SCRs), which are amplitude differences between the SCL and response peaks, are detected. In the present study, the average SCR (*Mean-SCR*) was calculated for each HRC configuration, representing a stress and arousal indicator. Regarding heart data, NN-intervals (i.e., time intervals between systolic peaks) were obtained from PPG. As HRV measure for stress, the Root Mean Square of Successive Differences between adjacent NN-intervals (RMSSD) was included due to its widespread usage [5, 41].

A series of Mixed-effect Ordinal Logistic Regression (MOLR) models were implemented in order to investigate the relationship of the fixed-factors of the experiment and their interactions with the subjective responses (i.e., interaction quality and SAM dimensions). The MOLR model was chosen for its suitability in (i) modelling dependent variables defined on an ordinal scale, and (ii) handling the participant effect as a random block effect. The “ordinal” package from the software R was used to fit these models [42]. Additional details on MOLR are provided in Appendix A.

Since the selected indicators for physiological response were continuous variables, Linear Mixed Models (LMMs) were implemented to explore the relationship with the considered factors and to handle the participant effect as a random block effect. Models were fitted using the “lmerTest” package from the software R.

The formula used for the models is reported below using the Wilkinson notation [43]:

Table 1 Questionnaire for interaction quality [11]

Item no	Dimension	Questionnaire item
Q1	Robot helpfulness	The robot was helpful in accomplishing the task
Q2	Interaction unsafety	I felt the interaction was not safe
Q3	Interaction naturalness	The collaboration felt natural
Q4	Team efficiency	The robot and I worked efficiently together
Q5	Team fluency	The robot and I worked fluently together
Q6	Discomfort	I felt uncomfortable with the robot
Q7	Robot trustworthiness	The robot was trustworthy

Table 2 Summary of the variables

Type of variable	Variable name	Description
Independent variables	<i>Speed</i>	Robot movement speed (<i>Low, Medium, High</i>)
	<i>Distance</i>	Proximity of the operator to the robot workspace (<i>Close, Far</i>)
	<i>Control</i>	Control of the task execution time (<i>Human, NoHuman</i>)
	<i>Trial</i>	Number of the trial performed by the participant. It represents the temporal progress of the experiment
	<i>Participant</i>	ID of the participant
	<i>Age</i>	Age of the participant
	<i>Gender</i>	Gender of the participant
	<i>Experience</i>	Level of prior experience with collaborative robots
	<i>NS1</i>	Score for “Negative Attitudes toward Situations and Interactions with Robots” from NARS
	<i>NS2</i>	Score for “Negative Attitudes toward Social Influence of Robots” from NARS
	<i>NS3</i>	Score for “Negative Attitudes toward Emotions in Interaction with Robots” from NARS
Dependant variables	<i>Q1_Helpful</i>	Assessment of perceived robot helpfulness (7-point scale)
	<i>Q2_NotSafe</i>	Assessment of perceived interaction unsafety (7-point scale)
	<i>Q3_Natural</i>	Assessment of perceived interaction naturalness (7-point scale)
	<i>Q4_Efficient</i>	Assessment of perceived team efficiency (7-point scale)
	<i>Q5_Fluid</i>	Assessment of perceived team fluency (7-point scale)
	<i>Q6_Uncomfortable</i>	Assessment of perceived discomfort (7-point scale)
	<i>Q7_Trustworthy</i>	Assessment of perceived robot trustworthiness (7-point scale)
	<i>Valence</i>	SAM dimension assessing how positive is the emotion (9-point scale)
	<i>Arousal</i>	SAM dimension assessing how much agitated a person feels (9-point scale)
	<i>Dominance</i>	SAM dimension assessing how strong is the dominance feeling (9-point scale)
	<i>MeanSCR</i>	Average of amplitudes of Skin Conductance Responses [μ S]
	<i>RMSSD</i>	Root Mean Square of Successive Differences between adjacent heart rate NN-intervals [ms]

$$Y \sim 1 + Speed * Distance * Control + Trial * Speed + Trial * Distance + Trial * Control + (1|Participant) \quad (1)$$

To explore the potential effect of participant characteristics, the same models were also fitted including the independent variables *Gender*, *Age*, *Experience*, *NS1*, *NS2*, and *NS3*.

4 Results

This section contains the analysis results for each considered response variable.

4.1 Perceived interaction quality

The relationships between the learning process, the experimental factors and the different aspects of the perceived quality of interaction with the robot will be discussed in this sub-section. In Appendix B, the estimated parameters of the fitted models are reported. Table 3 contains the results of the Analysis of Deviance (ANODE) for each MOLR model, showing the significance of each term. Figures 3, 4, 5, 6,

7, 8, and 9 provide a representation of the effects of the independent variables on the interaction quality dimensions. For each configuration of the fixed factors (i.e., *Speed*, *Distance*, and *Control*), the evolution of the expected response probabilities over the course of the experiment is shown. Through this representation, it is also possible to notice the significance of the fixed factors by comparing the patterns of the configurations (e.g., in Fig. 2, a distinct difference in patterns between the *Human* and *NoHuman* configurations can be seen).

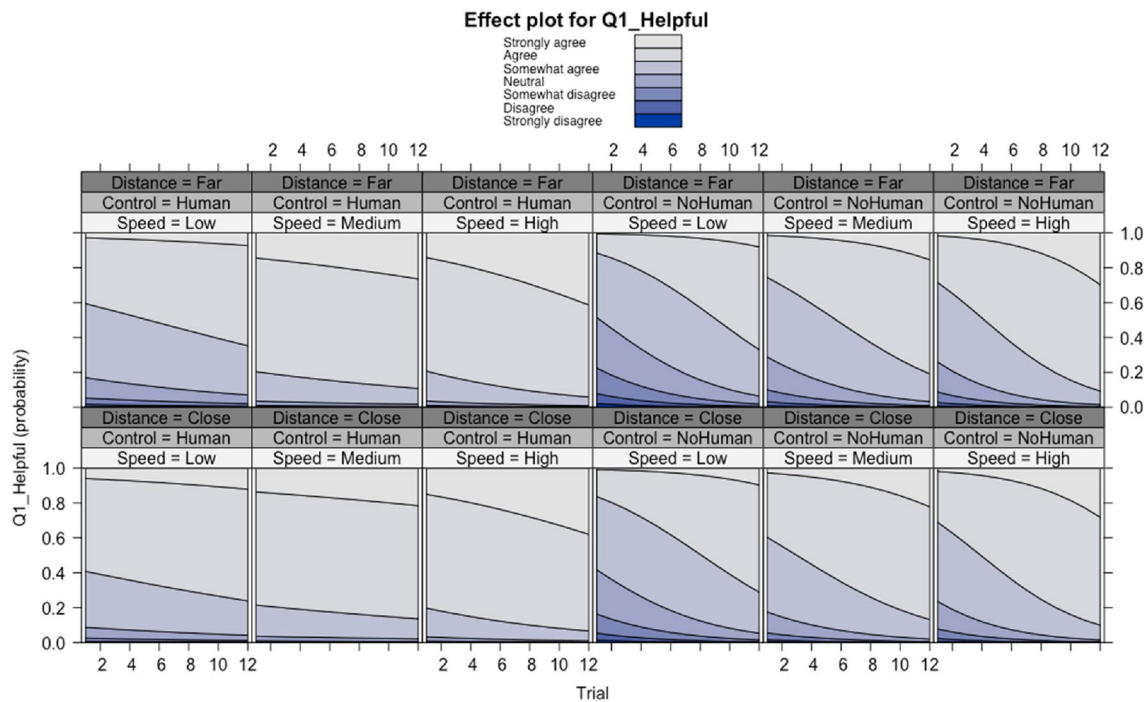
Speed, *Control*, *Trial*, and the interaction term *Control*·*Trial* had a significant effect on the perceived robot helpfulness (*Q1_Helpful*) and their effect can be seen in Fig. 2:

- *Speed* had a general positive effect ($\beta_{SpeedMed}=0.95$, $\beta_{SpeedHigh}=0.99$), meaning that when the robot moved faster it was perceived as being more helpful in completing the task.
- *Trial* had a significant positive effect ($\beta_{Trial}=0.07$), meaning that the perceived robot helpfulness slowly increased due to the learning process.
- The cobot was rated as being less helpful when the person had no control over the task execution time

Table 3 ANODE table with resulting p -values for quality of interaction

Model term	Q1_Helpful	Q2_NotSafe	Q3_Natural	Q4_Efficient	Q5_Fluid	Q6_Uncomf	Q7_Trustw
<i>Speed</i>	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	0.010**	0.413
<i>Distance</i>	0.209	0.259	0.584	0.881	0.643	0.206	0.031*
<i>Control</i>	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	0.001**
<i>Trial</i>	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***
<i>Speed·Distance</i>	0.533	0.470	0.449	0.772	0.637	0.352	0.096
<i>Speed·Control</i>	0.653	0.090	0.171	0.018*	0.018*	0.004**	0.244
<i>Distance·Control</i>	0.696	0.618	0.697	0.366	0.961	0.775	0.700
<i>Speed·Trial</i>	0.631	0.133	0.041*	0.889	0.359	0.023*	0.903
<i>Distance·Trial</i>	0.733	0.342	0.965	0.109	0.261	0.899	0.496
<i>Control·Trial</i>	0.004**	0.621	0.017*	0.005**	0.004**	0.058	0.844
<i>Speed·Distance·Control</i>	0.457	0.411	0.068	0.808	0.343	0.410	0.904

*0.05 > p ≥ 0.01, **0.01 > p ≥ 0.001, *** p < 0.001

**Fig. 2** Effect of the variables on the response probabilities for perceived robot helpfulness ($Q1_Helpful$). For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

($\beta_{ControlNoHum} = -2.17$). However, it is interesting to note that the interaction term $Control \cdot Trial$ had a positive effect ($\beta_{ControlNoHum \cdot Trial} = 0.16$). As the experiment progressed, the negative effect due to the lack of control was partially offset by the progressive acquisition of experience with the collaborative task.

Concerning the interaction unsafety ($Q2_NotSafe$), *Speed*, *Control*, and *Trial* were found to be significant, and Fig. 3 shows their effect on the response probabilities:

- The perceived safety degraded when the robot movement speed was high ($\beta_{SpeedMed} = -0.02$, $\beta_{SpeedHigh} = 0.30$).
- An increase of the perceived unsafety was observed when the participant had no control of task execution time ($\beta_{ControlNoHum} = 1.31$).
- As the experiment progressed (i.e., *Trial* increased), the participant became increasingly familiar with the collaborative task and the perceived unsafety gradually decreased ($\beta_{Trial} = -0.15$).

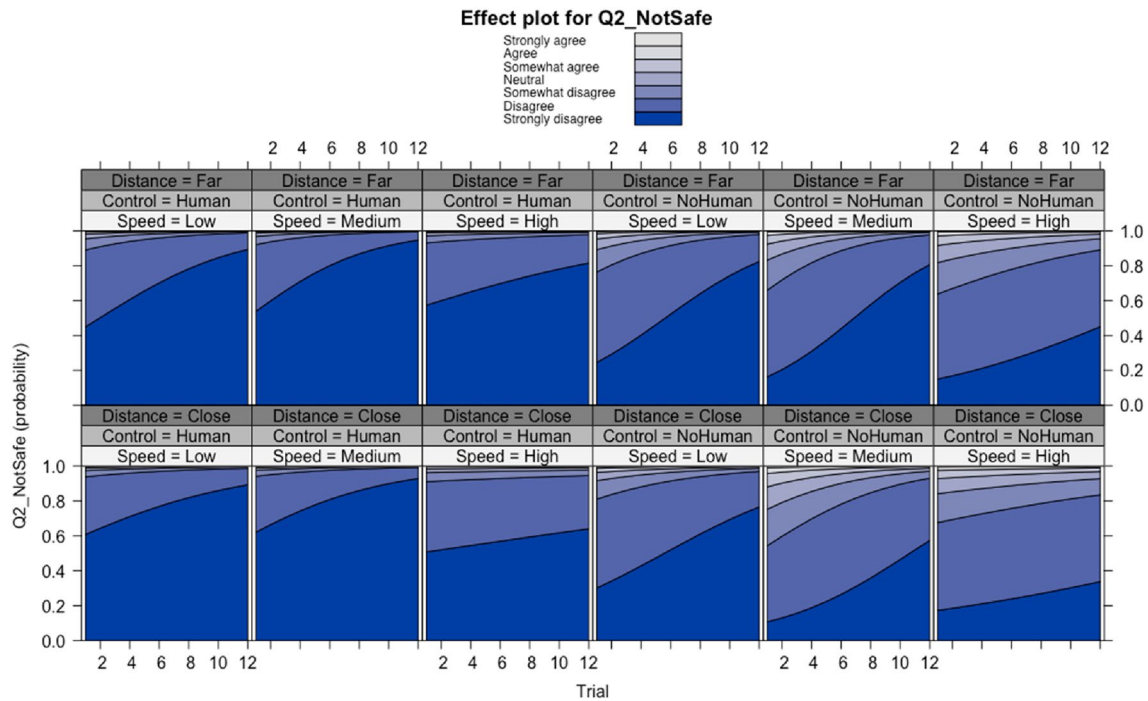


Fig. 3 Effect of the variables on the response probabilities for perceived interaction unsafety ($Q2_NotSafe$). For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

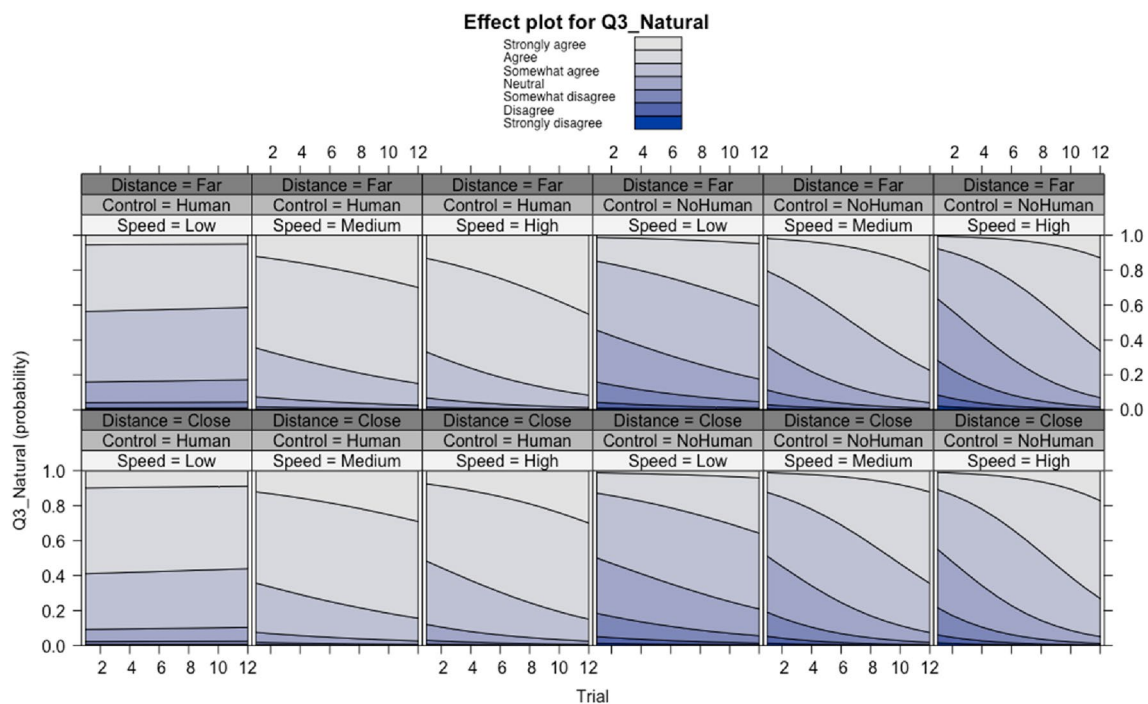


Fig. 4 Effect of the variables on the response probabilities for perceived interaction naturalness ($Q3_Natural$). For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

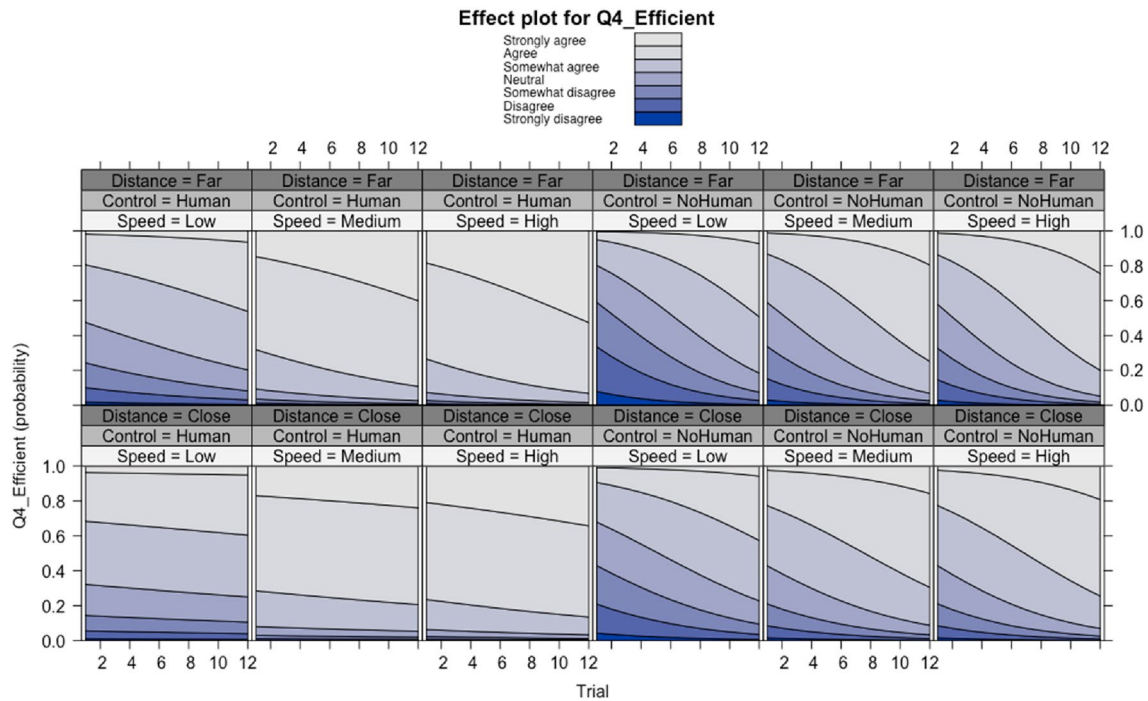


Fig. 5 Effect of the variables on the response probabilities for perceived team efficiency (*Q4_Efficient*). For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

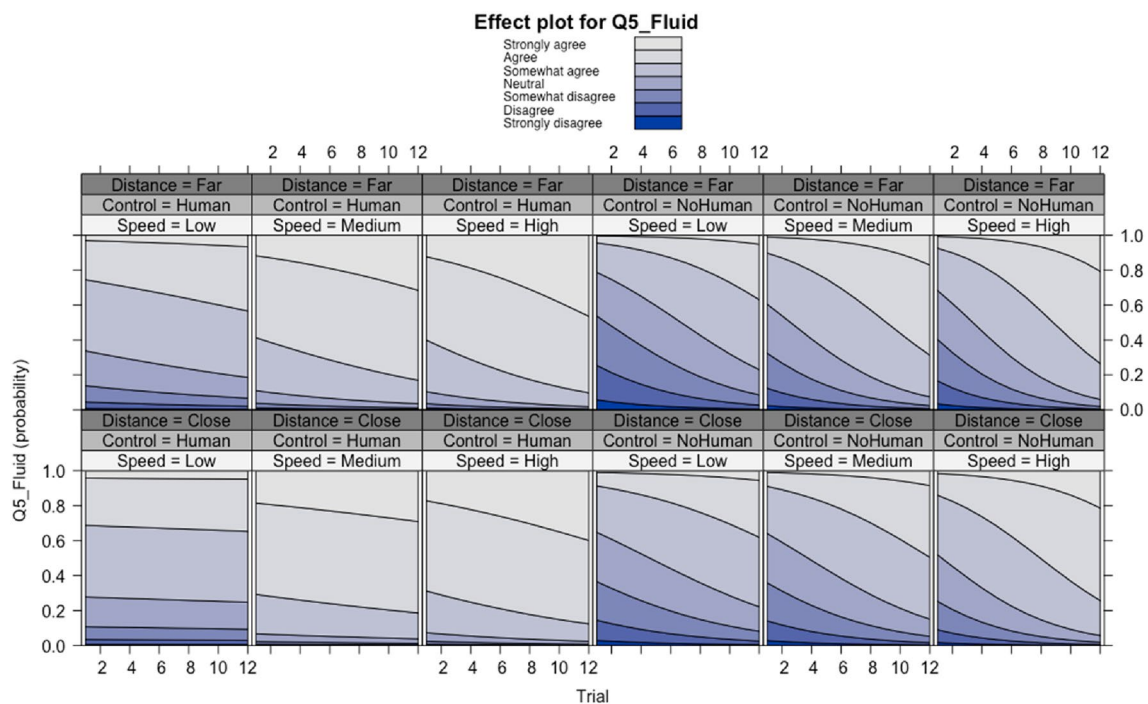


Fig. 6 Effect of the variables on the response probabilities for perceived team fluency (*Q5_Fluid*). For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

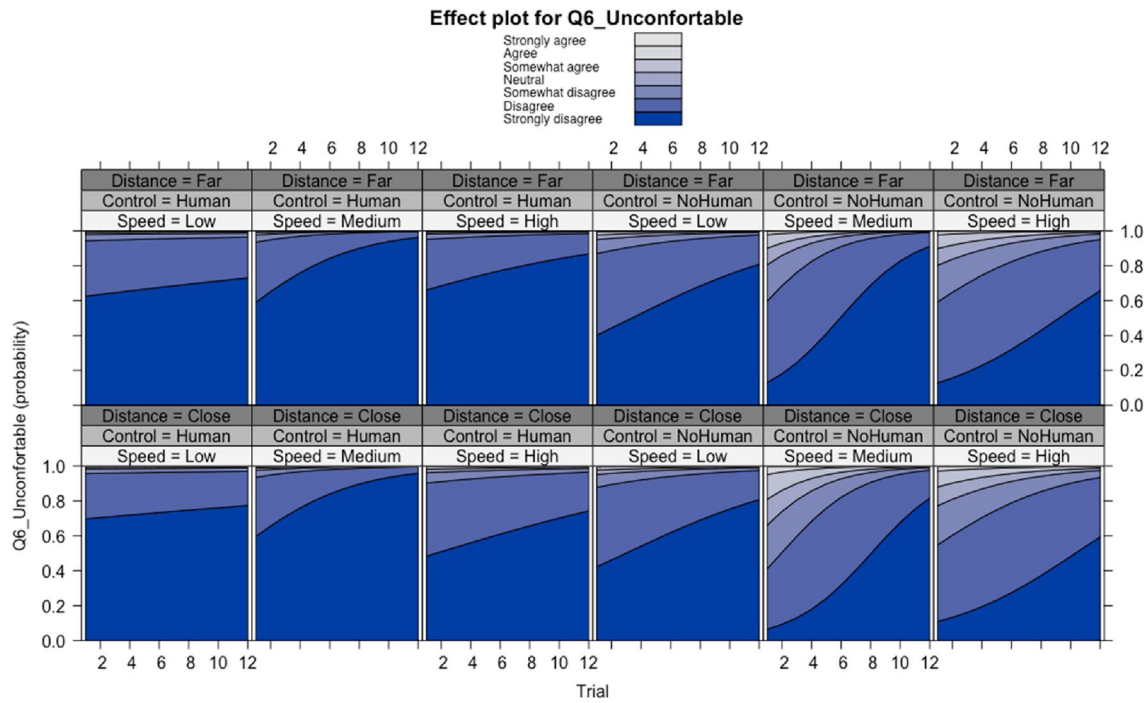


Fig. 7 Effect of the variables on the response probabilities for perceived discomfort ($Q6_Uncomfortable$). For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

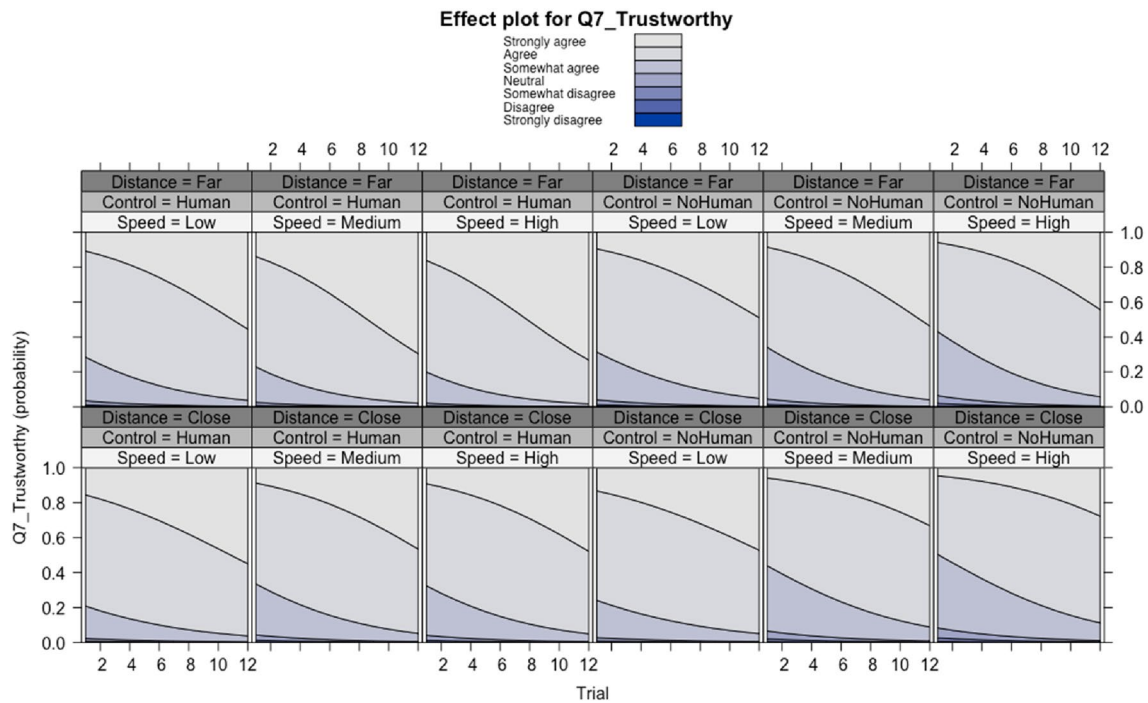


Fig. 8 Effect of the variables on the response probabilities for perceived robot trustworthiness ($Q7_Trustworthy$). For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

Table 4 ANODE table with resulting p -values for each SAM dimension

Model term	Valence	Arousal	Dominance
<i>Speed</i>	<0.001***	<0.001***	0.028*
<i>Distance</i>	0.931	0.797	0.137
<i>Control</i>	<0.001***	<0.001***	<0.001***
<i>Trial</i>	0.034*	<0.001***	<0.001***
<i>Speed·Distance</i>	0.703	0.246	0.569
<i>Speed·Control</i>	0.003**	0.675	0.615
<i>Distance·Control</i>	0.362	0.776	0.137
<i>Speed·Trial</i>	<0.001***	0.374	0.975
<i>Distance·Trial</i>	0.192	0.836	0.472
<i>Control·Trial</i>	0.446	0.195	0.394
<i>Speed·Distance·Control</i>	0.536	0.448	0.598

*0.05 > p ≥ 0.01, **0.01 > p ≥ 0.001, *** p < 0.001

The terms *Speed*, *Control*, *Trial*, *Speed·Trial*, and *Control·Trial* resulted significant in explaining the variability of perceived interaction naturalness with the cobot ($Q3_Natural$) and their effect are shown in Fig. 4:

- The medium speed of the robot had a slightly positive effect on interaction naturalness ($\beta_{SpeedMed} = 0.11$), while the high speed caused a decrease ($\beta_{SpeedHigh} = -0.46$). However, it is interesting to note

that as the participant became more familiar with the task, this negative effect was gradually completely compensated ($\beta_{SpeedMed·Trial} = 0.11$, $\beta_{SpeedHigh·Trial} = 0.16$).

- Participants felt the interaction less natural when they had no control over task execution ($\beta_{ControlNoHum} = -2.42$). However, as the participant became more familiar with the task, this negative effect was partially offset through the learning process ($\beta_{ControlNoHum·Trial} = 0.13$).

Regarding perceived team efficiency ($Q4_Efficient$), the terms *Speed*, *Control*, *Trial*, *Speed·Control*, and *Control·Trial* were found to be significant, and Fig. 5 shows their effect on the response probabilities:

- The robot movement speed had an overall positive effect ($\beta_{SpeedMed} = 1.68$, $\beta_{SpeedHigh} = 1.92$). However, the interaction term *Speed·Control* had a general negative effect ($\beta_{SpeedMed·ControlNoHum} = -0.65$, $\beta_{SpeedHigh·ControlNoHum} = -0.91$), meaning that a higher robot movement speed with the absence of execution time control by the participant implied a decrease in team efficiency.
- *Trial* had a slightly positive effect on team efficiency ($\beta_{Trial} = 0.03$).
- The perceived efficiency was significantly reduced when participants were not in control of task execution time ($\beta_{ControlNoHum} = -1.65$). However, this effect was gradu-

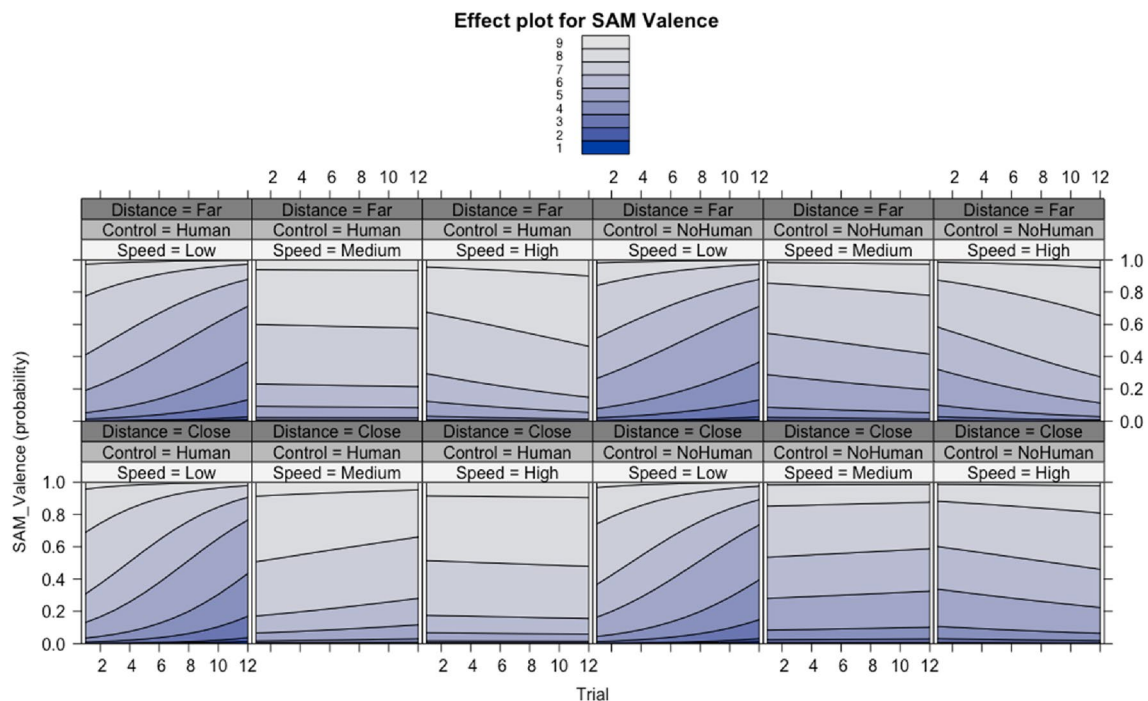


Fig. 9 Effect of the variables on the response probabilities for *Valence*. For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

ally mitigated by the experience gained by the participant as the experiment progressed ($\beta_{ControlNoHum \cdot Trial} = 0.15$).

Regarding team fluency ($Q5_Fluid$), *Speed*, *Control*, *Trial*, *Speed·Trial*, and *Control·Trial* were found to be significant and their effect can be seen in Fig. 6:

- Compared to low robot speed, medium and high robot speeds were related to increased team fluency ($\beta_{SpeedMed} = 1.63$, $\beta_{SpeedHigh} = 1.49$). However, the interaction term *Speed·Control* had a negative effect ($\beta_{SpeedMed \cdot ControlNoHum} = -1.64$, $\beta_{SpeedHigh \cdot ControlNoHum} = -1.04$), meaning that a medium or high robot movement speed with no execution time control by participants was related with a degradation of team efficiency. It is worth noting that a medium robot speed was associated with slightly greater team fluency than a high one, however the penalty due to the absence of execution time control was significantly greater for the medium speed.
- Lack of execution time control had a negative impact on team fluency ($\beta_{ControlNoHum} = -1.72$). However, this degradation was gradually compensated by the learning process of the participant ($\beta_{Trial} = 0.01$, $\beta_{ControlNoHum \cdot Trial} = 0.16$).

The terms *Speed*, *Control*, *Trial*, *Speed·Control*, and *Speed·Trial* resulted significant in explaining the variability of perceived discomfort ($Q6_Uncomfortable$), and Fig. 7 shows their effect:

- The increase of robot movement speed had a general positive effect on discomfort ($\beta_{SpeedMed} = 0.66$, $\beta_{SpeedHigh} = 0.97$). Moreover, the interaction term *Speed·Control* further enhanced this positive effect ($\beta_{SpeedMed \cdot ControlNoHum} = 1.90$, $\beta_{SpeedHigh \cdot ControlNoHum} = 0.89$).
- A greater discomfort was observed in configurations with no control of time execution by participants ($\beta_{ControlNoHum} = 1.27$).
- As the experiment progressed, perceived discomfort gradually decreased ($\beta_{Trial} = -0.04$) mainly due to learning process. This effect was further enhanced by the interaction term *Speed·Trial* for medium and high speeds ($\beta_{SpeedMed \cdot Trial} = -0.22$, $\beta_{SpeedHigh \cdot TrialNo} = -0.07$).

Concerning robot trustworthiness ($Q7_Trustworthy$), the terms *Distance*, *Control*, and *Trial* resulted significant, and their effect are shown in Fig. 8:

- The cobot was felt less trustworthy when a greater distance between the participant and robot workspace was present ($\beta_{DistanceFar} = -0.45$).

Table 5 ANOVA table with resulting p -values for each physiological response indicator

Model term	Mean SCR	RMSSD
<i>Speed</i>	0.024*	0.087
<i>Distance</i>	0.606	0.059
<i>Control</i>	0.054	0.011*
<i>Trial</i>	<0.001***	0.060
<i>Speed·Distance</i>	0.829	0.457
<i>Speed·Control</i>	0.610	0.119
<i>Distance·Control</i>	0.085	0.691
<i>Speed·Trial</i>	0.111	0.080
<i>Distance·Trial</i>	0.423	0.096
<i>Control·Trial</i>	0.061	0.013*
<i>Speed·Distance·Control</i>	0.919	0.276

* $0.05 > p \geq 0.01$, ** $0.01 > p \geq 0.001$, *** $p < 0.001$

- The absence of time execution control by participants implied a degradation of robot trustworthiness ($\beta_{ControlNoHum} = -0.17$).
- *Trial* had a general positive effect ($\beta_{Trial} = 0.17$), meaning that the learning process gradually increased perceived trustworthiness.

Likelihood Ratio (LR) tests with only fixed-term models were performed to check if the random effect *Participant* was significant. From the obtained results, the random effect *Participant* resulted highly significant ($p < 0.001$) for every fitted model.

4.2 Self-reported affective state

This subsection examines how the experimental variables influence the self-reported affective state collected through the SAM. In the Appendix B, the estimated parameters of the fitted models are reported. Table 4 contains the results of the ANOVA for each MOLR model, showing the significance of each model term. Figures 10, 11, and 12 provide a representation of the effects of the independent variables on the interaction quality dimensions.

Regarding *Valence* dimension, the terms *Speed*, *Control*, *Trial*, *Speed·Control*, and *Speed·Trial* were found significant and Fig. 9 shows their effect:

- Medium and high robot speed had a positive effect ($\beta_{SpeedMed} = 0.55$, $\beta_{SpeedHigh} = 0.45$), resulting in generally more pleasant emotions. Moreover, as the experiment progressed this effect was gradually increased ($\beta_{SpeedMed \cdot Trial} = 0.22$, $\beta_{SpeedHigh \cdot TrialNo} = 0.29$).
- Configurations with no time execution control by participants led to less pleasant emotions ($\beta_{ControlNoHum} = -0.30$). Moreover, this negative effect was further enhanced with

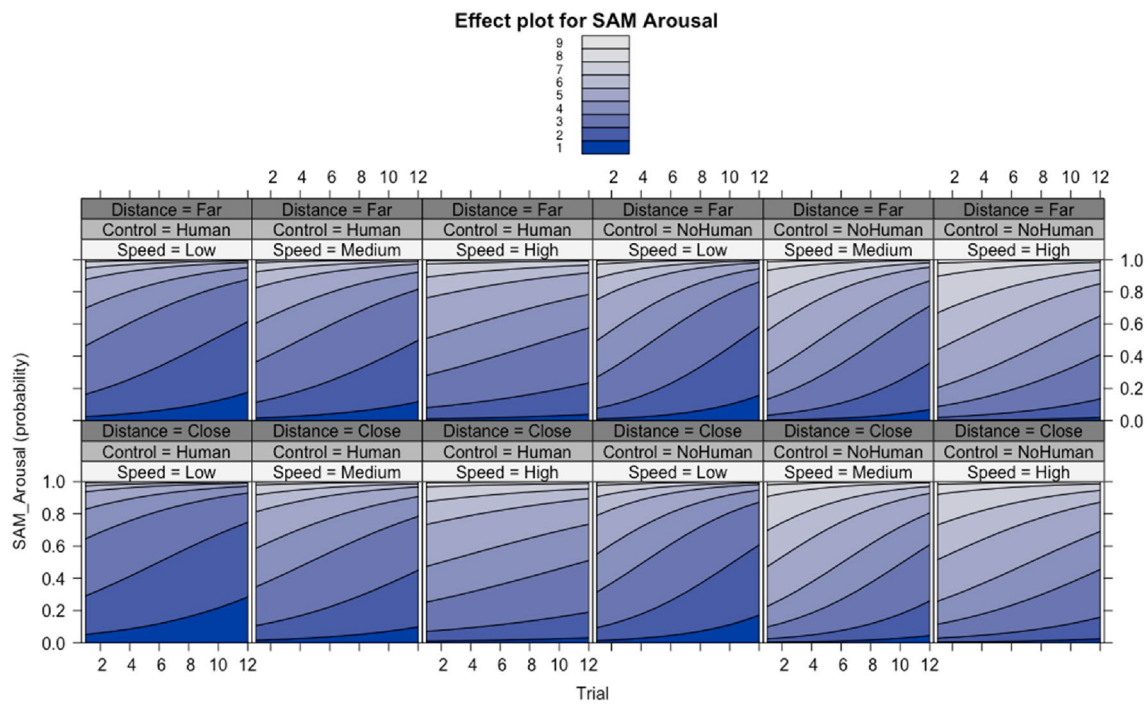


Fig. 10 Effect of the variables on the response probabilities for *Arousal*. For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

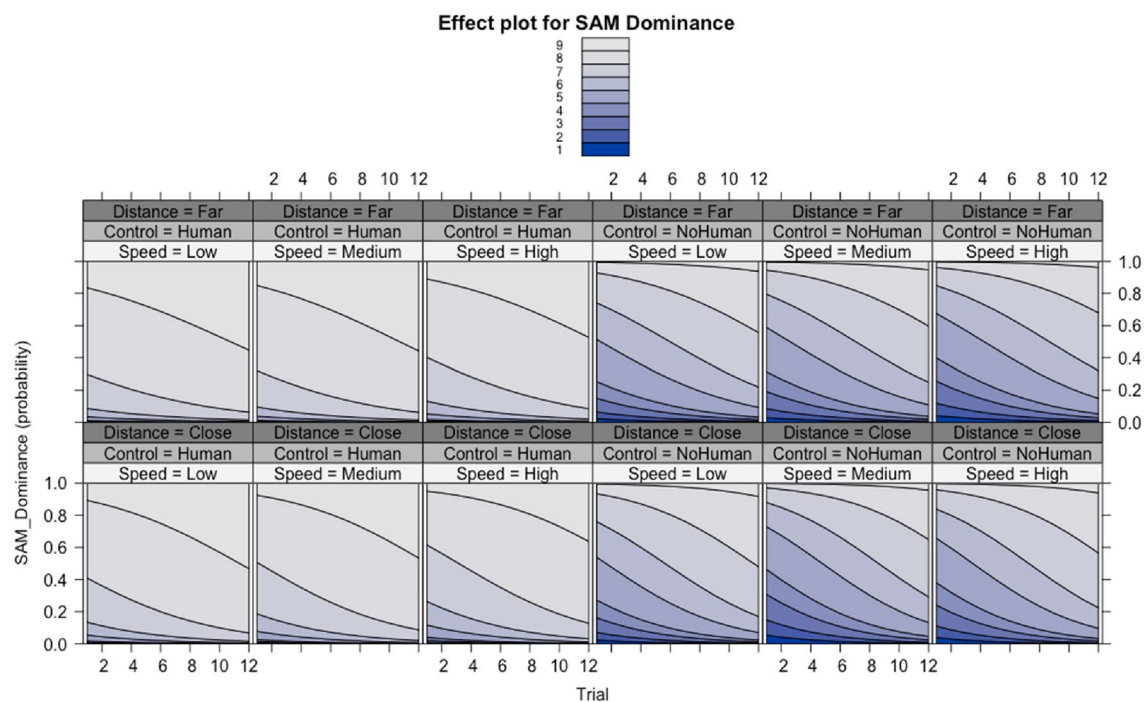


Fig. 11 Effect of the variables on the response probabilities for *Dominance*. For each configuration, the graph shows how response probability for each level (represented by the colored area) changes according to *Trial*

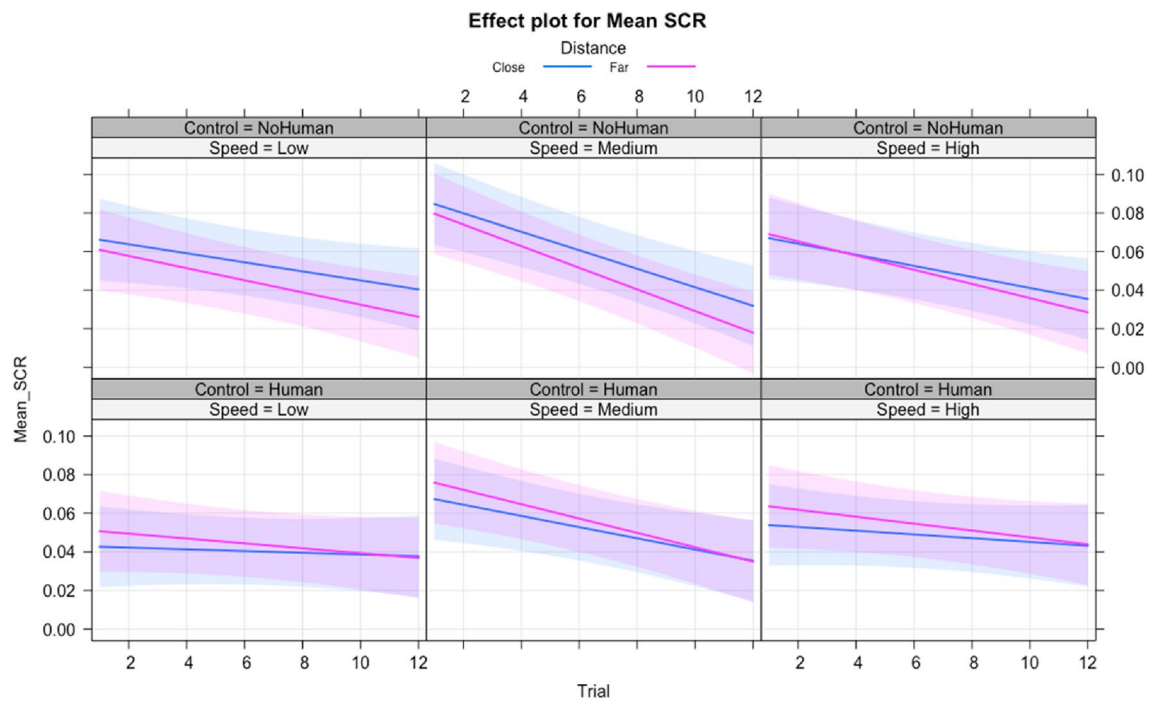


Fig. 12 Effect of the variables on the response probabilities for average SCR. Overlapping confidence intervals for different levels of *Distance* highlight the low significance of the factor on *MeanSCR*

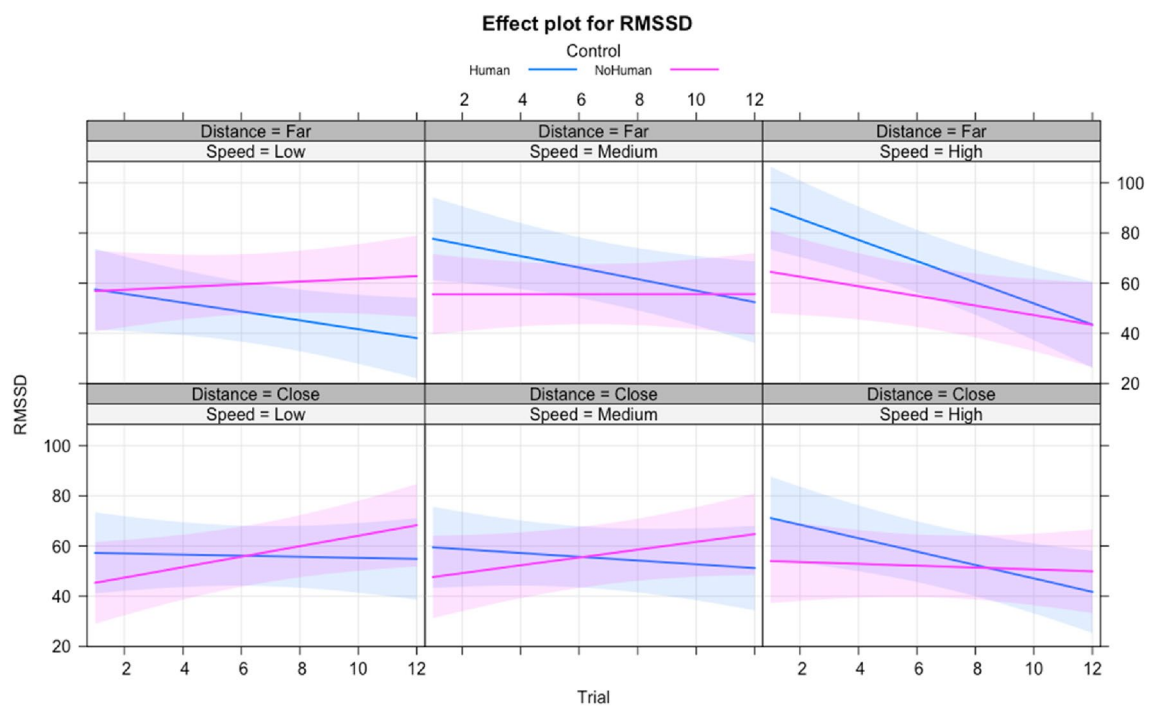


Fig. 13 Effect of the variables on the response probabilities for *RMSSD*. The change in slope for different levels of *Control* highlights the significant effect on *RMSSD*

Table 6 ANODE table with resulting p-values for quality of interaction including participant's characteristics

Model term	Q1_Helpful	Q2_NotSafe	Q3_Natural	Q4_Efficient	Q5_Fluid	Q6_Uncomf	Q7_Trustw
<i>Speed</i>	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	0.010**	0.420
<i>Distance</i>	0.205	0.259	0.589	0.881	0.655	0.201	0.031*
<i>Control</i>	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	0.001**
<i>Trial</i>	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***	<0.001***
<i>Age</i>	0.626	0.667	0.052	0.171	0.054	0.875	0.213
<i>Gender</i>	0.393	0.123	0.726	0.989	0.682	0.113	0.872
<i>Experience</i>	0.475	0.819	0.340	0.367	0.256	0.807	0.063
<i>NS1</i>	0.067	0.993	0.099	0.122	0.046*	0.813	0.149
<i>NS2</i>	0.351	0.939	0.794	0.798	0.885	0.834	0.491
<i>NS3</i>	0.868	0.917	0.617	0.345	0.372	0.758	0.294
<i>Speed·Distance</i>	0.533	0.471	0.456	0.782	0.635	0.354	0.102
<i>Speed·Control</i>	0.653	0.090	0.167	0.018*	0.018*	0.004**	0.247
<i>Distance·Control</i>	0.696	0.610	0.692	0.363	0.969	0.765	0.699
<i>Speed·Trial</i>	0.631	0.126	0.041*	0.870	0.324	0.023*	0.904
<i>Distance·Trial</i>	0.733	0.373	0.998	0.119	0.292	0.929	0.511
<i>Control·Trial</i>	0.004**	0.618	0.017*	0.005**	0.004**	0.055	0.808
<i>Speed·Distance·Control</i>	0.457	0.410	0.065	0.810	0.330	0.405	0.905

*0.05 > $p \geq 0.01$, **0.01 > $p \geq 0.001$, *** $p < 0.001$

Table 7 ANODE table with resulting p-values for SAM affective dimensions including participant's characteristics

Model term	Valence	Arousal	Dominance
<i>Speed</i>	<0.001***	<0.001***	0.029*
<i>Distance</i>	0.938	0.795	0.142
<i>Control</i>	<0.001***	<0.001***	<0.001***
<i>Trial</i>	0.035*	<0.001***	<0.001***
<i>Age</i>	0.339	0.511	0.007**
<i>Gender</i>	0.184	0.342	0.096
<i>Experience</i>	0.036*	0.507	0.290
<i>NS1</i>	0.399	0.756	0.412
<i>NS2</i>	0.174	0.720	0.725
<i>NS3</i>	0.478	0.669	0.932
<i>Speed·Distance</i>	0.694	0.240	0.569
<i>Speed·Control</i>	0.002**	0.678	0.633
<i>Distance·Control</i>	0.361	0.782	0.130
<i>Speed·Trial</i>	<0.001***	0.376	0.953
<i>Distance·Trial</i>	0.234	0.826	0.538
<i>Control·Trial</i>	0.514	0.190	0.432
<i>Speed·Distance·Control</i>	0.549	0.442	0.606

*0.05 > $p \geq 0.01$, **0.01 > $p \geq 0.001$, *** $p < 0.001$

Table 8 ANOVA table with resulting p-values for physiological response including participant's characteristics

Model term	Mean SCR	RMSSD
<i>Speed</i>	0.025*	0.098
<i>Distance</i>	0.632	0.055
<i>Control</i>	0.058	0.011*
<i>Trial</i>	<0.001***	0.059
<i>Age</i>	0.090	0.201
<i>Gender</i>	0.579	0.872
<i>Experience</i>	0.411	0.936
<i>NS1</i>	0.059	0.571
<i>NS2</i>	0.353	0.248
<i>NS3</i>	0.618	0.466
<i>Speed·Distance</i>	0.829	0.453
<i>Speed·Control</i>	0.610	0.118
<i>Distance·Control</i>	0.085	0.683
<i>Speed·Trial</i>	0.112	0.090
<i>Distance·Trial</i>	0.447	0.090
<i>Control·Trial</i>	0.065	0.014*
<i>Speed·Distance·Control</i>	0.919	0.273

*0.05 > $p \geq 0.01$, **0.01 > $p \geq 0.001$, *** $p < 0.001$

medium and high robot speeds ($\beta_{SpeedMed \cdot ControlNoHum} = -1.46$, $\beta_{SpeedHigh \cdot ControlNoHum} = -1.70$).

- *Trial* had a general negative effect ($\beta_{Trial} = -0.28$), meaning that the learning process potentially decreased the participant's positive feeling of novelty, even leading to boredom especially when the robot speed was low.

The terms *Speed*, *Control*, and *Trial* were significant in explaining the variability of *Arousal* and their effect can be seen in Fig. 10:

- Medium and high speeds were associated with a positive effect ($\beta_{SpeedMed} = 1.23$, $\beta_{SpeedHigh} = 1.61$), implying greater emotional activation as the robot speed increased.
- The lack of time execution control led to higher arousal of participants ($\beta_{ControlNoHum} = 1.44$).
- *Trial* had a general negative effect ($\beta_{Trial} = -0.17$), meaning that the learning process decreased the arousal, potentially leading to calmness or boredom.

With respect to *Dominance*, the terms *Speed*, *Control*, and *Trial* were significant and their effect are shown in Fig. 11:

- Participants sense of dominance was slightly reduced as the robot speed increased ($\beta_{SpeedMed} = -0.40$, $\beta_{SpeedHigh} = -0.84$).
- Dominance was significantly reduced in configurations with no time execution control ($\beta_{ControlNoHum} = -3.06$).
- As the experiment progressed, dominance gradually increased ($\beta_{Trial} = 0.20$) mainly due to the participant's learning process.

LR tests with only fixed-term models were performed to check if the random effect *Participant* was significant. From the obtained results, the random effect *Participant* resulted highly significant ($p < 0.001$) for every fitted model.

4.3 Physiological measures

The influence of the learning process and the experimental factors on physiological responses will be presented in this sub-section. In the Appendix B, the estimated parameters of the fitted models are reported.

Table 5 contains the results of the analysis of variance (ANOVA) for each linear mixed model, showing the significance of each model term.

Regarding the EDA response, the most significant terms for the average SCR (mean = 0.0497, sd = 0.0576) are *Speed*, *Control* and *Trial*. *Control* was still included among the significant variables because of its p-value close to the significance level ($\alpha = 0.05$). With reference to Fig. 12, the effects of the significant terms can be interpreted as follows:

- An increase of the average SCR was observed when robot speed was higher ($\beta_{SpeedMed} = 0.03$, $\beta_{SpeedHigh} = 0.01$), meaning that participants experienced more mental strain as the robot speed increased.
- As the experiment progressed, mental strain gradually decreased ($\beta_{Trial} = -0.0004$) mainly due to the participant's learning process.

- Configurations with no task execution control tended to lead through more stressful conditions, which raised the mean SCR ($\beta_{ControlNoHum} = 0.03$).

With respect to HRV, *Distance*, *Control*, and *Control·Trial* were the most significant terms in explaining the variability of RMSSD (mean = 56.89, sd = 40.01). *Distance* was still included among the significant variables because of its p-value close to the significance level ($\alpha = 0.05$). Referring to Fig. 13, the effects of the significant terms can be interpreted as follows:

- More stressful situations arose when participants had no control of time execution, which led to a significant decrease of RMSSD ($\beta_{ControlNoHum} = -14.25$). However, this negative effect was gradually compensated by the learning process of the participant ($\beta_{ControlNoHum·Trial} = 2.31$).
- A small positive effect on RMSSD was observed when participants were more distant from the robot's workspace ($\beta_{DistanceFar} = 1.80$), implying potentially slightly less stressful conditions.

LR tests with only fixed-term models were performed to check if the random effect *Participant* was significant. From the obtained results, the random effect *Participant* resulted highly significant ($p < 0.001$) for every fitted model.

4.4 Influence of participant's characteristics

The diversity of each participant turned out to be an important factor in order to explain the variability of the response variables considered. To gain some additional insights, the participant characteristics collected during the experiment (i.e., age, gender, prior experience with cobots, and attitudes towards robots) were included in the models.

Table 6 contains the ANODE for each interaction quality dimension. Only interaction naturalness (*Q3*) and team fluency (*Q5*) have at least one significant term related to participant characteristics. For interaction naturalness, *Age* had a positive effect ($\beta_{Age} = 0.10$) and was considered significant due to its p-value close to 0.05. Regarding team fluency, *Age* had a positive effect ($\beta_{Age} = 0.08$), with its p-value close to 0.05, and *NSI* had a significant negative effect ($\beta_{NSI} = -0.23$). This means that participants with a more negative attitude towards situations and interactions with robots tended to rate the team fluency lower.

Table 7 shows the ANODE for each SAM dimension. *Valence* was negatively affected by *Experience* ($\beta_{Experience} = -0.87$), meaning that participants with a high level of prior experience with cobots tended to report a less positive feeling, likely due to a decreased sense of novelty.

Dominance was found to be positively affected by *Age* ($\beta_{Age} = 0.09$). This may suggest that older individuals may tend to feel more in control of the situation during HRC.

In Table 8, the results of the ANOVA for the physiological responses are reported. None of the participant characteristics considered appeared to be significant for the average SCR and RMSSD.

LR tests with only fixed-term models were also carried out, revealing that the random effect *Participant* resulted still highly significant ($p < 0.001$) for every model. This suggests that there are important participant characteristics that have not been considered, and further study will be needed to identify them.

5 Discussion

The quantitative analysis revealed the influence of the learning process in conjunction with the configuration factors on the response variables, highlighting some interesting relationships and expanding the preliminary results of a previous work [11].

5.1 Effects of configuration factors and learning process

Among the fixed factors of robot configuration, *Speed* and *Control* were found to be most significant overall. In addition, the experience gained by interacting with the cobot during the experiment (*Trial*) significantly influenced the response variables. The main effects of the terms considered are summarized below.

- *Speed*. The robot movement speed had a general positive effect on the perceived robot helpfulness and efficiency, meaning that higher speeds were generally more appreciated by the participants from a performance viewpoint. However, *High* robot speed was also associated with slightly greater perceived unsafety and less interaction naturalness and comfort. Regarding the participants' affective state, the robot movement speed had a general high positive effect on arousal and a negative one on dominance. This means that participants, in general, felt slightly less in control of the situation and more aroused when robot movement speed was higher, leading to potentially more stressful situations. This was also confirmed from a physiological point of view, by an increase of average SCRs (*MeanSCR*) for higher speeds. Valence was positively influenced by higher robot movement speeds, leading to more positive feelings in general.
- *Control*. The absence of participant's control of the task execution time had a significant negative effect on the interaction quality. In this kind of configurations, the

robot was perceived by participants less helpful and trustworthy. In addition, the interaction was perceived less safe, natural, efficient, fluid, and comfortable. Regarding the participants' affective state, the lack of time execution control had a general negative effect: valence was slightly degraded, arousal increased, and dominance significantly decreased. This led to more stressful situations for the participants and a physiological confirmation was found in a significant decrease of HRV.

- *Speed·Control*. The interaction term *Speed·Control* played an interesting role. The lack of time execution control by the participant amplified or degraded the robot movement speed effect on some aspects of the interaction quality. Perceived team fluency was greater for higher speeds, with a slightly higher effect for the *Medium* speed. However, without time execution control the positive effect was almost nullified for higher speeds. *Speed·Control* had also a negative effect on perceived efficiency and especially on comfort. Concerning the affective state, the term *Speed·Control* had a significant negative effect on valence, meaning that the lack of time execution control by the participant led to more unpleasant emotions with higher robot speeds.
- *Distance*. The distance from the robot workspace was found to be not particularly influencing in general; just a slight decrease in robot trustworthiness and a slight increase in HRV were associated with a higher distance. This result could be also influenced by the relatively small size of the UR3e robot and future research will be required to verify this hypothesis.
- *Trial*. The experimental progress had a significant overall positive effect on the interaction quality. As the experiment progressed, participants tended to perceive the robot more helpful, safe, and trustworthy. In addition, the interaction was felt gradually more fluid, efficient, and comfortable. However, a constant slight decrease of perceived interaction naturalness emerged when robot speed was *Low*. Concerning the participants' affective state, the learning process had a negative effect on arousal and a positive effect on dominance, leading to a gradual relaxation of the participant. This effect was also confirmed physiologically by the constant decrease of the average SCR as the experiment progressed. The learning process negatively affected the valence when the robot speed was *Low*. This can be interpreted as an increasing sense of boredom in participants when the robot movement speed is perceived to be too slow. The overall positive effect of *Trial* on the response variables was further enhanced by the interaction terms *Speed·Trial* and *Control·Trial*.
- *Speed·Trial*. The negative effect of *Speed* on interaction naturalness and comfort was gradually compensated by the interaction term *Speed·Trial*, meaning that the

learning process particularly helped in improving these aspects for higher speeds. Regarding the affective state, valence was positively influenced by the interaction term *Speed·Trial*. *Medium* speed was initially associated with a slightly more positive feeling than the *High* one. However, as the experiment progressed, participants were in general more pleased with the *High* speed level.

- *Control·Trial*. As the experiment progressed, the interaction term *Control·Trial* partially compensated the negative effects on interaction naturalness, robot helpfulness, team fluency and efficiency due to the absence of execution time control. In addition, the interaction term *Control·Trial* showed also a significant gradual compensation of the negative effect of *Control* on HRV due to the learning process.

The interpretations of the results were also reflected in the unstructured interviews with participants. Initially, most participants were more comfortable with the *Low* and *Medium* robot movement speeds. However, towards the last few trials, the *Low* speed was perceived to be particularly tedious, even leading some participants to become more distracted. This may be due to the repetitiveness of the task and, consequently, the learning process that led most participants to prefer a *High* robot speed by the end of the experiment. The factor of distance from the robot's workspace, instead, was found to be more of a matter of preference due to the participant's comfort in performing their operations. Controlling the task execution time was generally preferred by participants, primarily due to less psychological pressure. Towards the end of the experiment, however, some participants were able to finish their tasks well in advance when they were not in control of the task execution time, implying a longer wait time before the robot proceeded with the next task. This resulted in a partial boredom effect but also in a degradation of the perceived efficiency, naturalness, and fluency of the interaction. Interestingly, at the end of the experiment, some participants expressed a preference for configurations in which the robot continued with its operations automatically, without waiting for a command, because this implied one less operation for the operator. This interesting cue raises the demand and the need to make collaborative robots more situationally aware, in order to allow a greater support of the operator also from the psycho-cognitive point of view. Moreover, the implementation of the ability of initiative by the cobot would allow to better support the operator and to create a symbiotic collaboration, characterized by a greater naturalness and fluency.

5.2 Importance of the operator's characteristics

The random effect *Participant* was found to be significant in all models, highlighting the importance of a more

personalized HRC. Moreover, this fact hinted that some user characteristics may significantly influence the perceived interaction quality and affective response. To further investigate this fact, all previous models were re-fitted by including the variables *Age*, *Gender*, *Experience*, and the three NARS-related scores (i.e., *NS1*, *NS2*, *NS3*) (see Tables 12, 13, and 14). Results revealed that of the characteristics considered only a few were significant for some response variables. *Age* was associated with a positive effect on self-reported dominance and perceived interaction naturalness and fluency. This may imply that an older age may result in feeling more confident and in control of the situation during the HRC. However, this effect may be due to the fact that most of the older people involved in the experiment knew at least what a cobot was. Future investigation of this will be needed by involving people with a more diverse background and a wider, more distributed age range. *Experience* had a significant negative effect on self-reported valence. This effect could be due to a decreased sense of novelty and excitement for those more familiar with collaborative robots, which can also lead to boredom. Finally, *NS1* was associated with a significant negative effect on perceived team fluency. This means that participants with a more positive attitude towards situations and interactions with robots tended to perceive greater team fluency. It is worth noting that even in the new models the *Participant* random effect still turned out to be quite significant, indicating that there are potentially significant aspects of the participant that were not accounted for. Further investigation of user characteristics is needed and will be the focus of future works.

6 Conclusions

The aim of this paper was to expand the previous preliminary results [11] by deepening the joint effect of the learning process with several HRC setting factors on the operator's affective state, perceived interaction quality, and physiological response. Among the configuration factors, cobot movement speed (*Speed*) and control of the task execution time (*Control*) were found to be the most influential ones. The learning process (*Trial*) also played a key role, especially in improving user experience and decreasing stressful conditions. It also emerged that the learning effect can change the perception of certain factors.

Regarding the robot movement speed, initially most participants preferred a low speed. However, as the experiment progressed, there was a greater preference toward higher speeds as they helped to be more engaged and efficient. This result highlights that when the robot trajectories are predictable and known, there is a preference to maintain a high speed of the robot's movements, mainly for performance and satisfaction reasons. On the other hand, when the operator

is not aware of the robot's trajectory, it may be preferred to maintain a lower movement speed.

The absence of participant's control of the task execution time led to a general degradation of perceived interaction quality and an increase of stress. However, it is interesting to note that these negative effects were mitigated through the learning process. In fact, some participants at the end of the experiment even expressed that they appreciated the absence of the control button, as it represented one less operation for them to do. This hint suggests that the implementation of a system that is able to automatically recognize the completion of human operations may be particularly appreciated in the HRC, contributing to establish a more natural and human-like interaction.

The distance between operator and robot workspace turned out to be merely a question of preference and to be not particularly influential. This fact might be due to the UR3e robot's relatively small size, although further research is needed to confirm this hypothesis. Future work will focus on introducing cobots of different payload (e.g., UR10e or UR16e) to investigate whether the cobot size may influence the perceived interaction quality and the overall user experience.

The non-invasive acquisition of physiological responses provided objective information about the participant's state and unconscious reactions. By analyzing HRV and EDA, confirmation of the participants' self-reported affective state and unstructured feedback was found. Such concordance is certainly promising for the design of a real-time monitoring system of operator well-being in HRC.

The differences among participants were found to be significant, highlighting the need to design and implement a more customizable HRC. An initial investigation into the effect of certain aspects of participants on response variables was conducted, however a more in-depth study is needed. Future work will also focus on a more in-depth study of the link between physiological responses and the affective state of the operator in the HRC.

Appendix A—Mixed-effect ordinal logistic regression (MOLR)

Ordinal Logistic Regression (OLR) model is a regression model for ordinal dependent variables based on Cumulative Link Models (CLMs). CLMs are a powerful class of

Generalized Linear Models (GLMs) for ordinal response variables, allowing them to be treated properly as categorical variables as well as exploiting their ordinal nature [42, 44]. In CLMs, the cumulative probability for each level j of the ordinal response (i.e., $\mathbb{P}(Y \leq j)$) is modeled via a link function. In this context, a commonly used link function (i.e., a function that links the probability to the linear function of the predictor variables) is the logit link. The logit link maps probability values from $(0, 1)$ to real numbers in $(-\infty, +\infty)$ and is defined as the logarithm of the odds $\frac{p}{1-p}$, where p is the probability. Logit link is often preferred to others (such as probit or log-log) because of mathematical convenience and higher interpretability of results. CLMs with the logit link function are called OLR models. Therefore, OLR model can be thought of as an extension of the logistic regression model (which applies to binary response variables), allowing for more than two (ordered) response categories.

Compared to OLR, other more common methods can be used to analyze an ordinal response variable, however they have several limitations. For instance, by treating the numerically coded ordinal variable as quantitative, typical least squares regression can be used. However, this common method, in addition to introducing a scale promotion for the response variable, often violates the assumptions of homoscedasticity and normality of the residuals. Another common method of modeling an ordinal variable is to treat it as nominal and use a multinomial logistic regression. This choice also introduces limitations, most notably the loss of information from ignoring the ordering which results in a loss of power for the model.

Mixed-effect Ordinal Logistic Regression (MOLR) models are based on Cumulative Link Mixed Models (CLMMs), which are an extension of CLMs that allow the inclusion of normally distributed random effects. The MOLR model can be specified in terms of cumulative logits as follows:

$$\begin{aligned} \text{logit}(\mathbb{P}(Y_{ik} \leq j)) &= \log\left(\frac{\mathbb{P}(Y_{ik} \leq j)}{1 - \mathbb{P}(Y_{ik} \leq j)}\right) \\ &= \theta_j - \mathbf{x}_{ik}^T \boldsymbol{\beta} - u_k \quad \text{for } j = 1, \dots, J-1 \end{aligned} \quad (2)$$

with $i = 1, \dots, n$ and $k = 1, \dots, K$, where J is the number of levels of the ordinal response variable, n the number of

observations, and K the number of participants (i.e., 42 in this case). The term $\mathbb{P}(Y_{ik} \leq j)$ represents the probability that the observation i of the participant k is associated with a rating equal or below j . Note that the logit is not defined for $j=J$, since $\mathbb{P}(Y_{ik} \leq J) = 1$. The vector β contains the model parameters, the vector $\mathbf{x}_{ik}^T$ contains data of the independent variables of observation i of participant k , and the terms θ_j are called threshold parameters and serves as intercepts of the model. The term u_k represents the random effect for participant k , where $u_k \sim N(0, \sigma^2)$. Note that the negative sign in front of $\mathbf{x}_{ik}^T \beta$ ensures that positive parameters are associated with increased probability for higher ratings as the explanatory variables increase.

Appendix B—List of estimated parameters of the models

In this section, the estimated parameters for the fitted models are reported. Tables 9, 10, and 11 contain the estimated parameters of the models for interaction quality dimensions, SAM dimensions, and physiological signals, respectively. In Tables 12, 13, and 14, instead, the estimated parameters of the models with participant characteristics for interaction quality dimensions, SAM dimensions, and physiological signals are reported.

Table 9 MOLR model parameter estimates for each interaction quality dimension

Model term	Model parameter	Coefficient						
		<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q5</i>	<i>Q6</i>	<i>Q7</i>
<i>Thresholds (intercepts)</i>	θ_1	−6.20	0.29	−6.64	−4.57	−5.06	0.80	−8.22
	θ_2	−4.81	2.59	−5.23	−2.80	−3.33	3.09	−6.37
	θ_3	−3.57	3.52	−3.78	−1.75	−2.10	4.12	−4.78
	θ_4	−2.27	4.42	−2.29	−0.71	−0.94	4.90	−3.57
	θ_5	−0.30	5.54	−0.37	0.81	0.81	6.46	−1.16
	θ_6	2.84	7.08	2.21	3.32	3.17	9.14	1.87
<i>Speed</i>	<i>Medium</i>	0.95	−0.02	0.11	1.68	1.63	0.66	−0.68
	<i>High</i>	0.99	0.30	−0.46	1.92	1.49	0.97	−0.63
<i>Distance</i>	<i>Far</i>	−0.77	0.70	−0.62	−0.73	−0.34	0.33	−0.45
<i>Control</i>	<i>NoHuman</i>	−2.17	1.31	−2.42	−1.65	−1.72	1.27	−0.17
<i>Trial</i>	—	0.07	−0.15	−0.01	0.03	0.01	−0.04	0.17
<i>Speed·Distance</i>	<i>Medium·Far</i>	0.82	−0.30	0.63	0.49	−0.25	−0.29	0.94
	<i>High·Far</i>	0.69	−0.91	1.25	0.49	−0.10	−1.05	1.07
<i>Speed·Control</i>	<i>Medium·NoHuman</i>	0.29	1.34	−0.26	−0.65	−1.64	1.90	−0.25
	<i>High·NoHuman</i>	−0.19	0.31	0.10	−0.91	−1.04	0.89	−0.56
<i>Distance·Control</i>	<i>Far·NoHuman</i>	0.36	−0.36	0.80	−0.01	−0.42	−0.24	0.05
<i>Speed·Trial</i>	<i>Medium·Trial</i>	−0.02	−0.04	0.11	0.01	0.04	−0.22	0.03
	<i>High·Trial</i>	0.04	0.10	0.16	0.03	0.09	−0.07	0.03
<i>Distance·Trial</i>	<i>Far·Trial</i>	0.02	−0.06	0.01	0.08	0.06	−0.01	0.04
<i>Control·Trial</i>	<i>NoHuman·Trial</i>	0.16	−0.03	0.13	0.15	0.16	−0.12	−0.01
<i>Speed·Distance·Control</i>	<i>Medium·Far·NoHuman</i>	−1.07	−0.47	−0.20	−0.49	1.10	−0.54	−0.16
	<i>High·Far·NoHuman</i>	−0.42	0.79	−1.79	−0.44	0.11	0.79	−0.41
Random effect —	St. deviation							
		<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q5</i>	<i>Q6</i>	<i>Q7</i>
<i>Participant</i>	—	2.42	2.12	2.27	1.91	1.95	2.41	2.35

Table 10 MOLR model parameter estimates for each SAM dimension

Model term	Model parameter	Coefficient		
		<i>Valence</i>	<i>Arousal</i>	<i>Dominance</i>
<i>Thresholds (Intercepts)</i>	θ_1	− 7.67	− 3.08	− 6.52
	θ_2	− 6.64	− 1.07	− 5.39
	θ_3	− 4.95	0.42	− 4.46
	θ_4	− 3.62	1.40	− 3.80
	θ_5	− 2.17	2.52	− 2.66
	θ_6	− 1.09	3.47	− 1.66
	θ_7	0.52	4.92	− 0.16
	θ_8	2.85	6.56	2.32
<i>Speed</i>	<i>Medium</i>	0.55	1.23	− 0.40
	<i>High</i>	0.45	1.61	− 0.85
<i>Distance</i>	<i>Far</i>	− 0.52	0.75	0.53
<i>Control</i>	<i>NoHuman</i>	− 0.30	1.44	− 3.06
<i>Trial</i>	–	− 0.28	− 0.18	0.20
<i>Speed·Distance</i>	<i>Medium·Far</i>	0.08	− 0.82	0.28
	<i>High·Far</i>	− 0.23	− 0.89	0.37
<i>Speed·Control</i>	<i>Medium·NoHuman</i>	− 1.46	0.20	− 0.44
	<i>High·NoHuman</i>	− 1.70	− 0.43	0.35
<i>Distance·Control</i>	<i>Far·NoHuman</i>	− 0.16	− 0.52	− 0.39
<i>Speed·Trial</i>	<i>Medium·Trial</i>	0.22	0.01	0.01
	<i>High·Trial</i>	0.29	0.08	0.01
<i>Distance·Trial</i>	<i>Far·Trial</i>	0.07	− 0.01	− 0.04
<i>Control· Trial</i>	<i>NoHuman·Trial</i>	0.04	− 0.07	0.04
<i>Speed·Distance·Control</i>	<i>Medium·Far·NoHuman</i>	0.49	0.25	0.25
	<i>High·Far·NoHuman</i>	0.91	0.97	− 0.55
Random effect	–	St. deviation		
		<i>Valence</i>	<i>Arousal</i>	<i>Dominance</i>
<i>Participant</i>	–	2.24	2.12	1.51

Table 11 LMM parameter estimates for each physiological indicator

Model term	Model parameter	Coefficient	
		<i>Mean SCR</i>	<i>RMSSD</i>
<i>Intercept</i>	–	0.0431	57.46
<i>Speed</i>	<i>Medium</i>	0.0279	2.75
	<i>High</i>	0.0117	16.36
<i>Distance</i>	<i>Far</i>	0.0089	1.80
<i>Control</i>	<i>NoHuman</i>	0.0254	– 14.25
<i>Trial</i>	–	– 0.0004	– 0.22
<i>Speed·Distance</i>	<i>Medium·Far</i>	0.0005	17.95
	<i>High·Far</i>	0.0017	18.48
<i>Speed·Control</i>	<i>Medium·NoHuman</i>	– 0.0061	0.12
	<i>High·NoHuman</i>	– 0.0104	– 5.22
<i>Distance·Control</i>	<i>Far·NoHuman</i>	– 0.0132	11.26
<i>Speed·Trial</i>	<i>Medium·Trial</i>	– 0.0025	– 0.53
	<i>High·Trial</i>	– 0.0005	– 2.46
<i>Distance·Trial</i>	<i>Far·Trial</i>	– 0.0008	– 1.55
<i>Control·Trial</i>	<i>NoHuman·Trial</i>	– 0.0019	2.31
<i>Speed·Distance·Control</i>	<i>Medium·Far·NoHuman</i>	– 0.0003	– 21.56
	<i>High·Far·NoHuman</i>	0.0056	– 19.51
Random effect	–	St. deviation	
		<i>Mean SCR</i>	<i>RMSSD</i>
<i>Participant</i>	–	0.0431	20.62

Table 12 MOLR model parameter estimates for each interaction quality dimension, including participant characteristics

Model term	Model parameter	Coefficient						
		<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q5</i>	<i>Q6</i>	<i>Q7</i>
<i>Thresholds (intercepts)</i>	θ_1	−8.75	0.60	−6.24	−5.24	−5.39	0.32	−8.64
	θ_2	−7.36	2.90	−4.83	−3.47	−3.65	2.61	−6.77
	θ_3	−6.11	3.83	−3.38	−2.43	−2.42	3.63	−5.17
	θ_4	−4.81	4.73	−1.89	−1.39	−1.25	4.42	−3.95
	θ_5	−2.84	5.85	0.03	0.14	0.49	5.97	−1.52
	θ_6	0.30	7.40	2.61	2.64	2.85	8.65	1.50
<i>Speed</i>	<i>Medium</i>	0.94	−0.03	0.11	1.68	1.61	0.66	−0.67
	<i>High</i>	0.97	0.28	−0.47	1.90	1.46	0.95	−0.62
<i>Distance</i>	<i>Far</i>	−0.76	0.68	−0.60	−0.71	−0.31	0.32	−0.43
<i>Control</i>	<i>NoHuman</i>	−2.16	1.31	−2.39	−1.63	−1.70	1.28	−0.16
<i>Trial</i>	–	0.07	−0.16	−0.01	0.03	0.01	−0.04	0.17
<i>Age</i>	–	0.02	0.02	0.10	0.06	0.08	0.01	0.06
<i>Gender</i>	<i>Female</i>	−0.70	−1.25	0.28	−0.01	−0.27	−1.48	0.13
<i>Experience</i>	–	−0.33	0.10	−0.43	−0.35	−0.43	−0.13	−0.90
<i>NS1</i>	–	−0.26	−0.01	−0.22	−0.18	−0.23	−0.04	−0.21
<i>NS2</i>	–	−0.12	−0.01	−0.03	−0.03	−0.02	−0.03	0.09
<i>NS3</i>	–	−0.02	0.01	−0.07	−0.11	−0.10	0.05	−0.15
<i>Speed·Distance</i>	<i>Medium·Far</i>	0.82	−0.29	0.63	0.48	−0.26	−0.29	0.93
	<i>High·Far</i>	0.70	−0.90	1.25	0.48	−0.10	−1.05	1.06
<i>Speed·Control</i>	<i>Medium·NoHuman</i>	0.29	1.34	−0.27	−0.65	−1.65	1.90	−0.25
	<i>High·NoHuman</i>	−0.18	0.32	0.10	−0.90	−1.03	0.88	−0.56
<i>Distance·Control</i>	<i>Far·NoHuman</i>	0.35	−0.36	0.80	−0.01	−0.41	−0.24	0.04
<i>Speed·Trial</i>	<i>Medium·Trial</i>	−0.02	−0.03	0.11	0.01	0.04	−0.22	0.03
	<i>High·Trial</i>	0.04	0.11	0.16	0.03	0.10	−0.06	0.03
<i>Distance·Trial</i>	<i>Far·Trial</i>	0.02	−0.05	0.01	0.08	0.06	−0.01	0.04
<i>Control·Trial</i>	<i>NoHuman·Trial</i>	0.16	−0.03	0.13	0.14	0.15	−0.12	−0.01
<i>Speed·Distance·Control</i>	<i>Medium·Far·NoHuman</i>	−1.06	−0.48	−0.20	−0.48	1.12	−0.55	−0.15
	<i>High·Far·NoHuman</i>	−0.42	0.79	−1.80	−0.44	0.00	0.79	−0.40
Random effect –	St. deviation							
		<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q5</i>	<i>Q6</i>	<i>Q7</i>
<i>Participant</i> –		2.16	2.01	2.06	1.76	1.72	2.34	2.17

Table 13 MOLR model parameter estimates for each SAM dimension, including participant characteristics

Model term	Model parameter	Coefficient		
		<i>Valence</i>	<i>Arousal</i>	<i>Dominance</i>
<i>Thresholds (Intercepts)</i>	θ_1	− 8.50	− 2.10	− 4.69
	θ_2	− 7.47	− 0.09	− 3.56
	θ_3	− 5.78	1.40	− 2.63
	θ_4	− 4.46	2.38	− 1.97
	θ_5	− 3.01	3.50	− 0.83
	θ_6	− 1.93	4.45	0.17
	θ_7	− 0.32	5.91	1.68
	θ_8	2.01	7.56	4.16
<i>Speed</i>	<i>Medium</i>	0.55	1.24	− 0.43
	<i>High</i>	0.45	1.61	− 0.88
<i>Distance</i>	<i>Far</i>	− 0.49	0.75	0.50
<i>Control</i>	<i>NoHuman</i>	− 0.27	1.45	− 3.04
<i>Trial</i>	–	− 0.27	− 0.18	0.20
<i>Age</i>	–	0.04	0.03	0.09
<i>Gender</i>	<i>Female</i>	0.96	0.74	− 0.87
<i>Experience</i>	–	− 0.87	− 0.29	− 0.31
<i>NS1</i>	–	− 0.10	0.04	− 0.07
<i>NS2</i>	–	− 0.15	− 0.04	0.03
<i>NS3</i>	–	0.09	0.06	0.01
<i>Speed·Distance</i>	<i>Medium·Far</i>	0.09	− 0.82	0.27
	<i>High·Far</i>	− 0.21	− 0.89	0.36
<i>Speed·Control</i>	<i>Medium·NoHuman</i>	− 1.45	0.20	− 0.44
	<i>High·NoHuman</i>	− 1.70	− 0.43	0.33
<i>Distance·Control</i>	<i>Far·NoHuman</i>	− 0.15	− 0.52	− 0.42
<i>Speed·Trial</i>	<i>Medium·Trial</i>	0.22	0.01	0.02
	<i>High·Trial</i>	0.29	0.08	0.02
<i>Distance·Trial</i>	<i>Far·Trial</i>	0.06	− 0.01	− 0.03
<i>Control· Trial</i>	<i>NoHuman· Trial</i>	0.03	− 0.07	0.04
<i>Speed·Distance·Control</i>	<i>Medium·Far·NoHuman</i>	0.48	0.25	0.27
	<i>High·Far·NoHuman</i>	0.89	0.98	− 0.53
Random effect	–	St. deviation		
		<i>Valence</i>	<i>Arousal</i>	<i>Dominance</i>
<i>Participant</i>	–	1.87	2.05	1.30

Table 14 LMM parameter estimates for each physiological indicator, including participant characteristics

Model term	Model parameter	Coefficient	
		<i>Mean SCR</i>	<i>RMSSD</i>
<i>Intercept</i>	–	0.0697	93.20
<i>Speed</i>	<i>Medium</i>	0.0271	2.44
	<i>High</i>	0.0118	15.88
<i>Distance</i>	<i>Far</i>	0.0085	2.06
<i>Control</i>	<i>NoHuman</i>	0.0252	– 14.09
<i>Trial</i>	–	– 0.0005	– 0.23
<i>Age</i>	–	– 0.0018	– 0.69
<i>Gender</i>	<i>Female</i>	– 0.0091	– 1.40
<i>Experience</i>	–	0.0077	0.40
<i>NS1</i>	–	0.0054	0.83
<i>NS2</i>	–	– 0.0024	– 1.59
<i>NS3</i>	–	0.0014	– 1.12
<i>Speed· Distance</i>	<i>Medium· Far</i>	0.0005	18.02
	<i>High· Far</i>	0.0017	18.59
<i>Speed· Control</i>	<i>Medium· NoHuman</i>	– 0.0061	0.13
	<i>High· NoHuman</i>	– 0.0104	– 5.19
<i>Distance· Control</i>	<i>Far· NoHuman</i>	– 0.0132	11.26
<i>Speed· Trial</i>	<i>Medium· Trial</i>	– 0.0025	– 0.49
	<i>High· Trial</i>	– 0.0005	– 2.39
<i>Distance· Trial</i>	<i>Far· Trial</i>	– 0.0008	– 1.59
<i>Control· Trial</i>	<i>NoHuman· Trial</i>	– 0.0019	2.28
<i>Speed· Distance· Control</i>	<i>Medium· Far· NoHuman</i>	– 0.0003	– 21.61
	<i>High· Far· NoHuman</i>	0.0056	– 19.65
Random effect	–	St. deviation	
		<i>Mean SCR</i>	<i>RMSSD</i>
<i>Participant</i>	–	0.0432	21.33

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement. This work has been partially supported by “Ministero dell’Istruzione, dell’Università e della Ricerca” Award “TESUN-83486178370409 finanziamento dipartimenti di eccellenza CAP. 1694 TIT. 232 ART. 6”.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Standards The authors respect the Ethical Guidelines of the Journal. Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. ISO/TS 15066:2016 (2016) Robots and robotic devices—collaborative robots. International Organization for Standardization, Geneva, CH
2. Vicentini F (2020) Collaborative robotics: a survey. J Mech Des. <https://doi.org/10.1115/1.4046238>
3. da Silva NA, Abreu JL, Orsolin Klingenberg C, Antunes Junior JAV, Lacerda DP (2022) Industry 4.0 and micro and small enterprises: systematic literature review and analysis. Prod Manuf Res 10:696–726. <https://doi.org/10.1080/21693277.2022.2124466>
4. Mishra R (2020) Confirmation of a measurement model for manufacturing flexibility development practices. Int J Qual Reliab Manag 38:317–338. <https://doi.org/10.1108/IJQRM-01-2019-0027>
5. Young MS, Brookhuis KA, Wickens CD, Hancock PA (2015) State of science: mental workload in ergonomics. Ergonomics 58:1–17. <https://doi.org/10.1080/00140139.2014.956151>
6. Gervasi R, Mastrogioacomo L, Franceschini F (2020) A conceptual framework to evaluate human–robot collaboration. Int

- J Adv Manuf Technol 108:841–865. <https://doi.org/10.1007/s00170-020-05363-1>
7. Gervasi R, Mastrogiacomio L, Maisano DA, Antonelli D, Franceschini F (2022) A structured methodology to support human–robot collaboration configuration choice. *Prod Eng Res Dev* 16:435–451. <https://doi.org/10.1007/s11740-021-01088-6>
8. Arai T, Kato R, Fujita M (2010) Assessment of operator stress induced by robot collaboration in assembly. *CIRP Ann* 59:5–8. <https://doi.org/10.1016/j.cirp.2010.03.043>
9. Kühnlenz B, Erhart M, Kainert M, Wang Z-Q, Wilm J, Kühnlenz K (2018) Impact of trajectory profiles on user stress in close human–robot interaction. *Automatisierungstechnik* 66:483–491. <https://doi.org/10.1515/auto-2018-0004>
10. Kulić D, Croft E (2007) Physiological and subjective responses to articulated robot motion. *Robotica* 25:13–27. <https://doi.org/10.1017/S0263574706002955>
11. Gervasi R, Aliev K, Mastrogiacomio L, Franceschini F (2022) User experience and physiological response in human–robot collaboration: a preliminary investigation. *J Intell Robot Syst* 106:36. <https://doi.org/10.1007/s10846-022-01744-8>
12. Gopinath V, Johansen K (2019) Understanding situational and mode awareness for safe human–robot collaboration: case studies on assembly applications. *Prod Eng Res Dev* 13:1–9. <https://doi.org/10.1007/s11740-018-0868-2>
13. Inkulu AK, Bahubalendruni MVAR, Dara A, SankaranarayanaSamy K (2021) Challenges and opportunities in human robot collaboration context of Industry 4.0—a state of the art review. *Ind Robot Int J Robot Res Appl*. <https://doi.org/10.1108/IR-04-2021-0077>
14. Faccio M, Granata I, Menini A, Milanese M, Rossato C, Bottin M, Minto R, Pluchino P, Gamberini L, Boschetti G, Rosati G (2022) Human factors in cobot era: a review of modern production systems features. *J Intell Manuf*. <https://doi.org/10.1007/s10845-022-01953-w>
15. Vink P (2012) *Advances in social and organizational factors*. CRC Press, Boca Raton
16. Faber M, Mertens A, Schlick CM (2017) Cognition-enhanced assembly sequence planning for ergonomic and productive human–robot collaboration in self-optimizing assembly cells. *Prod Eng Res Dev* 11:145–154. <https://doi.org/10.1007/s11740-017-0732-9>
17. Gualtieri L, Fraboni F, De Marchi M, Rauch E (2022) Development and evaluation of design guidelines for cognitive ergonomics in human–robot collaborative assembly systems. *Appl Ergon* 104:103807. <https://doi.org/10.1016/j.apergo.2022.103807>
18. Wang L, Liu S, Liu H, Wang XV (2020) Overview of human–robot collaboration in manufacturing. In: Wang L, Majstorovic VD, Mourtzis D, Carpanzano E, Moroni G, Galantucci LM (eds) *Proceedings of the 5th international conference on the industry 40 model for advanced manufacturing*. Springer International Publishing, Cham, pp 15–58
19. Gervasi R, Barravecchia F, Mastrogiacomio L, Franceschini F (2022) Applications of affective computing in human–robot interaction: state-of-art and challenges for manufacturing. *Proc Inst Mech Eng B J Eng Manuf*. <https://doi.org/10.1177/09544054221121888>
20. Odesanmi GA, Wang Q, Mai J (2023) Skill learning framework for human–robot interaction and manipulation tasks. *Robot Comput Integr Manuf* 79:102444. <https://doi.org/10.1016/j.rcim.2022.102444>
21. Neves M, Neto P (2022) Deep reinforcement learning applied to an assembly sequence planning problem with user preferences. *Int J Adv Manuf Technol* 122:4235–4245. <https://doi.org/10.1007/s00170-022-09877-8>
22. Mohammed A, Schmidt B, Wang L (2017) Active collision avoidance for human–robot collaboration driven by vision sensors. *Int J Comput Integr Manuf* 30:970–980. <https://doi.org/10.1080/0951192X.2016.1268269>
23. Buerkle A, Matharu H, Al-Yacoub A, Lohse N, Bamber T, Ferreira P (2022) An adaptive human sensor framework for human–robot collaboration. *Int J Adv Manuf Technol* 119:1233–1248. <https://doi.org/10.1007/s00170-021-08299-2>
24. Khamaisi RK, Brunzini A, Grandi F, Peruzzini M, Pellicciari M (2022) UX assessment strategy to identify potential stressful conditions for workers. *Robot Comput Integr Manuf* 78:102403. <https://doi.org/10.1016/j.rcim.2022.102403>
25. Wickens CD (2008) Multiple resources and mental workload. *Hum Factors* 50:449–455. <https://doi.org/10.1518/001872008X288394>
26. Reid GB, Nygren TE (1988) The subjective workload assessment technique: a scaling procedure for measuring mental workload. In: Hancock PA, Meshkati N (eds) *Advances in psychology*. North-Holland, Amsterdam, pp 185–218
27. 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*. North-Holland, Amsterdam, pp 139–183
28. Gawron VJ (2008) *Human performance, workload, and situational awareness measures handbook*. CRC Press, Boca Raton
29. Franceschini F, Galetto M, Maisano D (2019) Designing performance measurement systems: theory and practice of key performance indicators. Springer International Publishing, Cham
30. Marinescu AC, Sharples S, Ritchie AC, Sánchez López T, McDowell M, Morvan HP (2018) Physiological parameter response to variation of mental workload. *Hum Factors* 60:31–56. <https://doi.org/10.1177/0018720817733101>
31. Argyle EM, Marinescu A, Wilson ML, Lawson G, Sharples S (2021) Physiological indicators of task demand, fatigue, and cognition in future digital manufacturing environments. *Int J Hum Comput Stud* 145:102522. <https://doi.org/10.1016/j.ijhcs.2020.102522>
32. Charles RL, Nixon J (2019) Measuring mental workload using physiological measures: a systematic review. *Appl Ergon* 74:221–232. <https://doi.org/10.1016/j.apergo.2018.08.028>
33. Koppenborg M, Nickel P, Naber B, Lungfiel A, Huelke M (2017) Effects of movement speed and predictability in human–robot collaboration. *Hum Factors Ergon Manuf Serv Ind* 27:197–209. <https://doi.org/10.1002/hfm.20703>
34. Nomura T, Kanda T, Suzuki T, Kato K (2004) Psychology in human-robot communication: an attempt through investigation of negative attitudes and anxiety toward robots. In: RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759). IEEE, Kurashiki, Okayama, Japan, pp 35–40
35. Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry* 25:49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
36. Lang PJ (1980) Behavioral treatment and bio-behavioral assessment: computer applications. In: Sidowski JB, Johnson JH, Williams TA (eds) *Technology in mental health care delivery systems*. Ablex, Norwood, NJ, pp 119–137
37. Hoffman G (2019) Evaluating fluency in human–robot collaboration. *IEEE Trans Hum Mach Syst* 49:209–218. <https://doi.org/10.1109/THMS.2019.2904558>
38. Baraglia J, Cakmak M, Nagai Y, Rao R, Asada M (2016) Initiative in robot assistance during collaborative task execution. In: 2016 11th ACM/IEEE International conference on human–robot interaction (HRI), pp 67–74

39. Empatica: E4 wristband. <https://www.empatica.com/research/e4>
40. Benedek M, Kaernbach C (2010) A continuous measure of phasic electrodermal activity. *J Neurosci Methods* 190:80–91. <https://doi.org/10.1016/j.jneumeth.2010.04.028>
41. Kim H-G, Cheon E-J, Bai D-S, Lee YH, Koo B-H (2018) Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry Investig* 15:235–245. <https://doi.org/10.30773/pi.2017.08.17>
42. Christensen RHB (2019) Ordinal: regression models for ordinal data. <https://CRAN.R-project.org/package=ordinal>
43. Wilkinson GN, Rogers CE (1973) Symbolic description of factorial models for analysis of variance. *J Roy Stat Soc Appl Stat* 22:392–399. <https://doi.org/10.2307/2346786>
44. Agresti A (2002) Categorical data analysis. John Wiley & Sons, New Jersey

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.