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], and Arash Ajoudani[1]

***Abstract*—Handing objects to humans is an essential capability for collaborative robots. Previous research works on human-robot handovers focus on facilitating the performance of the human partner and possibly minimising the physical effort needed to grasp the object. However, altruistic robot behaviours may result in protracted and awkward robot motions, contributing to unpleasant sensations by the human partner and affecting perceived safety and social acceptance. This paper investigates whether transferring the cognitive science principle that "humans act *coefficiently* as a group" (i.e. simultaneously maximising the benefits of all agents involved) to human-robot cooperative tasks promotes a more seamless and natural interaction. Human-robot *coefficiency* is first modelled by identifying implicit indicators of human comfort and discomfort as well as calculating the robot energy consumption in performing the desired trajectory. We then present a reinforcement learning approach that uses the human-robot *coefficiency* score as reward to adapt and learn online the combination of robot interaction parameters that maximises such *coefficiency*. Results proved that by acting *coefficiently* the robot could meet the individual preferences of most subjects involved in the experiments, improve the human perceived comfort, and foster trust in the robotic partner.**

***Index Terms*—Human Factors and Human-in-the-Loop; Physical Human-Robot Interaction; Human-Centered Robotics**

I. INTRODUCTION

UNSTRUCTURED environments such as factories without work cells, households, and hospitals, where robots have the potential to assist humans, often involve robot-to-human handovers. Effective handovers are not limited to the accurate and precise transfer of objects from the robot to the human but require physical and cognitive coordination. Previous studies have proposed methods whereby the robot facilitates human action by reducing physical effort. According to various human ergonomic metrics (e.g. distance to a neutral position, overloading joint torque, posture-based observational methods), the robot adjusted the position [1], [2] and orientation [3], [4] of the object, and learned its optimal location in space [5], whole-body configuration [6], and accomplished trajectory [7].

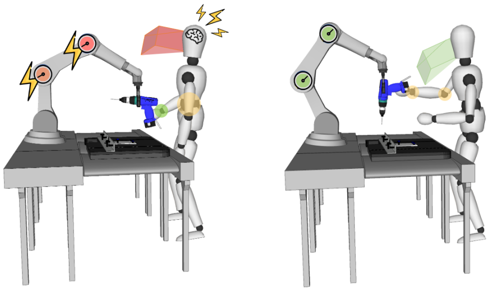

Fig. 1: Illustration of the rationale behind human-robot *coefficient* actions. Maximisation of human physical comfort may require awkward robot motion affecting perceived safety. Contrarily, considering efforts of both agents, a more seamless interaction can be achieved.

Nevertheless, to achieve seamless human-robot interactions, it is preferable that robots also display understanding toward the socio-cognitive aspects of the interaction and aim at matching human preferences and skills [8], [9]. The socio-cognitive patterns of the interaction are as important as the physical ones for robots to be considered partners and not only tools. For example, imagine a robot handing over a drill to a human coworker in a factory. The way the robot grasps and configures the object in the operating space affects the user's comfort, how convenient the tool's physical transfer is, and how efficiently the user can accomplish the subsequent action. An extreme maximisation of the human physical convenience could result in protracted and unnatural robot motions and negatively affect perceived safety and social acceptance [10] (see Fig.1). Although works on planning more legible and predictable robot motions exist in the literature [11], [12], gaps were identified in relation to holistic approaches, which examined all the aspects (socio-cognitive and physical) and phases (of the handover process) of human-robot interaction through a unified lens.

Cognitive science studies investigating human-human joint actions have highlighted that people are sensitive to the aggregate physical and cognitive effort of the dyad and tend to act *coefficiently* as a group [13]. In other words, when cooperating with others to reach a shared goal, people consider the dyadic interaction as a whole and select actions that maximise the overall efficiency of the joint action rather than any individual components [14], [15].

The present study aims to enable robots to learn to make *coefficient* decisions, as humans generally do in human-human interactions. Indeed, robot behaviours that are more natural or human-like are preferred inasmuch they are more readable and foster trust in the human partner [6]. We propose an innovative approach whereby the robot online assesses the comfort of the specific human it interacts with, both at the socio-cognitive and physical levels, and simultaneously takes into account its internal expenses (e.g. motion feasibility and energy consumption). Thereby, the robot uses this information to choose a handover configuration that results in *coefficient*

Manuscript received: January, 19, 2023; Revised March, 17, 2023; Accepted May, 8, 2023. This paper was recommended for publication by Editor Aniket Bera upon evaluation of the Associate Editor and Reviewers' comments. 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. M. Memeo, Dr. J.W. Ashmore Strachan, and M.Leonori for their help in the experiments.

[1]M.Lagomarsino, M.Lorenzini and A.Ajoudani are with Human-Robot Interfaces and Interaction Laboratory, Istituto Italiano di Tecnologia, Genoa, Italy. Corresponding author's email: marta.lagomarsino@iit.it

[2]M.Lagomarsino and E.De Momi are with Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy.

[3]M.D.Constable is with Department of Psychology, Northumbria University, Newcastle, United Kingdom.

[4,5]C.Becchio is with Department of Neurology, University Medical CenterHamburg-Eppendorf, Hamburg, Germany and Cognition, Motion and Neuroscience Laboratory, Istituto Italiano di Tecnologia, Genoa, Italy.

Digital Object Identifier (DOI): NA.

actions for the two agents (i.e. concurrently efficient for the robot passer and the human receiver).

The human-robot *coefficiency* is modelled by online capturing implicit comfort and discomfort body signals from the human partner as well as the robot energy consumption. Specifically, for the former, we estimate human cognitive and physical ergonomics during the interaction by analysing the reaction time, the attention distribution, and the upper-body kinematics. This allows us to estimate the aggregate expenses of the dyad during the interaction, define a human-robot *coefficiency* score, and 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 reward obtained, which is based on the aforementioned human-robot *coefficiency* score. Then, it decides whether to exploit the gathered information to maximise the short-term reward (by selecting the subsequent interaction parameters accordingly) or keep exploring the environment.

Nevertheless, planning robot motions that match human preferences is tricky. Learning biologically inspired robot trajectories is often not enough since human preferences are profoundly subjective and often vary when familiarising the task. For example, it has been shown that the comfortable distance perceived by each specific-subject changes with the feeling of menace in the robot actions and the smoothness of the accomplished trajectory [16]. Hence, we design a system that does not learn each interaction parameter separately but takes into account the considered interaction parameters simultaneously to ultimately find the combination that best fits users' personal preferences.

To the best of our knowledge, this is the first time that the human tendency to act *coefficiently*, which differs from altruistic behaviour commonly adopted in the literature [1], [3], is designed and developed in human-robot handovers. The proposed handover learning and adaptation system is tested on twelve subjects in a daily collaborative activity, where the robot hands over a mug to the human counterpart to prepare some coffee (see multimedia attachment[1]).

II. Human-Robot Coefficiency Model

In this section, we define a set of variables to model *coefficiency* in human-robot joint actions, such as handovers, and describe how to evaluate it online without affecting the natural flow of the interaction.

Regarding human contribution, we investigate variables related to human cognitive and physical ergonomics during task execution. The only requirement here is the online collection of data about the object position and human motion through a suitable sensor system (e.g. visual tracking or IMU-based motion capture systems). The musculoskeletal system of the human body can be modelled by N articulations (joints) and $N+1$ body segments (rigid links). A frame Σ_i is associated with each joint $i \in \{1,\dots,N\}$ and its configuration over time with respect to a world frame Σ_{W} is known thanks to an initial calibration. Each joint can feature $D \leq 3$ degrees of freedom (DoFs), and the angle with respect to its parent link is denoted by $q_i^{\mathrm{H}} \in \mathbb{R}^D$. The kinematic chain imposes a set of constraints on the link's motion patterns, which leads to the definition of the joints Range of Motion (RoM):

$$q_{i,k,\min}^{\mathrm{H}} < q_{i,k}^{\mathrm{H}} < q_{i,k,\max}^{\mathrm{H}},\; i \in \{1,\dots,N\},\, k \in \{1,\dots,D\} \quad (1)$$

For the robot, we measure the energy consumed to complete the planned trajectory. We consider a robotic manipulator constituted by a serial collection of rigid links, which are connected by M revolute joints exhibiting one DoF each (represented by the angle $q_j^{\mathrm{R}} \in \mathbb{R}, j \in \{1,\dots,M\}$).

A. Human Cognitive Ergonomic Cost

Regarding the social-cognitive level of the interaction, we analyse the reaction time τ and the attention that the human receiver gives to the object that he/she has to handle. Indeed, studies on the control of human body motion in social contexts highlight that human actions requiring a more significant amount of planning result in motion initiation latencies [17]. To detect changes in the kinematics of human movements during human-robot interaction, we consider the time elapsed between the time instant in which the robot starts its motion t_0^{R} and the human motion initiation time t_0^{H}, and we normalise the value to the total execution time Δt of the robot's trajectory

$$\tau = \frac{t_0^{\mathrm{H}} - t_0^{\mathrm{R}}}{\Delta t}. \quad (2)$$

Standardising reaction times against the duration of the observed movement is a common practice when considering joint actions with movements of different duration [18].

Moreover, behavioural and neuroscientific studies have provided evidence that discomfort and cognitive load usurp executive resources, which otherwise could be used for attentional control, thus increase distraction [19]. 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 [20] (denoted as Σ_{gaze} from now on, see Fig.2). 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 [19] is then applied to normalise the measured attention angles at each time instant t ($\theta(t)$ and $\varphi(t)$ angles, indicated in Eq.(3) as $\alpha(t)$) in the range $[\alpha_{\min}(t), \alpha_{\max}(t)]$.

$$f(\alpha(t)) = \begin{cases} 1, & \text{if } |\alpha(t)| \leq \alpha_{\min}(t) \\ \frac{1}{2}\left[1 - \cos\left(\frac{|\alpha(t)| - \alpha_{\min}(t)}{\alpha_{\max}(t) - \alpha_{\min}(t)}\pi\right)\right], & \text{if } |\alpha(t)| > \alpha_{\min}(t) \\ & \& \, |\alpha(t)| \leq \alpha_{\max}(t) \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

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. Specifically, the lower limit $\alpha_{\min}(t)$ and upper limit $\alpha_{\max}(t)$ are computed as

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

$$\alpha_{\text{min}}(t) = \tan^{-1}\left(\frac{(1-\gamma)\,r}{d(t)}\right), \quad \alpha_{\text{max}}(t) = \tan^{-1}\left(\frac{(1+\gamma)\,r}{d(t)}\right), \tag{4}$$

where r is the radius of the area in which the object is assumed to be contained (i.e. the fixed distance between the origin of the frame Σ_{object} associated with the object and its boundary), while $d(t)$ is the distance between the origins of Σ_{gaze} and Σ_{object}. γ is set to 0.4 to determine a smooth behaviour (from 1 to 0) of the function $f(\alpha(t))$ in the angle range corresponding to the distance range $[(1-\gamma)\,r,\ (1+\gamma)\,r]$ with respect to the origin of Σ_{object}.

The attention level $\Lambda(t)$ toward the task is therefore 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

$$\Lambda(\theta(t), \varphi(t)) = f(\theta(t))\, f(\varphi(t)). \tag{5}$$

B. Human Physical Ergonomic Cost

The physical comfort is monitored by considering the joints RoM values (see Eq.(1)). We take inspiration from "pen-and-paper" ergonomics indexes, such as Rapid Upper Limb Assessment (RULA) [21] and Rapid Entire Body Assessment (REBA) [22], which claim that a human is exposed to physical effort if one of his/her joints is close to the RoM extrema. Similarly to [23], we parametrise the ergonomic cost for each k-th DoF of i-th joint at instant t as

$$\zeta_i^k(t) = \frac{2\min\left\{\left|q_{i,k}^{\text{H}} - q_{i,k,\text{min}}^{\text{H}}\right|, \left|q_{i,k}^{\text{H}} - q_{i,k,\text{max}}^{\text{H}}\right|\right\}}{\left|q_{i,k,\text{max}}^{\text{H}} - q_{i,k,\text{min}}^{\text{H}}\right|}. \tag{6}$$

Note that $\zeta_i^k(t)$ spans in the interval $[0, 1]$, and the closer the value is to 1.0, the more comfortable posture the human is experiencing. Then, we identify the most stressed DoF for each joint and we average the stress effects over all N joints

$$\bar{\zeta}(t) = \frac{1}{N}\sum_{i=1}^{N}\left[\min_{k=1,\dots D}\zeta_i^k(t)\right]. \tag{7}$$

C. Robot Consumption Cost

The robot efficiency is parametrised in this work by the robot power consumption [24]. Indeed, the latter is sensitive to robot behaviours that could affect human comfort (and thus would like to avoid), e.g. unnatural and too fast motions (resulting in higher torques and velocities, respectively). At a specific instant t, the power consumed by the j-th robot joint is

$$P_j(t) = \left|\tau_j(t)\ \dot{q}_j^{\text{R}}(t)\right|, \tag{8}$$

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

$$P(t) = \sum_{j=1}^{M} P_j(t). \tag{9}$$

D. Coefficiency of Human-Robot Joint Actions

A human-robot *coefficiency* score is associated with each conjoint action a executed in $t \in [t_0, \dots, t_{\text{f}}]$ 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

$$C_{\text{coefficiency}}^{\text{HR}}(a) = \frac{1}{3}\left[C_{\text{cognitive erg}}^{\text{H}} + C_{\text{physical erg}}^{\text{H}} + C_{\text{energy cons}}^{\text{R}}\right] \tag{10}$$

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

$$C_{\text{cognitive erg}}^{\text{H}}(a) = \frac{1}{2}\left[(1-\tau) + \mathop{\mathbb{E}}_{t=t_0,\dots t_{\text{f}}}[\Lambda(t)]\right]. \tag{11}$$

where $\mathop{\mathbb{E}}_{t=t_0,\dots t_{\text{f}}}$ indicates the mean value of human attention level during the handover entire execution. This formulation is based on a study of human movements behaviour in human-robot interaction [25]. The latter analysed human body movements in several aspects (such as eye contact, distance, synchronisation to the robot, and touches) and computed the correlations with subjective evaluations of robot behaviour. Results indicated that higher-ranked human-robot interactions were characterised by periods of intensive attention and motion synchronisation to the robot. In particular, multiple linear regression analysis revealed that these two aspects were the most relevant and equally significant ones (standardised partial regression coefficients were 0.476 and 0.535, respectively). Hence, taking inspiration from that study, we formulate our cognitive cost to be positively correlated to the average attention an individual gives to the task and negatively correlated with the reaction time (indeed the higher is τ, the less synchronised the human is to the robot motion).

On the other hand, the physical ergonomic cost

$$C_{\text{physical erg}}^{\text{H}} = \min_{t=t_0,\dots t_{\text{f}}}\bar{\zeta}(t) \tag{12}$$

identifies the worst posture assumed during the interaction. We did not consider accumulative values since the interaction was quick, nor the averaging factor to avoid underestimating bad postures assumed for a short period of time.

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

$$C_{\text{energy cons}}^{\text{R}}(a) = 1 - \frac{1}{E_{\text{max}}}\int_{t_0}^{t_{\text{f}}} P(t)dt. \tag{13}$$

The reader should note that the costs defined in Eq.(10), (11), (12), and (13) are normalised in $[0, 1]$, and values of these indexes closer to 1 denote high comfort for the agents.

III. Handover Adaptation System

Our robot behaviour adaptation system observes subconscious responses from the user revealing his/her comfort as well as the energy efficiency of the robot itself, and online adjusts the interaction parameters to maximise these monitored signals. The latter are combined in the human-robot *coefficiency* score (described in the previous section) and then used as the reward that the system should maximise. Exploiting a RL approach based on a Multi-Armed Bandit (MAB) algorithm, the robot explores and learns on the fly without separating the data collection and learning phases (system overview in Fig.2).

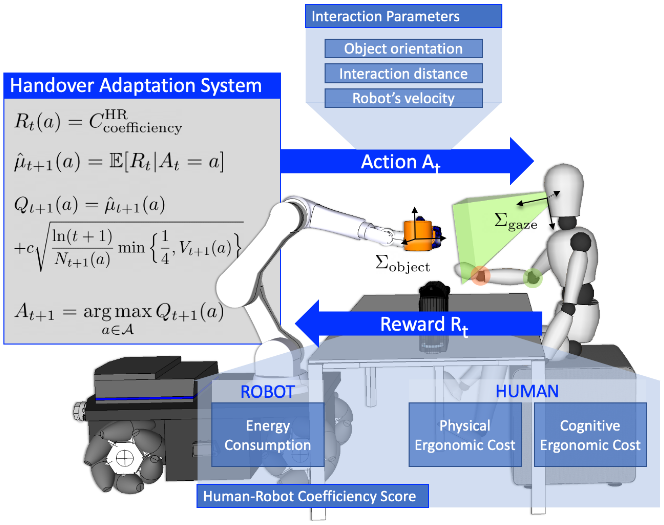
Interaction Parameters
Object orientation
Interaction distance
Robot's velocity
Handover Adaptation System
$R_t(a) = C^{\text{HR}}_{\text{coefficiency}}$
$\hat{\mu}_{t+1}(a) = \mathbb{E}[R_t|A_t = a]$
$Q_{t+1}(a) = \hat{\mu}_{t+1}(a)$
$+c\sqrt{\frac{\ln(t+1)}{N_{t+1}(a)} \min\left\{\frac{1}{4}, V_{t+1}(a)\right\}}$
$A_{t+1} = \arg\max_{a \in \mathcal{A}} Q_{t+1}(a)$
Action A_t
Σ_{gaze}
Σ_{object}
Reward R_t
ROBOT
Energy Consumption
HUMAN
Physical Ergonomic Cost
Cognitive Ergonomic Cost
Human-Robot Coefficiency Score

Fig. 2: Overall structure of the proposed framework to transfer human paradigm of acting *coefficiently* in human-robot handovers.

A. Adapted Parameters

In this work, the adaptation of the handover strategy operates on three interaction parameters, i.e. (i) the object orientation on the horizontal plane, (ii) the interaction distance, and (iii) the robot velocity profile. The selection of these parameters was motivated by their straightforward implementation, the desire to keep the dimensionality of the search space limited, and, above all, their irrefutable impact on the interaction. Indeed, how the robot positions and orients the transferring object determines the ergonomics of the physical exchange and how convenient is the completion of the subsequent action. However, human comfort deeply depends on person-specific physical characteristics such as the length of body segments and joints RoM. Moreover, the confidence level with technology influences human perceived comfort during robot motion and affects the amount of planning required by the human [26]. Some subjects may prefer to incur a slightly extra physical effort if it would make the robot keep a distance perceived as safer [27]. These feelings also vary with the velocity profiles adopted by the robot to approach the human [28]. Hence, to summarise, optimising the interaction to the user preferences requires considering all the combinations of these parameters simultaneously.

B. Observed Reward

Despite the numerous studies in the literature, fully human-centred approaches are not guaranteed to be the best choice to achieve seamless human-robot interactions [10]. One could claim that the closer the object is and with the grasp affordance region more oriented to the user, the more comfortable the interaction is. However, an extreme maximisation of the user's physical convenience could result in protracted and unnatural robot motions, contributing to unpleasant sensations by the human partner and affecting perceived safety and social acceptance. For this reason, in our RL scenario, the reward observed R_t at each iteration t keeps into account variables related to human ergonomics but also the convenience for the robot to execute the action. Namely, we considered

$$R_t(a) = C^{\text{HR}}_{\text{coefficiency}}(a). \tag{14}$$

The sensitivity to excessive robot expenses is motivated by the desire to enable natural and pleasant human-robot interactions.

C. Multi-Armed Bandit Problem

An agent learning how to optimally interact with a new human partner quickly faces the exploration-exploitation dilemma. Namely, it must decide whether to continue exploring new actions or to perform the one that has earned it the highest rewards so far. Dealing with this trade-off efficiently is crucial in human-in-the-loop systems like ours, where testing time is limited. To tackle this challenge, the RL community introduced the principle of *optimism in the face of uncertainty*. This heuristic states that, despite the lack of knowledge about the environment, the agent makes an optimistic guess about how good the expected reward of each action is and selects the action with the highest guess. If the model is correct, the agent has no regrets, otherwise it updates its internal knowledge, diminishing the optimistic guess associated with that action and thus inducing the exploration of other actions. As the agent resolves its uncertainty, the effects of optimism diminish, and the agent's policy approaches optimality.

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. More 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 (as the agent) chooses an action (i.e. arm $a \in \mathcal{A}$) to perform and receives a reward $R_t(a)$. Then, the robot updates its internal knowledge about the expected reward

$$\hat{\mu}_{t+1}(a) = \hat{\mu}_t(a) + \frac{R_t(a) - \hat{\mu}_t(a)}{N_t(a)}. \tag{15}$$

Note that the expected reward $\hat{\mu}_{t+1}(a)$ is no more than the average reward associated to 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. This formulation is memory efficient since it avoids storing all past rewards and re-calculating the average at each time step.

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

$$Q_t(a) = \hat{\mu}_t(a) + c\sqrt{\frac{\ln(t)}{N_t(a)} \min\left\{\frac{1}{4}, V_t(a)\right\}}, \tag{16}$$

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, computed on the basis of the rewards obtained until t,

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

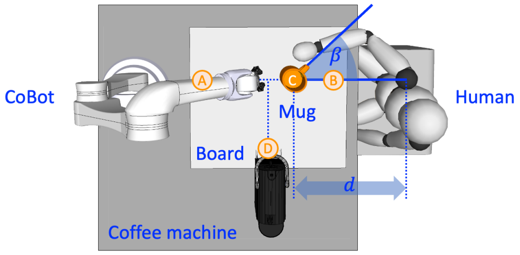

Fig. 3: Overview of the experimental setup involving collaborative robot, human partner, and object for handover (i.e. mug).

$$V_t(a) = \sum_{k=\{t|A_k=a\}} \frac{\hat{\mu}_k(a)^2}{N_t(a)} - \hat{\mu}_t(a)^2 + \sqrt{\frac{2\ln(t)}{N_t(a)}}, \quad (17)$$

and the factor $1/4$ is the upper bound on the variance of a Bernoulli random variable. This means that the action a, which has been played $N_t(a)$ times during the first t plays, has a variance that is at most the sample variance plus $\sqrt{2\ln(t)/N_t(a)}$. On each subsequent pull, the agent picks the action A_t that maximises $Q_t(a)$, namely

$$A_t = \arg\max_{a\in\mathcal{A}} Q_t(a). \quad (18)$$

It can be noticed that UCB moves from focusing primarily on exploration (when the actions that are tried the least are preferred) to instead concentrating on exploitation (selecting the action with the highest estimated rewards).

IV. Experiments

In this section, we describe an experimental analysis conducted to assess the capability of the proposed framework in learning to make *coefficient* decisions and evaluate its potential for improving human-robot interaction[3]. More specifically, two research questions (RQs) were tested:

RQ1. *Are the interaction parameters learned by our framework resulting in efficient actions for the involved agents?*

RQ2. *Does the proposed coefficiency-based decision-making strategy allow aligning the robot behaviour to the preferences of the human partner?*

Note that preferences are meant for interaction parameters, and perceived naturalness, appropriateness and trust.

A. Experimental Protocol

Twelve healthy subjects, three men, eight women, and one non-binary (26.1±3.3 years), were recruited in the experiments. Participants signed written informed consent, declared to have never interacted with a manipulator before and were naïve to the experimental purpose. We considered a scenario where a robot performs a day-to-day task next to a person (setup in Fig.3). Specifically, the collaborative robot (Franka Emika Panda) picked a mug from position A and handed it to the human (in C) sitting at the table. The latter, starting from B, grasped the object and placed it under the coffee machine (i.e. D). We asked participants to repetitively perform the action fifty times in the most natural way as they would not be observed. To measure the kinematics of human motion, we exploited a wearable MVN Biomech suit (Xsens Tech.BV) based on inertial measurement unit sensors. For this study, we restricted the analysis of physical ergonomics to the right wrist and elbow joints since they are the most involved in this task. Note that all participants were right-handed. Besides, the object position was estimated online by considering a fixed transformation from the robot grasping position during the handover phase and from the tracked human wrist during the post-handover action. A board with buttons in the locations mentioned above was also designed to assess human motion initiation time and thus the reaction time τ more precisely. Indeed, participants were trained to keep pressing the button until the action of grasping the mug.

B. Parameters Adaptation

The robot ran an impedance controller and tracked trajectories computed by smoothly interpolating a sequence of desired configurations. Different robot behaviours were implemented by adapting online the performed trajectory. The trajectory starting point was fixed to the robot configuration to grasp the mug on the table in A. The following poses 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 $\beta_3 = 5\pi/6$, obtained by evenly sampling the range identified in [30] by the final angle a human passer places the handle to facilitate the grasp of a human receiver, see Fig.3), two interaction distances ($d_1 = 0.30$m and $d_2 = 0.45$m, i.e. the middle and the limit of the *intimate distance* range proposed in [27]) and two total execution times of the robot trajectory ($\Delta t_1 = 5.0$s and $\Delta t_2 = 8.0$s, in line with natural and fast human movements registered in [31]). Thus, a twelve-armed bandit problem based on UCB1-tuned is defined as presented in Sec.III-C. 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 (decrease of second term in Eq.16) 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.

C. Subjective Questionnaires

After the experiment, we asked participants to select the parameters they felt more comfortable for the interaction (i.e. preferred object orientation, distance and robot velocity). Moreover, they ranked the naturalness and seamlessness of the handover technique using five-point Likert scale questions (see Table I where scoring 1 indicates that the subject strongly disagrees and 5 strongly agrees with the statement). The evaluation included a technique developed by NASA to assess the relative importance of factors in determining the final

[3] The 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).

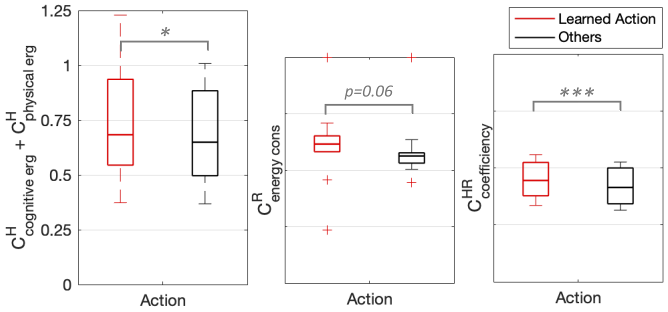

Fig. 4: Comparison of human ergonomics, robot energy consumption and human-robot *coefficiency* score running the action learned by the framework and all the other iterations.

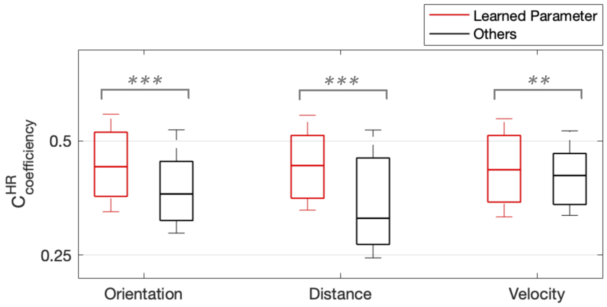

Fig. 5: Comparison of human-robot *coefficiency* score obtained exploiting a specific interaction parameter value learned by the framework and all the other iterations. Wilcoxon test significance levels are indicated at *p<0.05, **p<0.01, ***p<0.001.

score, i.e. the coefficients of their combination. Pairs of costs involved in the *coefficiency* score were presented, and subjects were asked to select which of the two should be taken more into consideration while planning robot motions. From the pattern of choices, we computed the weights that each subject would associate with costs in Eq.10. A copy of the custom questionnaire can be found as supplementary materials for this paper. Finally, they filled the trust scale defined in [32][4]to assess the perceived appropriateness of the robot motion and pick-up speed, cooperation safety, and reliability.

V. Experimental Results

A. Learning and Adaptation Results

A statistical analysis using the non-parametric Wilcoxon signed-rank test (WSRT) was conducted to compare the efficiency metrics computed when the robot exploited the learned parameters and in all other iterations (RQ1.). As can be seen in Fig.4, overall, the efficiency of the actions performed by both the involved agents improved in value thanks to the learning. 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 when executing the optimal action learned by the system for each specific user ($p^A = 0.027$)[5]. Moreover, the human-robot *coefficiency* cost experienced a growth of 10.6% in the median ($p^A < 0.001$). From Fig.5, we can also notice a significant

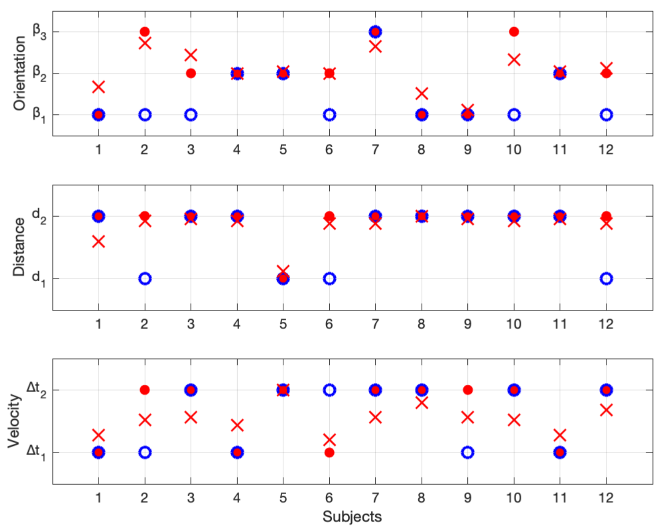

Fig. 6: Learned parameters (red full circles) and preferences (blue circles) for twelve subjects. Red crosses indicate the weighted average over last twenty-five steps.

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}, $p^{d} < 0.001$). The same can be stated for the robot velocity ($p^{\Delta t} = 0.002$) although with lower significance.

Figure 6 shows the result of the adaptation system, i.e. the learned mug's orientation, interaction distance and robot velocity to maximise human-robot *coefficiency*, in comparison to the preferences stated by the twelve subjects involved in the experiments (RQ2.). In the plots, the full red circles represent the learned parameters, which are the most selected values (i.e. the variable mode) over the last twenty-five iterations of each experiment run. Since the algorithm keeps searching for the optimal value, we also computed the average parameter value in the same interval and reported it in the figure through red crosses. Finally, we depicted the preferred values indicated by each participant in the questionnaire as blue circles.

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, after 21.1 iterations, on average. In this work, we consider the convergence achieved when the human-preferred value is selected by the policy most of the time (i.e. the mode over all iterations performed until then is the human-preferred value) and at least five times in a row. The learning converged to the preferred orientation for seven subjects and to the preferred interaction distance and robot velocity for nine out of twelve subjects (see Fig.6). To determine the contribution of each parameter to a fruitful interaction, we computed the mean distance over all subjects between the learned parameter (red circle) and the average parameter value during the last iterations (red cross) and we obtained 0.13, 0.10, 0.33 for orientation, distance, and velocity respectively. The higher this distance, the lower the parameter contribution since it means the algorithm keeps jumping among its possible values. Besides, an average improvement of 4.6% in the *coefficiency* score was registered with preferred parameters ($p^{A_{\text{pref}}} = 0.07$).

[4] Note that the authors of [32] provide the average outcome and standard deviation of each questionnaire item following the experimental trials on one hundred and fifty-five subjects at Cranfield and Loughborough Universities.

[5] p^V denotes the p-value between iterations exploiting learned values and all the others and V is a specific interaction parameter, i.e. orientation β, distance d or velocity Δt, or a combination of them, i.e. an action A.

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

Charalambous Trust Scale	Mean	Std
Perceived robot motion and pick-up speed	3.79	1.41
Perceived cooperation safety	4.31	0.73
Perceived robot reliability	4.25	1.06
Total	12.35	1.97

B. Subjective Questionnaires

Table I reports the results of post-study survey. From the top half of the table, we can see that participants agreed that the robot motion's appropriateness and fluency should be taken into account. Interestingly, patterns of choices in the custom questionnaire indicate that participants consider the proposed costs equally relevant to plan well-coordinated behaviours. On average, the weights given 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.

The bottom half of the table shows the Charalambous trust scores normalised in the interval $[1, 5]$. Note that gripper-related items were removed. With respect to the reference values in [32], an increase of 110.7% was registered for the perceived appropriateness (not normalised) of the robot motion and pick-up speed. Moreover, the proposed robot adaptation system slightly improved the perceived cooperation safety and reliability (1.5% and 3.7%, respectively). Overall, well-coordinated robot behaviours determined a predominant trust growth (17.8%) in naïve participants.

VI. Discussion

Results highlighted the ability of the proposed learning strategy to online maximise the benefits and minimise the effort of the specific agents involved in the cooperative task. Indeed, a statistically significant improvement in human cognitive and physical ergonomics was registered by exploiting the optimal interaction parameters learned by the system, and the robot expenses were noticeably reduced. This means that we succeeded in embedding a concept of *coefficiency* based on cognitive and physical factors inspired by theories of human joint actions into human-robot interactions.

Interestingly, acting *coefficiently*, the robot was able to meet the individual preferences of most of the subjects who participated in the experiments. The proposed metrics of human comfort and discomfort (presented in Sec.II) were found to be appropriate for adjusting the robot interaction parameters on-the-fly and learning the personalised behaviour that best fits the user needs. However, not all parameters have the same impact on defining a fruitful interaction. Parameters more affecting human-robot *coefficiency* are learned faster and more accurately, while less relevant ones may even not converge. As seen in Fig.6, the mean distance over all subjects between the learned velocity and the average parameter value during the last iterations is higher than the ones obtained for the orientation and distance. We can deduce that the robot velocity is less relevant to the decision-making strategy.

Nevertheless, measuring and expressing actual human preferences is not straightforward. For example, subject 5 claimed that the robot did not facilitate his grasping even if the proposed object orientations evenly covered the operating space and parameters converged to the stated preferences. This makes us question the reliability of the self-reported values and complicates the evaluation of the system's accuracy.

The main limitation of the framework is related to the assumptions in the definition of human-robot *coefficiency*. For example, relying on the behavioural analysis in [25], we expect that participants divert attention from the mug when the interaction is annoying and not legible (i.e. the robot moving too slowly or performing an extensive rotation after the handover to come back 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 manner although the parameters were far from her preferences. Hence, the interaction parameters were learned only based on the robot expenses. On the contrary, subject 2 tended to get distracted and protracted the motion initiation when the robot ran actions that were more legible and predictable for him. To solve these issues, we may extend the concept of *coefficiency*, including more variables in addition to those currently exploited to fulfil learning inabilities encountered by the framework for some participants. Moreover, at this stage, we limited the search space's dimensionality given the potential countertrend of the efficiency costs, which may affect the convergence of the algorithm, and we exploited previous knowledge in the field of robot-to-human handovers. The promising results of this study encourage us to consider a wider range of interaction parameters in follow-up works and quantify the gain of capturing *coefficiency* rather than any individual components optimisation through a within-subjects experiment. In general, the reader should note that this work does not aim to substantially improve the score *per sé*. Still, it mainly focuses on presenting new metrics for a more natural interaction and providing initial proof of their potential utility.

Although the questionnaire revealed that, on average, subjects ranked the costs equally important to plan a seamless interaction, it would also be interesting to investigate the benefits of a subject-specific model of human-robot *coefficiency* score. Assessing the weights that each subject would associate with each cost, we could define a reward function as a personalised weighted combination to address the individual demands and characteristics of the user.

In addition, the subjective impressions reported in the questionnaires suggested that well-adapted robot behaviours are perceived as natural, improve the perceived motion appropriateness and foster trust in the automated partner with respect to the values provided in [32] for early comparison. This remarkable outcome indicates that the proposed *coefficiency* framework is a step forward on the way to developing robots that can be interacted with as easily as humans.

VII. Conclusions

This study investigated whether transferring the human paradigm of acting *coefficiently*, i.e. maximising the partner's benefits while being sensitive to own expenses, to human-robot cooperative tasks promotes a more seamless and natural interaction. We first modelled human-robot *coefficiency* by detecting implicit indicators of human comfort and discomfort and computing the energy expended by the robot to accomplish the desired trajectory. Then, we proposed an RL approach to online adapt the robot behaviour, which exploits the human-robot *coefficiency* score as a reward to learn the actions that maximise such *coefficiency*. More specifically, the robot starts exploring by adjusting the value of different interaction parameters, then learns and selects the combination of them that ultimately best fits human preferences.

The framework performed well for ten out of twelve participants, indeed at least an interaction parameter converged to the preferences specified in the questionnaires. However, occasional contradictory results raise doubts about the reliability of self-reported values and encourage us to investigate more variables related to human-body language and emotional cues presented in the literature. Future studies may also consider designing a subject-specific model of the reward function to cope with situations where our costs are not equally relevant or assumptions are not completely fulfilled. Moreover, it would be interesting to analyse the impact of variations of every parameter on the proposed costs and investigate the system adaptation over time.

Overall, the adaptation mechanism developed in this study showed promising features to be applied in more complex collaborative tasks, involving, for instance, direct physical contact with the robot, and human whole-body movements and analysing stress-related motion patterns. Finally, the study of *coefficiency* in human-robot handovers built the foundation for future applications of cognitive psychology to hybrid interaction settings.

References

[1] A. Bestick, R. Pandya, R. Bajcsy, and A. Dragan, "Learning human ergonomic preferences for handovers," in *Proceedings of International Conference on Robotics and Automation (ICRA)*, pp. 3257–3264, 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] B. Busch, G. Maeda, Y. Mollard, M. Demangeat, and M. Lopes, "Postural optimization for an ergonomic human-robot interaction," in *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2017.

[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 International Conference on Intelligent Robots and Systems (IROS)*, pp. 1986–1993, IEEE, 2011.

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

[8] V. Ortenzi, A. Cosgun, T. Pardi, W. P. Chan, E. Croft, and D. Kulic, "Object handovers: A review for robotics," *IEEE Transactions on Robotics*, pp. 1855–1873, 2021.

[9] 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.

[10] D. A. Norman, "Human-centered design considered harmful," *Interactions*, pp. 14–19, 2005.

[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] F. Stulp, J. Grizou, B. Busch, and M. Lopes, "Facilitating intention prediction for humans by optimizing robot motions," in *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, pp. 1249–1255, IEEE, 2015.

[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] M. Lagomarsino, M. Lorenzini, E. De 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 International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2022.

[17] 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.

[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," *Cognition*, pp. 140–145, 2017.

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

[20] 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.

[21] 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.

[22] S. Hignett and L. McAtamney, "Rapid entire body assessment (REBA)," *Applied Ergonomics*, pp. 201–205, 2000.

[23] S. Gholami, M. Lorenzini, E. De Momi, and A. Ajoudani, "Quantitative physical ergonomics assessment of teleoperation interfaces," *IEEE Transactions on Human-Machine Systems*, pp. 169–180, 2022.

[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 *Proceedings of Int. Joint Conference on Artificial Intelligence*, 2003.

[26] M. Lagomarsino, M. Lorenzini, P. Balatti, E. De 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] R. J. Kirschner, H. Mayer, L. Burr, N. Mansfeld, S. Abdolshah, and S. Haddadin, "Expectable motion unit: Avoiding hazards from human involuntary motions in human-robot interaction," pp. 2993–3000, 2022.

[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. D. Constable, A. P. Bayliss, S. P. Tipper, A. P. Spaniol, J. Pratt, and T. N. Welsh, "Ownership status influences the degree of joint facilitatory behavior," *Psychological Science*, pp. 1371–1378, 2016.

[31] I. Georgiou, C. Becchio, S. Glover, and U. Castiello, "Different action patterns for cooperative and competitive behaviour," *Cognition*, pp. 415–433, 2007.

[32] G. Charalambous, S. Fletcher, and P. Webb, "The development of a scale to evaluate trust in industrial human-robot collaboration," *International Journal of Social Robotics*, pp. 193–209, 2016.