Maximising Coefficiency of Human-Robot Handovers through Reinforcement Learning

Marta Lagomarsino[1,2], Marta Lorenzini[1], Merryn Dale Constable[3], Elena De Momi[2], Cristina Becchio[4,5], Arash Ajoudani[1]

***Abstract*— Collaborative robots need to possess the ability to hand objects properly to humans. Earlier studies on robot-to-human handovers have centred around enhancing the human partner's performance and reducing the physical exertion required to grip the object. Nonetheless, robots exhibiting overly altruistic behaviours may generate protracted and awkward movements that create uncomfortable feelings for humans and affect perceived safety and social acceptance. This paper examines whether applying the cognitive science principle that "humans act *coefficiently* as a group" in human-robot collaboration - i.e. maximising the benefits for all parties involved simultaneously - leads to a smoother and more natural interaction. Human-robot *coefficiency* is modelled by online monitoring of human comfort and discomfort indicators and computing robot energy consumption. This score is used by a reinforcement learning problem to adaptively learn the optimal combination of robot interaction parameters to maximise such *coefficiency* during the task execution. Results demonstrated that by acting *coefficiently*, the robot accommodated the individual preferences of the majority of participants and enhanced the human perceived comfort.**

***Index Terms*— Human Factors and Human-in-the-Loop; Mutual Human-Robot Adaptation; Explainable Robotics**

I. INTRODUCTION

Handover is a crucial ability for robots assisting humans in unstructured environments such as factories, households, and hospitals. Previous research has focused on optimising physical aspects of the interaction, such as accurate object transfer and reduced physical exertion for the human partner. Based on a variety of human ergonomic metrics (e.g. distance to a neutral position, overloading joint torque, posture-based observational methods), the robot adapted the position [1], [2] and orientation [3], [4] of the object, and learned its optimal location in space [5], whole-body configuration [6], [7], and accomplished trajectory [8], [9]. However, for human-robot interaction to be seamless and trustworthy, robots should also consider the socio-cognitive aspects of the interaction and aim to match human preferences and skills [10], [11]. This is particularly important in scenarios such as object handovers, where how a robot grasps and configures the object affects the user's perceived safety, comfort, and efficiency in accomplishing the subsequent action.

[1]Human-Robot Interfaces and Interaction Laboratory, Istituto Italiano di Tecnologia, Genoa, Italy.

[2]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy.

[3]Department of Psychology, Northumbria University, Newcastle, UK.

[4]Cognition, Motion and Neuroscience Laboratory, Istituto Italiano di Tecnologia, Genoa, Italy.

[5]Department of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

Corresponding author's email: marta.lagomarsino@iit.it

Research in cognitive science investigating human joint actions has indicated that individuals are sensitive to the aggregate dyad effort and typically act in a coordinated manner as a team [12], [13]. To clarify, when working with others towards a common objective, individuals consider the dyadic interaction as a whole and opt for actions that optimise the overall efficiency of the joint action (also known as *coefficiency*) rather than focusing solely on individual components [14], [15].

The objective of this preliminary research is to enable collaborative robots to learn to make *coefficient* decisions akin to those selected by humans in social contexts. Indeed, robot behaviours that are more natural or human-like tend to be transparent, predictable, and explainable and foster trust in their human partners [6], [16]. Thus, we propose a novel approach in which the robot evaluates the comfort level of the specific human it interacts with, both on a socio-cognitive (i.e. analysing human reaction time and attention distribution) and physical (i.e. monitoring the upper-body kinematics) level, while also considering its internal costs (i.e. energy consumption). Based on this information, we define the human-robot *coefficiency* score, estimating the aggregate efficiency of the two agents (i.e. the robot passer and the human receiver) during the interaction, and we learn through a reinforcement learning (RL) approach the actions that maximise such *coefficiency*. At each robot-to-human handover iteration, the robot explores different values of the considered interaction parameters, i.e. (i) the object orientation, (ii) the interaction distance, and (iii) the velocity in approaching the human partner. It reads the obtained reward, i.e. the aforementioned human-robot *coefficiency* score, and decides whether to exploit the collected information to maximise the short-term reward (by selecting the subsequent interaction parameters accordingly) or keep exploring the environment.

However, achieving robot motions that align with human preferences is challenging. Research has demonstrated that human preferences are highly subjective and can change as individuals become familiar with the task. Moreover, the perceived comfortable distance depends on factors such as the velocity and smoothness of the executed trajectory [17]. Therefore, we propose a system that learns the interaction parameters together to identify the combination that best suits users' individual preferences. The proposed handover learning and adaptation system, which differs from altruistic behaviour commonly adopted in the literature [1], [3], is tested on twelve subjects in a daily activity, where the robot hands over a mug to the human for making coffee[1].

[1]The video can also be found at youtu.be/VYwnkW5AIJU.

II. Human-Robot Coefficiency Model

The section describes how to evaluate *coefficiency* online without disrupting the natural flow of the interaction. To compute human cognitive and physical ergonomics, we adopted a reduced-complexity representation of the human musculoskeletal structure, characterised by a floating-based sequence of rigid links interconnected by N joints featuring $D \leq 3$ degrees of freedom (DoFs) denoted by $\mathbf{q}^{\mathrm{H}} \in \mathbb{R}^{N \times D}$. For the robot, we measure the energy consumed by a M-DoFs manipulator (i.e. $\mathbf{q}^{\mathrm{R}} \in \mathbb{R}^{M}$) to accomplish the trajectory.

A. Human Cognitive Ergonomic Cost

Concerning the social-cognitive aspect of the interaction, the human receiver's reaction time τ and their attention towards the object they need to handle are examined. We measure the time elapsed between the robot motion start and the human motion initiation time, normalised to the total execution time of the robot trajectory [18]. This is because studies on the control of human body motion in social contexts show that human actions that require more planning result in motion initiation latencies [19].

Moreover, behavioural and neuroscientific studies suggested that discomfort and cognitive load usurp executive resources responsible for attentional control, thus increase distraction [20]. To estimate the level of attention toward the task, we consider the head frame, translate it in correspondence to the centre of the head link and tilt it ten degrees to approximate the gaze direction [21] (denoted as Σ_{gaze} from now on, see Fig.1). Consequently, the Cartesian vector expressing the relative position between Σ_{gaze} and Σ_{object}, namely the frame associated with the object that should be handled, is mapped into spherical coordinates (azimuth angle θ, elevation angle φ and radial distance). A fuzzy logic membership function exploiting the Raised-Cosine Filter [22] is then applied to normalise the measured attention angles at each time instant t ($\theta(t)$ and $\varphi(t)$ angles, indicated in Eq.(1) as $\alpha(t)$) in the range $[\alpha_{\min}(t), \alpha_{\max}(t)]$.

$$f(\alpha(t)) = \begin{cases} 1, & \text{if } |\alpha(t)| \leq \alpha_{\min}(t) \\ \frac{1}{2}\Big[1 - \cos\Big(\frac{|\alpha(t)| - \alpha_{\min}(t)}{\alpha_{\max}(t) - \alpha_{\min}(t)}\pi\Big)\Big], & \text{if } |\alpha(t)| > \alpha_{\min}(t) \\ & \;\&\; |\alpha(t)| \leq \alpha_{\max}(t) \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Note that the threshold values on $\alpha(t)$ (i.e. control points $\alpha_{\min}(t), \alpha_{\max}(t) > 0$) depend on the current distance of the moving object from the human operator.

The attention level $\Lambda(t)$ toward the task is thus defined as the product between the normalised azimuth and elevation indicators and values of $\Lambda(t)$ closer to 1 indicate a total focus on the region of interest

$$\Lambda(\theta_{(t)}, \varphi_{(t)}) = f(\theta_{(t)})\, f(\varphi_{(t)}) \in [0, 1] \tag{2}$$

B. Human Physical Ergonomic Cost

Physical comfort is based on the common claim that a human is exposed to physical effort if one of the joints is close to its Range of Motion (RoM) extrema [23]. We parametrise the ergonomic cost for each k-th DoF of the i-th joint at the instant t as

$$\zeta_i^k(t) = \frac{2\min\left\{\left|q^{\mathrm{H}}_{i,k} - q^{\mathrm{H}}_{i,k,\min}\right|, \left|q^{\mathrm{H}}_{i,k} - q^{\mathrm{H}}_{i,k,\max}\right|\right\}}{\left|q^{\mathrm{H}}_{i,k,\max} - q^{\mathrm{H}}_{i,k,\min}\right|} \in [0, 1]. \tag{3}$$

Then, we identify the most stressed DoF for each joint (i.e. the minimum $\zeta_i^k(t)$) and average the effects over N joints

$$\bar{\zeta}(t) = \frac{1}{N}\sum_{i=1}^{N}\Big[\min_{k=1,\dots D} \zeta_i^k(t)\Big]. \tag{4}$$

C. Robot Consumption Cost

The robot efficiency is parametrised in this work by the robot power consumption [24]. At a specific time instant t, the power consumed by the j-th robot joint is obtained by

$$P_j(t) = \left|\tau_j(t)\ \dot{q}^{\mathrm{R}}_j(t)\right|, \tag{5}$$

where $\tau_i(t)$ is the torque applied at j-th joint and $\dot{q}^{\mathrm{R}}_j(t)$ is the j-th joint velocity. We sum up the contributions of all the M robot joints

$$P(t) = \sum_{j=1}^{M} P_j(t). \tag{6}$$

D. Coefficiency of Human-Robot Joint Actions

A human-robot *coefficiency* score is associated with each conjoint action a, representing how efficient the latter is in terms of aggregate costs of the involved agents. More specifically, the score is modelled by integrating the quantities described in the above sections over the entire interaction duration (e.g. pre-handover phase, physical exchange and subsequent action) as follows

$$C^{\mathrm{HR}}_{\text{coefficiency}}(a) = \frac{1}{3}\Big[C^{\mathrm{H}}_{\text{cognitive erg}} + C^{\mathrm{H}}_{\text{physical erg}} + C^{\mathrm{R}}_{\text{energy cons}}\Big] \tag{7}$$

where $C^{\mathrm{H}}_{\text{cognitive erg}}$ and $C^{\mathrm{H}}_{\text{physical erg}}$ parametrise the human efficiency while $C^{\mathrm{R}}_{\text{energy cons}}$ refers to the robot efficiency. In particular, the human cognitive ergonomic cost is defined as

$$C^{\mathrm{H}}_{\text{cognitive erg}}(a) = \frac{1}{2}\Big[(1 - \tau) + \mathop{\mathbb{E}}_{t=t_0,\dots t_{\mathrm{f}}}[\Lambda(t)]\Big]. \tag{8}$$

The formulation is based on a study of human-robot interaction [25], where human body movements were analysed and correlated with subjective evaluations of robot behaviour. Results showed that higher-ranked interactions were characterised by intensive attention and motion synchronisation to the robot and these aspects were equally significant in determining user appreciation. Thus, drawing from that study, we establish that our cognitive cost is positively correlated to the average attention an individual dedicates to the task and is inversely associated with the reaction time (as a higher τ implies less synchronisation to the robot motion).

On the other hand, the physical ergonomic cost

$$C^{\mathrm{H}}_{\text{physical erg}} = \min_{t=t_0,\dots t_{\mathrm{f}}} \bar{\zeta}(t) \tag{9}$$

identifies the worst posture assumed during the interaction.

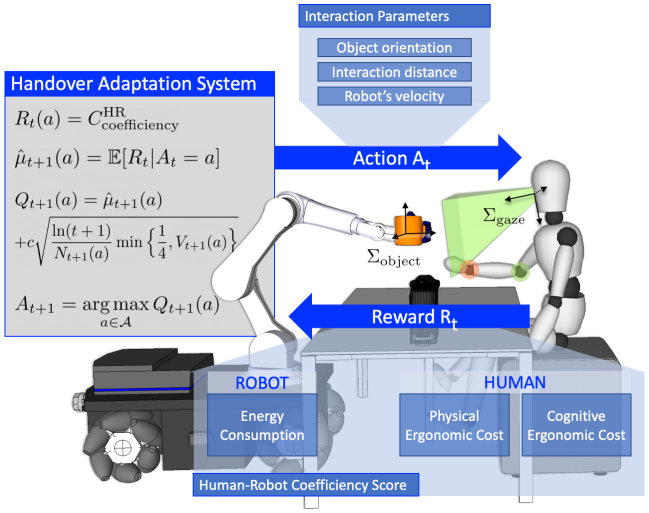

Fig. 1: Overall structure of the proposed framework to transfer human paradigm of acting *coefficiently* in human-robot handovers.

Finally, the robot efficiency is computed as the normalised energy consumption to execute the desired trajectory, e.g.

$$C^{R}_{\text{energy cons}}(a) = 1 - \frac{1}{E_{\max}} \int_{t_0}^{t_f} P(t)dt. \tag{10}$$

The reader should note that the costs defined in Eq.(7), (8), (9), and (10) are normalised in $[0, 1]$, and values of these indexes closer to 1 denote high comfort for the agents.

III. Handover Adaptation System

Our system adapts robot behaviour based on implicit user responses and the robot energy consumption, using a combined score called human-robot *coefficiency*. Specifically, this score is used as reward for a RL algorithm, allowing the robot to learn and adapt online without separating data collection and learning phases (see Fig.1 for a system overview).

A. Adapted Parameters and Observed Reward

The handover strategy in this work involves adjusting three interaction parameters: (i) the object's orientation on the horizontal plane, (ii) the interaction distance, and (iii) the robot velocity profile. These parameters were chosen for easy implementation, the desire to limit the search space's dimensionality, and, above all, their impact on the interaction. Indeed, the ergonomics of the handover process and the ease of completing subsequent actions depend on how the robot positions and orients the transferred object. However, human comfort is also affected by personal characteristics and technology confidence levels [26]. Some individuals prefer extra physical effort for a perceived safer distance [27] and are sensitive to changes in robot velocity profiles [28].

To summarise, optimising the interaction to the user preferences requires considering all the combinations of these parameters simultaneously. Nevertheless, an extreme focus on maximising user convenience could result in protracted and unnatural robot motions and negatively affect perceived safety and social acceptance. Thus, in our RL scenario, we use the aforementioned human-robot *coefficiency* as reward $R_t(a)$ to optimise the handover execution both considering human ergonomics and robot convenience.

B. Multi-Armed Bandit Problem

When learning to interact with a new human partner, an agent must balance between exploring new actions and performing the ones that have earned the highest rewards so far. This is particularly important in human-in-the-loop systems where testing time is limited. To address this dilemma, the RL community proposed the principle of *optimism in the face of uncertainty*, where the agent makes an optimistic guess about the expected reward of each action and selects the one with the highest guess. If the guess is incorrect, the agent updates its knowledge and explores other actions. As the agent learns more about the environment, the effects of optimism decrease, and the policy improves.

Our work focuses on a finite-horizon MAB problem, i.e. a specific form of RL enabling the exploration-exploitation of the environment without changing the state. Other RL techniques would require defining a set of possible states and transition probabilities in a Markov decision process that can not be done for our application. Specifically, we consider a finite set of possible values for each parameter and define a K-armed bandit problem, where each arm corresponds to a robot action with a different combination of interaction parameters. At each iteration t, among the actions $\mathcal{A} \in \mathbb{R}^K$, the robot performs an action (i.e. arm $a \in \mathcal{A}$) and receives a reward $R_t(a)$. Then, the robot updates its internal knowledge about the expected reward

$$\hat{\mu}_{t+1}(a) = \hat{\mu}_t(a) + \frac{R_t(a) - \hat{\mu}_t(a)}{N_t(a)}. \tag{11}$$

Note that the expected reward $\hat{\mu}_{t+1}(a)$ is no more than the average reward associated with the action a estimated iteratively on the basis of the observed reward $R_t(a)$ and the number $N_t(a)$ of times a was taken prior to t.

To improve the robot policy required to select the subsequent action, we use the Upper Confidence Bound (UCB) algorithm [29] that asymptotically achieves the logarithmic regret[2]. For each action a, we compute UCB1-tuned value

$$Q_t(a) = \hat{\mu}_t(a) + c\sqrt{\frac{\ln(t)}{N_t(a)}\min\left\{\frac{1}{4}, V_t(a)\right\}}, \tag{12}$$

where the second term denotes the confidence level of the estimate ($c > 0$). $V_t(a)$ is the upper confidence bound on the variance of the action a, based on rewards obtained until t,

$$V_t(a) = \sum_{k=\{t|A_k=a\}} \frac{\hat{\mu}_k(a)^2}{N_t(a)} - \hat{\mu}_t(a)^2 + \sqrt{\frac{2\ln(t)}{N_t(a)}}, \tag{13}$$

and the factor $1/4$ is the upper bound on the variance of a Bernoulli random variable. On each subsequent pull, the agent picks the action A_t that maximises $Q_t(a)$, namely

$$A_t = \arg\max_{a\in\mathcal{A}} Q_t(a). \tag{14}$$

[2]The regret for a policy is defined as the difference between the reward obtained and the highest expected reward.

The UCB algorithm shifts its focus from prioritising exploration, which involves selecting the least attempted actions, to emphasising exploitation, which chooses actions with the highest estimated rewards.

IV. Experiments

Twelve participants, three men, eight women, and one non-binary (26.1 ± 3.3 years), with no prior experience with robots were recruited to test the proposed framework's ability to make coefficient decisions and improve human-robot interaction[3]. In particular, two research questions were tested: (i) *Are the interaction parameters learned by our framework resulting in efficient actions for the involved agents?* (ii) *Does the proposed coefficiency-based decision-making strategy allow aligning the robot behaviour to the preferences of the human partner?* The experiment involved a collaborative robot (Franka Emika Panda) handing a mug to a human, who then placed it under a coffee machine. We exploited a button board to measure human reaction times and a wearable MVN Biomech suit (Xsens Technologies BV) to measure the kinematics of the right wrist and elbow joints.

The robot utilised an impedance controller to track trajectories computed by smoothly interpolating a sequence of desired configurations. Different robot behaviours were implemented, adapting online the performed trajectory. The starting point of the trajectory was fixed to the robot configuration to grasp the mug on the table. The following configurations the robot passes through and the associated timing law varied according to the parameters learned by the adaptation system. The participants experienced three different final object orientations ($\beta_1 = \pi/6$, $\beta_2 = \pi/2$ and $\beta_3 = 5\pi/6$), two interaction distances ($d_1 = 0.30\text{m}$ and $d_2 = 0.45\text{m}$) and two total execution times of the robot trajectory ($\Delta t_1 = 5.0\text{s}$ and $\Delta t_2 = 8.0\text{s}$). Thus, a twelve-armed bandit problem based on UCB1-tuned is defined. The confidence value was set to infinity ($c = +\infty$) for the unexplored arms, inducing an initial priming round to be performed, in which each action a was sampled once to obtain the initial value of $\hat{\mu}_t(a)$. This avoided divide-by-zero errors in the exploration term $Q_t(a)$ when actions have not yet been tried and $N_t(a)$ is equal to zero. The policy then explored with $c = 0.1$, reduced the uncertainty and learned the optimal combination of parameters for the specific user the robot is collaborating with. For the test, $E_{\max}$ was set to the maximum energy consumed among the proposed trajectories.

After the experiment, participants were asked to select their preferred interaction parameters and rate the naturalness and seamlessness of the handover using a Likert scale. They also used a NASA-developed technique to assess the relative importance of factors in determining the final score.

A. Experimental Results

A statistical analysis using the non-parametric Wilcoxon signed-rank test (WSRT) was conducted to compare the implicit comfort signals acquired while the robot exploited the learned parameters and in all other iterations. Overall, the action efficiency improved in value thanks to the learning (see Fig.2). A significant increase in the human ergonomic cost (as the sum of $C^{\text{H}}_{\text{cognitive erg}}$ and $C^{\text{H}}_{\text{physical erg}}$) was registered by executing the optimal action learned by the system for each specific user ($\text{p}^A = 0.027$)[4]. Moreover, the human-robot *coefficiency* cost experienced a growth of 10.6% in the median ($\text{p}^A < 0.001$). From Fig.3, we can also notice a significant effect of each interaction parameter on the reward of our RL algorithm. Indeed, the mug's orientation and distance learned by the presented policy predominately increased the *coefficiency* of the human-robot dyad (p^β, $\text{p}^d < 0.001$). The same can be stated for the robot velocity ($\text{p}^{\Delta t} = 0.002$) although with lower significance.

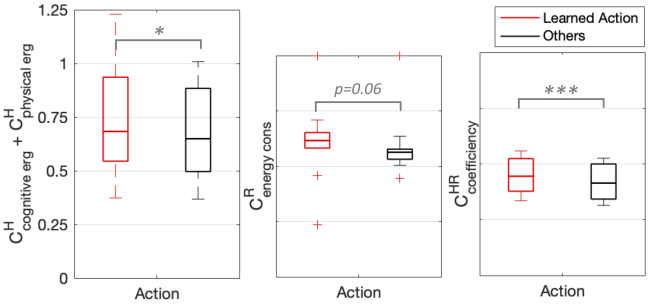

Fig. 2: Comparison of human ergonomics, robot energy consumption and human-robot *coefficiency* score running the action learned by the framework and all the other iterations. Significance levels of Wilcoxon's test are indicated at *p<0.05, **p<0.01, ***p<0.001.

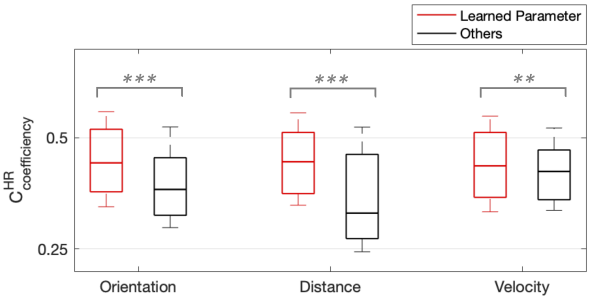

Fig. 3: Comparison of human-robot *coefficiency* score obtained exploiting a specific interaction parameter value (i.e. object orientation, interaction distance and robot velocity) learned by the framework and all the other iterations.

Figure 4 shows the adaptation system's results for 12 participants involved in the experiments. Full red circles represent the learned parameters (i.e. the most selected values over the last twenty-five iterations), red crosses indicate the average parameter value, and blue circles show the preferred values indicated by each participant in the questionnaire.

Considering all subjects, at least two of the parameters reached convergence with the stated preferences within about 7 minutes from the beginning of the interaction, which is,

[3]Experiments were carried out at HRII Lab in accordance with the Declaration of Helsinki, and the protocol was approved by the ethics committee ASL Genovese N.3 (IIT_ERC_IMOVEU version 03.1 29/06/2022).

[4]We denote as p^V the p-value obtained from Wilcoxon test between iterations exploiting learned values and all the others. It should be noticed that V could refer to a specific interaction parameter, i.e. orientation β, distance d or velocity Δt, or a combination of them, i.e. an action A.

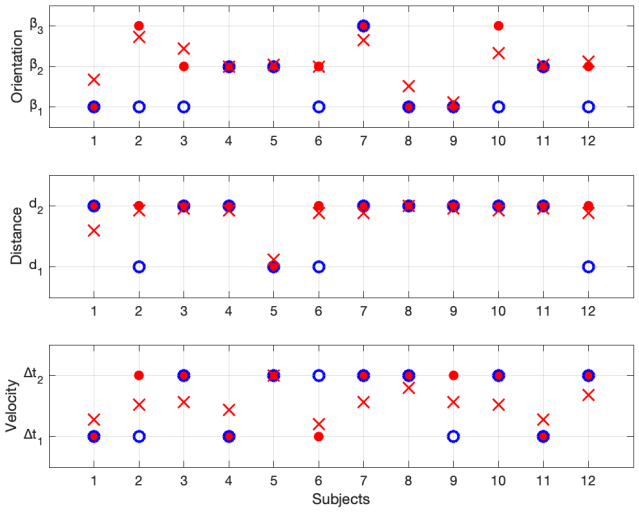

Fig. 4: Learned parameters and preferences for twelve subjects. Red full circles indicate the learned parameters, i.e. mode, and red crosses the weighted average over last twenty-five steps. Reported preferred values are depicted through blue circles.

TABLE I: Results of post-study subjective questionnaires.

Custom Questionnaire	Mean	Std
Q1: The way the robot moved at the end of the experiment met my preferences.	3.58	1.00
Q2: The robot behaved in an awkward and unnatural way.	2.25	1.22
Q3: I felt comfortable while performing the task as I would be with another human.	4.17	0.83
Q4: In planning how to configure the object and hand it over to you, the robot should take into account:		
your mental load and perceived safety.	3.67	0.87
your physical effort.	3.33	1.41
the appropriateness and fluency of robot motion.	4.11	0.93

after 21.1 iterations, on average. Convergence was defined as the human-preferred value being selected by the policy most of the time and chosen at least five times in a row. The learning converged to the preferred orientation for 7 subjects and to the preferred distance and velocity for 9 out of 12 subjects. The contribution of each parameter to a fruitful interaction was determined by computing the mean distance between the learned parameter and the average parameter value during the last iterations, which was 0.13, 0.10, and 0.33 for orientation, distance, and velocity, respectively. The higher this distance, the lower the parameter contribution indicating that the algorithm jumps between its possible values.

Table I presents the results of post-study subjective questionnaires. Participants agreed that the robot motion's appropriateness should be taken into account and considered the proposed costs equally relevant to plan well-coordinated robot behaviours. Indeed, patterns of choices in the custom questionnaire indicated that the means of the weights given by participants to $C^{\mathrm{H}}_{\text{cognitive erg}}$, $C^{\mathrm{H}}_{\text{physical erg}}$, and $C^{\mathrm{R}}_{\text{energy cons}}$ were 0.33, 0.26, and 0.41, respectively.

V. Discussion

The findings demonstrated the effectiveness of the proposed online learning approach in maximising the benefits and reducing the effort required by the agents involved in the collaborative task. The system's ability to learn optimal interaction parameters led to a significant improvement in human cognitive and physical ergonomics, as well as a noticeable decrease in robot expenses. As a result, we successfully incorporated the notion of *coefficiency* based on cognitive and physical factors inspired by human joint action theories into human-robot interactions.

It is worth noting that through acting in a *coefficient* manner, the robot successfully met the individual preferences of most of the subjects involved in the experiments. Despite expressing actual human preferences is not straightforward, the metrics proposed to measure human comfort were found to be suitable for adjusting the robot interaction parameters on the fly and learning the personalised behaviour that best suits the user's needs. Nevertheless, not all parameters have the same impact on determining a fruitful interaction. Parameters that have a more significant effect on *coefficiency* were learned more quickly and accurately, while less significant ones may not even converge. As shown in Fig. 4, the mean distance between the learned velocity and the average parameter value during the last iterations is higher than that obtained for orientation and distance across all subjects. This implies that the robot velocity is less relevant to the decision-making strategy.

The main limitation of the proposed framework is linked to the assumptions made in defining human-robot *coefficiency*. For example, relying on the behavioural analysis in [25], it is expected that participants would shift their attention from the mug when the interaction is annoying and not legible (e.g. the robot moves too slowly or rotates excessively after the handover to return to the homing configuration). But, two participants exhibited behaviours far from our expectations, thus preventing the system from appropriately learning. Subject 12 forced herself to be overfocused and always performed the task in the same way, despite the parameters being far from her preferences, leading to learning interaction parameters based only on robot expenses. Conversely, subject 2 tended to get distracted and delayed the motion initiation when the robot ran actions that were more legible and predictable for him. To address these issues, we could expand the concept of *coefficiency* by including additional variables beyond those currently used to overcome learning difficulties encountered by the framework for some participants.

Although the questionnaire revealed that, on average, subjects ranked the costs equally important to plan a seamless interaction, it would also be valuable to explore the advantages of a personalised model of human-robot *coefficiency* score. By evaluating the weights that each subject would assign to each cost, we could formulate a reward function as a customised weighted combination to address the individual demands and characteristics of the user.

Furthermore, according to subjective impressions reported in the questionnaires, participants perceived well-adapted robot behaviours as natural and appropriate. This positive outcome is a clear indication that the proposed *coefficiency* framework represents a significant step towards developing robots that can be interacted with as seamlessly as humans.

VI. Conclusions

This study examined whether transferring the human paradigm of acting *coefficiently*, i.e. simultaneously maximising the benefits of all involved agents, to human-robot cooperative tasks facilitates a more seamless and natural interaction. We first modelled human-robot *coefficiency* by monitoring implicit indicators of human comfort and discomfort and calculating the energy expended by the robot to accomplish the desired trajectory. Then, we proposed a RL strategy to adapt online the behaviour of the robot, which exploits the human-robot *coefficiency* score as a reward to learn the actions that maximise such *coefficiency*. Initially, the robot explores different interaction parameters, then learns and selects the combination of parameters that best fits human preferences.

The proposed framework showed satisfactory results for ten out of twelve participants, where at least one interaction parameter converged to the preferences stated in the questionnaires. Nonetheless, the occasional contradictory outcomes cast doubt on the dependability of the self-reported values and motivate further exploration of additional variables, such as those related to human body language and emotional cues discussed in literature [30]. Future studies could consider developing a personalised reward function model for each subject to address situations where our costs are not equally relevant or the assumptions are not completely fulfilled.

Overall, the adaptation mechanism developed in this study showed promising features to be applied in more complex cooperative tasks, analysing, for instance, human whole-body movements and stress-related motion patterns such as hyperactivity and self-touching. Ultimately, the study of *coefficiency* in human-robot handovers presented in this paper laid the groundwork for future applications of cognitive psychology to hybrid interaction settings.

Acknowledgment

This work was supported by the ERC-StG Ergo-Lean (Grant No.850932) and The Royal Society (Grant No.IES\R3\203086). The authors thank Dr. Mariacarla Memeo, Dr. James William Ashmore Strachan and Mattia Leonori for their help in experiments.

References

[1] A. Bestick, R. Pandya, R. Bajcsy, and A. D. Dragan, "Learning human ergonomic preferences for handovers," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3257–3264, IEEE, 2018.

[2] W. Kim, M. Lorenzini, P. Balatti, P. D. Nguyen, U. Pattacini, V. Tikhanoff, L. Peternel, C. Fantacci, L. Natale, G. Metta, and A. Ajoudani, "Adaptable workstations for human-robot collaboration: A reconfigurable framework for improving worker ergonomics and productivity," *IEEE Robotics & Automation Magazine*, pp. 14–26, 2019.

[3] J. Aleotti, V. Micelli, and S. Caselli, "An affordance sensitive system for robot to human object handover," *International Journal of Social Robotics*, pp. 653–666, 2014.

[4] P. Ardon, M. E. Cabrera, E. Pairet, R. P. A. Petrick, S. Ramamoorthy, K. S. Lohan, and M. Cakmak, "Affordance-aware handovers with human arm mobility constraints," *IEEE Robotics and Automation Letters*, pp. 3136–3143, 2021.

[5] J. Mainprice, M. Gharbi, T. Simeon, and R. Alami, "Sharing effort in planning human-robot handover tasks," in *Proceedings of IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 764–770, IEEE, 2012.

[6] M. Cakmak, S. S. Srinivasa, M. K. Lee, J. Forlizzi, and S. Kiesler, "Human preferences for robot-human hand-over configurations," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1986–1993, IEEE, 2011.

[7] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human hand-overs," in *Proceedings of International Conference on Human-Robot Interaction (HRI)*, p. 489, ACM Press, 2011.

[8] E. A. Sisbot and R. Alami, "A human-aware manipulation planner," *IEEE Transactions on Robotics*, pp. 1045–1057, 2012.

[9] A. Jain, B. Wojcik, T. Joachims, and A. Saxena, "Learning trajectory preferences for manipulators via iterative improvement," *Advances in Neural Information Processing Systems*, 2013.

[10] M. Lorenzini, M. Lagomarsino, L. Fortini, S. Gholami, and A. Ajoudani, "Ergonomic human-robot collaboration in industry: A review," *Frontiers in Robotics and AI*, p. 262, 2023.

[11] A. Dragan, K. Lee, and S. Srinivasa, "Legibility and predictability of robot motion," in *Proceedings of International Conference on Human-Robot Interaction (HRI)*, pp. 301–308, IEEE, 2013.

[12] J. P. Santamaria and D. A. Rosenbaum, "Etiquette and effort: Holding doors for others," *Psychological Science*, pp. 584–588, 2011.

[13] G. Török, B. Pomiechowska, G. Csibra, and N. Sebanz, "Rationality in joint action: Maximizing coefficiency in coordination," *Psychological Science*, pp. 930–941, 2019.

[14] J. W. Strachan and G. Török, "Efficiency is prioritised over fairness when distributing joint actions," *Acta Psychologica*, p. 103158, 2020.

[15] G. Török, O. Stanciu, N. Sebanz, and G. Csibra, "Computing joint action costs: Co-actors minimize the aggregate individual costs in an action sequence," *Open Mind*, pp. 1–13, 2021.

[16] E. De Momi, L. Kranendonk, M. Valenti, N. Enayati, and G. Ferrigno, "A neural network-based approach for trajectory planning in robot–human handover tasks," *Frontiers in Robotics andAI*, p. 34, 2016.

[17] M. Lagomarsino, M. Lorenzini, E. D. Momi, and A. Ajoudani, "Robot trajectory adaptation to optimise the trade-off between human cognitive ergonomics and workplace productivity in collaborative tasks," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2022.

[18] J. Podda, C. Ansuini, R. Vastano, A. Cavallo, and C. Becchio, "The heaviness of invisible objects: Predictive weight judgments from observed real and pantomimed grasps," vol. 168, pp. 140–145, 2017.

[19] M. Khan, I. Franks, D. Elliott, G. Lawrence, R. Chua, P. Bernier, S. Hansen, and D. Weeks, "Inferring online and offline processing of visual feedback in target-directed movements from kinematic data," *Neuroscience & Biobehavioral Reviews*, pp. 1106–1121, 2006.

[20] R. W. Quinn, G. M. Spreitzer, and C. F. Lam, "Building a sustainable model of human energy in organizations: Exploring the critical role of resources," *Academy of Management Annals*, pp. 337–396, 2012.

[21] U. Weidenbacher, G. Layher, P.-M. Strauss, and H. Neumann, "A comprehensive head pose and gaze database," in *Proceedings of International Conference on Intelligent Environments (IE)*, pp. 455–458, IEEE, 2007.

[22] M. Lagomarsino, M. Lorenzini, E. D. Momi, and A. Ajoudani, "An online framework for cognitive load assessment in industrial tasks," *Robotics and Computer-Integrated Manufacturing*, p. 102380, 2022.

[23] L. McAtamney and E. N. Corlett, "RULA: a survey method for the investigation of work-related upper limb disorders," *Applied ergonomics*, pp. 91–99, 1993.

[24] A. Mohammed, B. Schmidt, L. Wang, and L. Gao, "Minimizing energy consumption for robot arm movement," *Procedia*, pp. 400–405, 2014.

[25] T. Kanda, H. Ishiguro, M. Imai, and T. Ono, "Body movement analysis of human-robot interaction," in *International Joint Conference on Artificial Intelligence*, 2003.

[26] M. Lagomarsino, M. Lorenzini, P. Balatti, E. D. Momi, and A. Ajoudani, "Pick the right co-worker: Online assessment of cognitive ergonomics in human-robot collaborative assembly," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–1, 2022.

[27] E. Hall, *The hidden dimension*. Anchor, 1966.

[28] D. Kulić and E. Croft, "Physiological and subjective responses to articulated robot motion," *Robotica*, pp. 13–27, 2007.

[29] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, pp. 235–256, 2002.

[30] M. Karg, A.-A. Samadani, R. Gorbet, K. Kuhnlenz, J. Hoey, and D. Kulic, "Body movements for affective expression: A survey of automatic recognition and generation," *IEEE Transactions on Affective Computing*, pp. 341–359, 2013.