Modulations of one's sense of agency during human-machine interactions: a behavioral study using a full humanoid robot

Aïsha Sahaï[1,2,*], Emilie Caspar[3], Albert De Beir[4], Ouriel Grynszpan[5], Elisabeth Pacherie[1], Bruno Berberian[2]

[1]Institut Jean-Nicod, Département d'Etudes Cognitives, ENS, EHESS, CNRS, PSL Research University, Paris, France

[2]ONERA, The French Aerospace Lab, Département Traitement de l'Information et Systèmes, Salon-de-Provence, France

[3] Social & Moral Brain Lab, Department of Experimental Psychology, Ghent University, Belgium

[4]Robotics & Multibody Mechanics Research Group, Vrij Universiteit Brussel (VUB), Bruxelles, Belgium

[5]Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, LIMSI-CNRS, Université Paris-Sud, Orsay, France

*** Correspondance :**
Aïsha Sahaï
asahai.pro@gmail.com

Keywords: shared agency, joint Simon task, humanoid robot, servomotor, human-robot interaction, cooperation.

Abstract

Although previous investigations reported a reduced sense of agency when individuals act with traditional machines, little is known about the mechanisms underpinning interactions with human-like automata. The aim of this study was twofold: (1) to investigate the effect of the machine's physical appearance on the individuals' sense of agency, and (2) to explore the cognitive mechanisms underlying the individuals' sense of agency when they are engaged in a joint task. Twenty-eight participants performed a joint Simon task together with another human or an automated artificial system as a co-agent. The physical appearance of the automated artificial system was manipulated so that participants could cooperate either with a servomotor or a full humanoid robot during the joint task. Both participants' response times and temporal estimations of action-output delays (i.e., an implicit measure of agency) were collected. Results showed that participants' sense of agency for self- and other-generated actions sharply declined during interactions with the servomotor compared to the human-human interactions. Interestingly, participants' sense of agency for self- and other-generated actions was reinforced when participants interacted with the humanoid robot compared to the servomotor. These results are discussed further.

1 Introduction

The sense of agency refers to the experience of being in control of a voluntary performed action (Gallagher, 2000; Pacherie, 2007). During the last decades, a significant amount of research examining the sense of agency has been carried out in the context of individual self-generated actions (e.g., Barlas, Hockley, & Obhi, 2018; Haggard, Clark, & Kalogeras, 2002; Renes, van Haren, Aarts, & Vink, 2015; Sato & Yasuda, 2005; Sidarus & Haggard, 2016, Wen & Haggard, 2020). Thus, it has been pointed out that during self-generated individual actions, individuals' brain actively constructs their sense of agency by using a combination of both internal sensorimotor signals (e.g., feed-forward cues, proprioception, and sensory feedbacks) and circumstantial signals (e.g., intentions, thoughts, and contextual cues) (Moore & Fletcher, 2012; Synofzik, Vosgerau, & Voss, 2013).

Yet, in recent years, there has been an increasing interest in understanding the emergence of this sense of agency for cooperative behavior where actions are intentionally produced by two persons acting together. Initial results have suggested the possible transformation of the agentive awareness and identity in such a cooperative context, from a sense of individual agency to a sense of joint agency (e.g., Dewey, Pacherie & Knoblich, 2014, Bolt, Poncelet, Schultz, & Loehr, 2016; Grynszpan, Sahaï, Hamidi, Pacherie, Berberian, et al., 2019; Jenkins, Esemezie, Lee, Mensingh, Nagales, et al., 2021; Le Bars, Devaux, Nevidal, Chambon, & Pacherie, 2020; Obhi & Hall, 2011a; Strother, House, & Obhi, 2010). Indeed, on the one hand, it is suggested that during the joint task, both individuals would experience a sense of agency for their own actions and their outcomes (i.e., their individual parts of the joint task), here proposed to be called sense of *self-agency*. Concurrently, on the other hand, both individuals would experience a form of agency for the actions and outcomes generated by their co-agent (i.e., their co-agent's parts of the joint task), here proposed to be called sense of *vicarious agency*. This dual presence of both the sense of *self-agency* and the sense of *vicarious agency* over the partner's contributions during a cooperative task where the self–other distinction remains intact is therefore taken as evidence for the emergence of a form of joint agency, here proposed to be called sense of *shared agency* (Pacherie, 2012; Silver, Tatler, Chakravarthi, & Timmermans, 2020). It is to be noted that different types of joint agency have been highlighted by prior work according to the degree of cooperation between the actors during the joint task (Silver, et al. 2020). Hence, Pacherie (2012) and Silver and colleagues (2020) proposed that the sense of joint agency could be regarded as a sense of *we-agency* when the self–other distinction is blurred during the joint task (e.g., Obhi & Hall, 2011a) and rather regarded as a sense of *shared agency* when the self–other distinction remains intact during the joint task. In this study, we will focus on this second form of interaction.

This experience of shared agency during these cooperative joint tasks is an essential aspect of human cooperativeness. Indeed, the development of an agentive experience during a joint task can influence both the objective outcome quality (Babcock & Loewenstein, 1997) and the subjective perception of the outcome quality thereby influencing whether people continue to engage in the joint task (Caruso, Epley, & Bazerman, 2006). Nevertheless, due to the increasing place of automated artificial systems in our daily lives, an important issue remains the emergence of this sense of *shared agency* during interactions that involve artificial partners. Previous work has highlighted individuals' difficulties in developing a sense of *shared agency* during joint tasks with computer co-agents (Obhi & Hall, 2011b; Sahaï, Desantis, Grynszpan, Pacherie, & Berberian, 2019). Actually, it has been proposed that these difficulties could stem from humans' inability to simulate or represent the computer-generated actions in their cognitive system (see Sahaï, Pacherie, Grynszpan, & Berberian, 2017 for a comprehensive review). Indeed, the ability to simulate an observed (or guessed) other-generated action allows the simulation content to be used to predict the consequence of the observed (or guessed) other-generated action, improving implicit action understanding and the experience of being in control as during individual actions (Blakemore & Frith, 2005; Frith, Blakemore, & Wolpert, 2000; Kilner, Friston, & Frith, 2007; Picard & Friston, 2014). At the empirical level, the representation in one's own cognitive system of a co-agent generated actions can be assessed using the joint version of the Simon task (Sebanz, Knoblich, & Prinz, 2003). In the standard Simon task, participants had to detect two types of targets with two different response keys. Results showed that their performance decreased when the target appeared in an incongruent location with respect to their response key (Simon and Small, 1969). This occurred because two action representations (i.e., the correct action to perform and the spatially-induced automatic activated action) are activated and the participant has to solve the conflict in order to select the accurate behavior. By contrast, when participants had to detect only one type of target, there was no effect of location congruency. Intriguingly, during the joint version of the Simon task (Sebanz, et al., 2003) in which the double target detection task was distributed across two persons (i.e., each agent was responsible for only one type of target), the interference effect for the incongruent target-response key mapping reappeared.

Accordingly, previous investigation aimed at examining empirically the possible link between the representation of a co-agent's action and the development of the sense of *shared agency* during a joint Simon task. Indeed, in a previous experiment, Sahaï and colleagues (2019) coupled together a joint Simon target detection task wherein participants' response times (RTs) served as an index of action co-representation (Sebanz, Knoblich, & Prinz, 2003; but see Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz, et al., 2011) and an intentional binding task wherein time estimation served as an implicit measure of participants' sense of agency (Haggard, et al., 2002). The intentional binding phenomenon refers to a subjective temporal compression between a voluntary action and its sensory outcome. Importantly, this temporal binding seems to reliably occur in situations in which the participant is an intentional agent, but not with passive movements (Haggard, Clark, & Kalogeras, 2002). In the authors' task (Sahaï et al., 2019), participants had to perform the Simon task jointly with a co-agent. They were requested to detect colored dots (e.g., green dots) that appeared on a screen either on the same side as the accurate response key (e.g., right key) or on the opposite side (e.g., left side), which corresponded to a co-agent's current location. Throughout the task, the co-agent had to alternately detect a different type of dots (e.g., red dots) with a different response key (e.g., left key). The type of co-agent was manipulated so that participants could interact either with another human being or with an unseen computer. Accurate target detections were always followed by an auditory tone after a particular delay. Participants were requested to estimate the delay between the onset of the target detection (that could be either self- or other-generated) and the onset of the subsequent auditory tone. The originality of this joint task consisted in the fact that the two agents performed actions alternately so that temporal estimations for the other-generated actions only indicated the participant' sense of *vicarious agency* for the co-agent's actions. In fact, in previous studies that focused on individuals' sense of agency during joint actions (e.g., Dewey, et al., 2014; Obhi & Hall, 2011a), the participant's action and the co-agent's action were simultaneously performed. As a consequence, this made it difficult to specifically explore the participants' sense of *vicarious agency* for the actions generated by the co-agent excluding their own performance. The results of Sahaï and colleagues (2019)'s study indicated that participants exhibited a stronger sense of agency for their partner-generated actions than for their own self-generated actions during the human-human cooperation, suggesting a loss of sense of *self-agency* in this particular context of joint action. Importantly, participants were able to exhibit a sense of *vicarious agency* for the other-generated action when the co-agent was another human being but not when it was a computer. This paralleled the RTs results demonstrating faster self-generated responses when the target appeared at the same location as the response key in comparison with the opposite location when they were cooperating with another human being but not with a computer. This stimulus-response congruency effect (or Social Simon Effect, SSE) has been shown to derive from the cognitive interference that occur when two different representations of actions are concurrently activated (Simon & Wolf, 1963). Hence, it could be said that participants co-represented the actions performed by the human co-agent but this ability was impaired for the computer-generated actions.

Given the important role of prediction in both joint action (Sebanz, Bekkering, & Knoblich, 2006; Sebanz & Knoblich, 2009) and agency development (Sahaï, et al. 2017), these difficulties in representing the action of the artificial partner may disturb action understanding and prediction, which may explain the difficulties in developing the sense of *shared agency* when interacting with a computer co-agent. Moreover, recent neurophysiological investigations have underlined that the sense of *shared agency* exhibited by two human individuals during a joint task was correlated with inter-brain synchronization (Shiraishi & Shimada, 2021). Yet, this cerebral activity has been shown to be decreased during human-computer cooperation (Hu, Pan, Shi, Cai, Li, & Chen, 2018), suggesting that individuals were unable to neurally bind with the computer, as well as a lack of engagement (Schilbach, Timmermans, Reddy, Costall, Bente, et al., 2013) with this type of machine.

Nevertheless, the large variety of automated artificial systems facing us, with varying complexities from single-unit levers as well as desktop computers to full human-like machines, must be taken in consideration. More in details, little is known about the specific contribution of the external appearance of the machine in the alteration of the sense of *shared agency* during human-machine interactions. Yet, there is evidence that during a joint task, anthropomorphized robots, in contrast to traditional machines, can elicit the representation of the machine-generated actions in the human brain. Indeed, human-like appearance favors the attribution of an intentional agency to robots and evokes attitudes similar to those governing human social interactions (Wiese, Meta, Wykowska, 2017). For instance, studies in neuroimaging have shown that, under certain constraints (the human-like appearance, notably), the neural mechanisms involved in action understanding are activated for both human-robot and human-human interactions (Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012; Gallagher, Jack, Roepstorff, & Frith, 2002; Krach, Hegel, Wrede, Sagerer, Binkofski, & Kirsher, 2008; Takahashi, Terada, Morita, Suzuki, Haji, et al., 2014; Wang & Quadflieg, 2015). Moreover, it has been shown that during hand-over interactions between a human and a robotic arm, the predictability of the robotic arm motions for the human was strongly dependent on the automaton's motion laws and physical appearance (Glasauer, Hubert, Basili, Knoll, & Brandt, 2010). Indeed, the authors showed that when the robotic arm was handing on a cube to the human seated in front of it, the human's RTs to grasp the cube were faster when the robot assumed human-like kinematics in comparison with a trapezoidal joint velocity (i.e., a typical robotic motion), meaning that individuals were able to better predict the observed human-like movement endpoints. Interestingly, the effect of the kinematic profile on the RTs was modulated by the external appearance of the robot: when the robotic arm had a humanoid appearance, its human partner had faster RTs than when the robotic arm had an industrial appearance, suggesting a better motion prediction. Moreover, the human's RTs tended to be faster when the robotic arm had a typical robotic motion profile but a humanoid appearance than when the robotic arm had a human-like kinematic but an industrial appearance. Finally, previous work on social robotics investigated the human ability to represent actions that have been performed by a humanoid robot in one's cognitive system

(Stenzel, Chinellato, Bou, del Pobil, Lappe, et al., 2012). In the study by Stenzel and colleagues (2012), the participants were sitting next to a full humanoid robot described either as an intelligent and active agent or as a passive machine acting in a deterministic way. The participants had to detect one type of target (e.g., a white square) that could appear on the left or the right side of a screen. The task of the robot was to detect another type of target (e.g., a white diamond) on the same screen. Interestingly, the authors found a SSE in the participants' RTs when the robot was introduced as a human-like robot who can actively act but not when the robot was introduced as a deterministic machine. Hence, this finding pointed out that representation of machine-generated actions could also occur during a joint task with a humanoid robot provided that the robot was considered as an active partner. Possibly, to envisage the other as similar as oneself is needed in order to map their actions into one's own cognitive system during a joint task. Therefore, the first objective of the current study consisted in investigating the impact of an artificial system's physical appearance (i.e., human-like or not) on both individuals' sense of *self-agency* and *vicarious agency* using a paradigm that allows the measurement of action representation.

In addition, the second objective of the current study was to investigate the underlying cognitive mechanisms of the modulation of the sense of *self-agency* towards an experience of a sense of *shared agency* during a joint task with a machine. Indeed, it has been established that individuals could build a sense of "we-ness" during human-human joint tasks (Crivelli & Balconi, 2010; Dewey, et al., 2014; Obhi & Hall, 2011a). Moreover, previous work has highlighted that egocentric sensory predictions were less involved in the construction of the agentive experience during joint action, with respect with individual actions. For example, some authors reported that individuals had a general bias towards claiming more explicit control than they objectively had over a performed joint action (Dewey, et al., 2014; van der Wel, Sebanz, & Knoblich, 2012), indicating a modulation of the *self-agency* experience during human-human joint actions. However, whether such a new "we-identity" is constructed during human-robot interactions remains unclear as most of the study have focused on actions that have been generated by a computer (Obhi & Hall, 2011b, Sahaï et al., 2019) or by low-level robotics (Grynszpan, et al., 2019).

In this context, the aim of the current study was twofold: (1) to investigate the effect of the robot's physical appearance on the individuals' sense of *self-agency* and *vicarious agency*, (2) to explore the cognitive mechanisms underlying the sense of *shared agency* when individuals are engaged in a joint task with a machine. We ran a modified version of Sahaï and colleagues (2019)' paradigm. In the current study, the type of co-agent was manipulated so that the participants could perform the task jointly with another human, a full humanoid robot, or a servomotor. All accurate target detection triggered an auditory tone after a certain delay. We investigated the participants' sense of agency for the individual parts of the joint task: the sense of agency over the participant's *own part* of the joint task (here called sense of *self-agency*), and the sense of agency over their *partner's part* of the joint task (here called sense of *vicarious agency*). Particularly, the participants had to estimate the temporal delay between the onset of the target detection (either self- or other- generated) and the onset of the tone. This measure served as an implicit measure of participants' sense of agency (intentional binding phenomenon, Haggard, et al., 2002). We hypothesized that the more similar to the participants the co-agent would be, the stronger the participants' sense of *vicarious agency* would be, mainly due to their ability to better simulate and predict their co-agent's actions and outcomes. We also hypothesized a shift from a sense of *self-agency* to a sense of *shared agency* with the human and human-like co-agents but not with the servomotor due to the foreseeable construction of the "we-identity" with the first two agents.

2 Method

2.1 Ethic statement

This study was approved by the institutional ethical research committee of the Université libre de Bruxelles (Belgium, N° 008/2016). The investigation was carried out in accordance with the Declaration of Helsinki and all participants provided their written informed consent before starting the experiment. All participants were assigned a number in order to ensure the anonymity of the data.

2.2 Participants

Participants were recruited trough social medias. Twenty-eight healthy adults volunteered to take part in the experiment (22 women, 24 right-handed, mean age 23.61 years, SD: 3.52 years). Two power analyses tested for repeated measures and within-factors ANOVA were run using G*Power application (Faul, Erdfelder, Lang, & Buchner, 2007) in order to estimate the minimal required sample size to highlight differences on participants' RTs and temporal interval estimations. The significance threshold was set at $\alpha = .05$ and the power at $1\text{-}\beta = .90$ for both power analyses. Based on the parameters reported in the previous study by Sahaï and colleagues (2019), the first power analysis revealed that a sample of 9 participants was needed to exhibit a SSE on participants' RTs when considering three types of Co-agent (Human, Human-like, Servomotor), two levels of Congruency (Congruent, Incongruent), and an effect size defined by partial $\eta^2 = .42$ (SSn = 3200.03 and SSd = 4484.45). Moreover, because the authors' study did not report any significant Co-agent x Congruency x Agent interaction on participants' temporal estimations in their previous investigation, an a priori medium effect size defined by partial $\eta^2 = .09$ was considered in the second power analysis. In this later analysis, we found that a sample size of 21 participants was needed to exhibit differences on participants' temporal estimations when considering three types of Co-agent (Human, Human-like, Servomotor), two levels of Congruency (Congruent, Incongruent), and two levels of Agent (Self, Other). Therefore, the minimal required sample size in the current study consisted of a sample of 21 participants. A little over participants were finally tested in order to compensate for potential data loss. All participants had normal or corrected-to-normal vision. None of them had prior knowledge about the purpose of the experiment. Participants were paid €30 for their participation in the experiment.

2.3 Materials and stimuli

Two desktop computers were used to allow pairs of participants to run the experiment in parallel.

Visual stimuli consisted of three dots of 0.5 cm diameter: a white dot, a blue dot, and a yellow dot. An auditory tone (1000 Hz and 200 ms duration), presented via a headphone, was used during the experiment as a sensory consequence of the agent's key presses for measuring intentional binding. Moreover, the use of headphones made it possible to mask the sound naturally generated by the co-agent's actions (e.g., the sound outputted from the effector in motion or from the key presses).

The type of co-agent participants interacted with was manipulated using a within-participants design so that participants successively interacted with another human, a full humanoid robot

named Pepper, and a servomotor in a counterbalanced order (see **Figure 1** for pictures of the two robots). Robots such as Pepper belong to a class of robots designed to engage people at an interpersonal and socio-affective level (Breazeal, Takanishi, & Kobayashi, 2008), and are called social robots (see Fong, Nourbakhsh, & Dautenhahn, 2003 for a discussion of the concept of social robot). However, in order to control prior belief or expectations about the robots, neither Pepper nor the servomotor interacted with the participants before the testing phase. Hence, both machines were already powered up and placed at the suitable location when participants entered in the testing room. Participants were told that they would have to perform a joint task with different co-agents, without being introduced to each other. During the task, Pepper's key presses were performed with the help of its fist, and the servomotor's key presses were performed according to a toggle mechanism of a pivoting bar.
When the humanoid robot or the servomotor performed key presses, their RTs were randomly taken from a normal distribution computed from the mean and standard deviation of naïve participants' RTs during a previous similar experiment (Sahaï, et al., 2019). Hence, the co-agents' RTs to detect the target were similar in all experimental conditions, that is to say, when the co-agent was another human, a humanoid robot, and a servomotor.

Stimuli presentation and robot-generated actions were controlled using PsychoPy software (2_PY3 version).

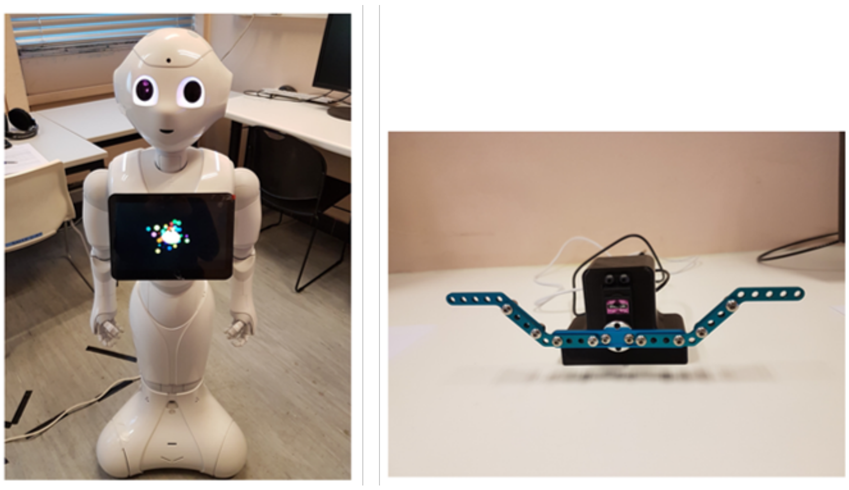

Figure 1. The humanoid robot and the servomotor used as co-agents during the experiment.

2.4 Procedure

Once arrived in the experimental room, participants were asked to give their informed and written consent before to take part in the experiment. Participants and their co-agent were seated on each side of a screen. They had to detect, as quickly and as accurately as possible, colored dots that appeared either on the left or the right side of a central fixation cross. Participants' co-agent could be either another naïve participant, a full humanoid robot, or a servomotor (see **Figure 2**). During the human-human interactions, the participant and his/her co-agent were matched both by gender and handiness.

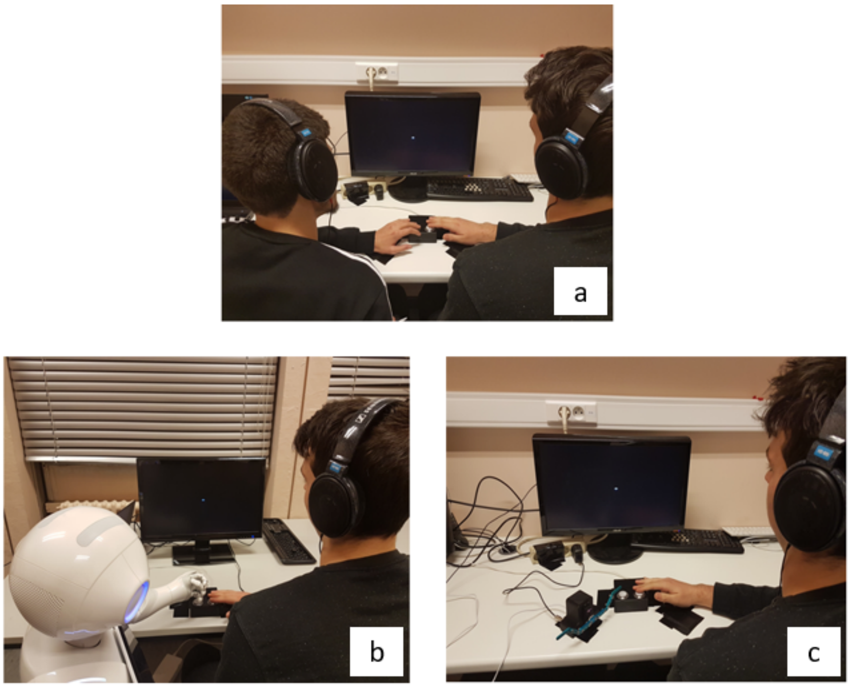

Figure 2. The experimental display of the experiment when the participants performed the joint task with another human (a), the humanoid robot (b), and the servomotor (c).

Each trial started with a fixation cross that appeared at the center of the screen for 500 ms followed by the immediate apparition of the target dot. According to the color of the target dot (either blue or yellow), the participants or their co-agent (i.e., the other human, the humanoid robot, or the servomotor) had at most 1000 ms in order to press their assigned response key (e.g., left or right key, counterbalanced across participants) otherwise an error message appeared, and the trial was canceled. Participants were informed of the onset of their own action and those of their co-agent in real-time with the help of the presentation of a white dot (with the same size as the target dot) that was displayed above the target dot for a duration of 200 ms. Participants were required to look at the computer screen throughout the experiment and not to look at the actions performed by their co-agents. Each correct target detection was followed by an auditory tone presented after the key press at one of two possible Stimulus Onset Asynchronies (SOA) of 400 or 1200 ms, randomly selected. However, participants were told that this delay was totally random and that it could vary between 100 ms and 2000 ms. After the presentation of the auditory tone, participants had to write on a sheet of paper the perceived duration between the onset of the target detection (self- or other- generated, indicated by the white dot appearing on the target) and the onset of the auditory tone (see **Figure 3** for a summary). The sheet of paper consisted of several empty rows, each corresponding to a specific trial of the experiment. Participants were requested to report the temporal estimation for the corresponding trial in the accurate row. These time interval estimations served as an implicit measure of the participants' sense of agency (Haggard et al., 2002). Moreover, at each change of partner, participants were trained at the beginning of the block to estimate and report their perceived duration of the action-tone intervals. During this training, they were presented with two different colors dots that appeared sequentially with a random time interval comprised between 100 ms and 2000 ms. Participant were told to write on a sheet of paper the perceived duration of this interval in milliseconds. Then, participants were given a feedback with the correct delay that appeared on the screen in order to accurately recalibrate their internal clock. This training session consisted of 25 trials. Thereafter, participants performed 16 trials of the forthcoming experimental condition as training. The aim of this training was to familiarize participants with the task so that they would associate their key presses with the subsequent auditory tone.

The mappings between the color of the target dot (e.g., blue or yellow) and the accurate response key that was associated with (e.g., left or right key) were counterbalanced across participants but stayed constant throughout all the experiment for a given participant. For the participants' trials, every trial was coded as "Congruent" when the target appeared on the side of the participant's response key, and as "Incongruent" when the target appeared on the opposite side of the participant's response key. Moreover, for the co-agent's trials, every trial was coded as "Congruent" when the target appeared on the side of the partner's response key, and as "Incongruent" when the target appeared on the opposite side of the partner's response key. Participants completed a total of 720 trials (3 Co-agents (Human, Humanoid robot, Servomotor) ×2 Agents (Self, Other) × 2 Congruency levels (Congruent, Incongruent) × 2 Delays (400, 1200) × 30 trials).

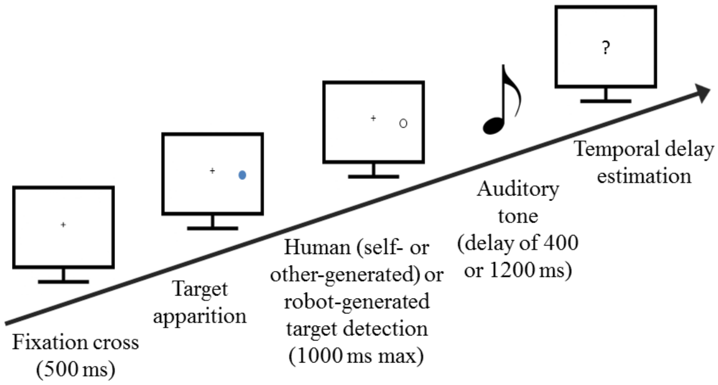

Figure 3. A trial timeline. The trial started with a fixation cross that appeared for 500 ms. Then, the target dot (either blue- or yellow- colored) appeared. According to the color of the target, participants or their co-agents had to detect the target by pressing a specific key (either the left or the right key) within a time window of 1000 ms. Every target detection was signaled by the target becoming white-colored. An auditory tone was generated at one of the two possible a SOA (either 400 ms or 1200 ms) following the target detection. Participant had to report the perceived temporal delay between the onset of the target detection (self- or other-generated) and the onset of the tone.

3 Data analyses

Our first dependent measure was the participants' mean target detection RT. Our second dependent measure was the participants' mean perceived action-tone temporal interval. To distinguish the participants' trials from the co-agent's trials, the participants' trials were labeled "*Self* trials", and the co-agents' trials were labeled "*Other* trials". Statistical analyses were performed with R software (3.3.1 version). Extreme values (i.e., the values that were below or above 2 standard deviations from the mean) of the participants' RTs were excluded from further analyses in order to eliminate outliers and allow for robust statistical analyses. These rejections represented 7% of the raw data. Previous work using the Simon task have already used a similar approach in the pre-processing of the RTs (Wylie, Ridderinkhof, Bashore, & van den Wildenberg, 2010; Sahaï, et al., 2019). The significance level was set at α = .05. In addition, post-hoc comparisons were made using the False Discovery Rate (FDR) correction (Benjamini & Hochberg, 1995).

3.1 Response times (RTs) analyses

This analysis was based exclusively on the data gathered in the conditions wherein participants performed an action (i.e. the *Self* trials). An analysis of variance (ANOVA) was computed on the participants' mean RTs with Co-agent (Human, Pepper, Servomotor) and Congruency (Congruent, Incongruent) as within-factors, and Hand (Dominant, Non-dominant) as a between-factor. The Hand factor was included in the ANOVA because some studies reported that handiness asymmetries could impact the stimulus-response congruency effect (Rubichi & Nicoletti, 2006; Seibold, Chen, & Proctor, 2016). Note should be taken that the Target (Blue dot, Yellow dot) factor was not included in the ANOVA because the stimulus-response congruency effect does not rely on the target identity (i.e., its color) but rather on the congruency between the location of the target and the location of the response key. Lastly, the Delay (400 or 1200) factor was irrelevant for this analysis as the auditory tone was produced after the participants' response and therefore could not influence their RTs.

3.2 Temporal interval estimations analyses

The so-called Intentional Binding (IB) phenomenon was used as an implicit measure of the sense of agency. This phenomenon refers to the individuals' illusory temporal attraction between the onset of a generated action (e.g., a key press) and the onset of its sensory consequence (e.g., a tone) which occurs only when the action has been intentionally triggered (Haggard, et al., 2002). IB is known as a robust implicit measure of sense of agency (for a review, see Moore & Fletcher, 2012).

An analysis of variance (ANOVA) was computed on the participants' mean temporal estimations with Co-agent (Human, Pepper, Servomotor) and Congruency (Congruent, Incongruent) as within-factors and Agent (Self, Other) as a between-factor. The Congruency factor was included in the ANOVA in order to investigate foreseeable effects of the conflictual action selection context (e.g., on incongruent trials) (Sidarus & Haggard, 2016) on the participants' mean temporal estimations. The action-tone Delay (400 or 1200 ms) was not included in the ANOVA as a separate factor. Indeed, the participants' temporal estimations for both delays were averaged as we were interested in the way the social context could influence the temporal interval estimations, rather than the influence of different temporal interval lengths on the reported temporal estimations. In the current study, the variations in interval lengths were made to avoid a predictability bias.

4 Results

4.1 Response times (RTs)

We aimed to examine the occurrence of the Social Simon Effect (SSE) during a joint task with regard to the nature of the co-agent (Human, Humanoid robot, Servomotor). We assessed the normality of the RTs distributions of the differences between the congruent trials and the incongruent trials, separately for each type of co-agent, using the Shapiro-Wilk test. The analyses showed that none of the RTs distributions deviated from normality (all $W > 0.90$ and all $p > .10$). Hence, a 3 × 2 x 2 analysis of variance (ANOVA) with Co-agent (Human, Humanoid robot, Servomotor) and Congruency (Congruent, Incongruent) as within factors and Hand (Dominant, Non-dominant) as a between factor was computed. A significant main effect of Congruency was found on the participants' mean RTs, indicating longer participants' mean RTs on Incongruent trials compared to Congruent trials ($F(1,26) = 43.98$, $p < .001$, partial $\eta^2 = .63$). Moreover, there was no significant main effect of Co-agent ($F(2,52) = 1.90$, $p = .16$, partial $\eta^2 = .07$) or Hand ($F(1,26) = .04$, $p = .85$, partial $\eta^2 = .001$) on the participants' mean RTs. However, a significant interaction between Congruency and Co-agent was found

($F(2,52) = 6.53$, $p = .003$, partial $\eta^2 = .20$, **Figure 4**) on the participants' mean RTs. Other interactions did not reach significance (all *p*s > .05).

Post-hoc comparisons investigating the Congruency x Co-agent interaction revealed that the participants' mean RTs on Incongruent trials were significantly longer than the participants' mean RTs on Congruent trials when the Co-agent was a Human (respectively, 344.08 ms (95%CI = [328.87; 359.30]) and 325.16 ms (95%CI = [310.21; 340.10]), *pFDR* < .001, Cohen's $d = .32$) and a Humanoid robot (respectively, 355.39 ms (95%CI = [339.75; 371.03]) and 336.72 ms (95%CI = [321.31; 352.13]), *pFDR* < .001, Cohen's $d = .26$) but not when the Co-agent was the Servomotor (respectively, 355.71 ms (95%CI = [340.21; 371.21]) and 349.86 ms (95%CI = [331.29; 368.44]), *pFDR* = .12, Cohen's $d = .11$). Hence, the SSE was observed both when the participants performed the task with another human and with the humanoid robot. On the contrary, no SSE was observed when the participants interacted with the servomotor.

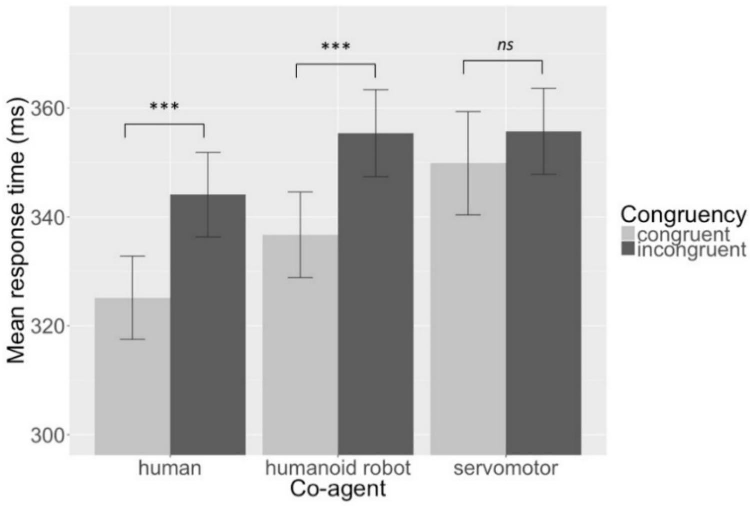

Figure 4. The Congruency x Co-agent interaction on the participants' mean response times in the joint Simon task according to the type of Co-agent. Error bars represent standard errors. All tests were two-tailed. *** corresponds to a p value < .001.

4.2 Temporal interval estimations

We aimed to examine the influence of the social context and the target congruency on the participant's perceived temporal interval estimations between the onset of a performed action, either self- or other- generated, and the onset of a subsequent auditory tone. A 3 × 2 x 2 analysis of variance (ANOVA) with Co-agent (Human, Pepper, Servomotor), Congruency (Congruent, Incongruent), and Agent (Self, Other) as within-factors. A significant main effect of Co-agent ($F(2,54) = 5.36$, $p = .008$, partial $\eta^2 = .17$, see **Figure 5**) was found on the participants' mean temporal estimations. This main effect indicated that the participants' mean temporal estimations were shorter during the joint task with another human compared to the humanoid robot (respectively, 631.42 ms (95%CI = [563.13; 699.71]) and 676.92 ms (95%CI = [578.00; 775.85]), *pFDR* = .03, Cohen's $d = .44$)) and to the servomotor (758.89 ms (95%CI = [664.22; 853.56]), *pFDR* <.001, Cohen's $d = .64$). In addition, the participants' mean temporal estimations were shorter during the joint task with the humanoid robot compared to the Servomotor (*pFDR* = .001, Cohen's $d = .57$). Furthermore, no significant main effect of Congruency ($F(1,27) = 3.48$, $p = .07$, partial $\eta^2 = .11$) or Agent ($F(1,27) = .64$, $p = .43$, partial $\eta^2 = .02$) was found on the participants' mean temporal estimations. However, a significant Congruency x Agent x Co-agent interaction was found on the participants' mean temporal estimations (*HFe* = .78, p = .02, partial η^2 = .15, see **Figure 6**).

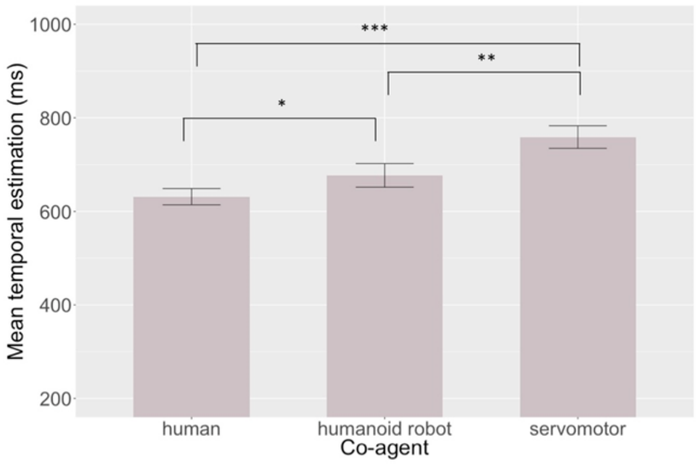

Figure 5. The main effect of Co-agent. Error bars represent standard errors. All tests were two-tailed. * corresponds to a p value < .05, ** corresponds to a p value < .01 and *** corresponds to a p value < .001

Post-hoc comparisons investigating the Congruency x Agent x Co-agent interaction revealed that during Congruent trials, the participants' mean temporal estimations did not differ between the Self and the Other trials when the Co-agent was another human (respectively 633.50 ms (95%CI = [573.90; 693.09]) and 634.30 ms (95%CI = [554.84; 713.77]), *pFDR* = .97, Cohen's d = .01), the humanoid robot (respectively 678.50 ms (95%CI = [586.00; 771.00]) and 679.28 ms (95%CI = [574.86; 783.70]), *pFDR* = .97, Cohen's d = .01), and the Servomotor (respectively 735.49 ms (95%CI = [640.80; 830.17]) and 732.26 ms (95%CI = [632.06; 832.45]), *pFDR* = .90, Cohen's d = .02). However, during the Incongruent trials, a significant difference indicated that the participants' mean temporal estimations were longer during the Self trials compared to the Other trials when the Co-agent was the Servomotor (respectively 832.62 ms (95%CI = [745.65; 919.59]) and 735.18 ms (95%CI = [639.03; 831.34]), *pFDR* = .02, Cohen's d = .48). By contrast, the participants' mean temporal estimations during the Incongruent trials did not differ between the Self trials and the Other trials when the Co-agent was another human (respectively 635.28 ms (95%CI = [571.47; 699.09]) and 622.59 ms (95%CI = [550.33; 694.84]), *pFDR* = .45, Cohen's d = .14) or the humanoid robot (respectively 660.05 ms (95%CI = [567.37; 752.72]) and 689.87 ms (95%CI = [579.79; 799.95]), *pFDR* = .23, Cohen's d = .23). Hence, the results indicated that the participants' mean temporal estimations were modulated by the nature of the Agent (Self, Other) only during the interactions with the servomotor, on the incongruent trials.

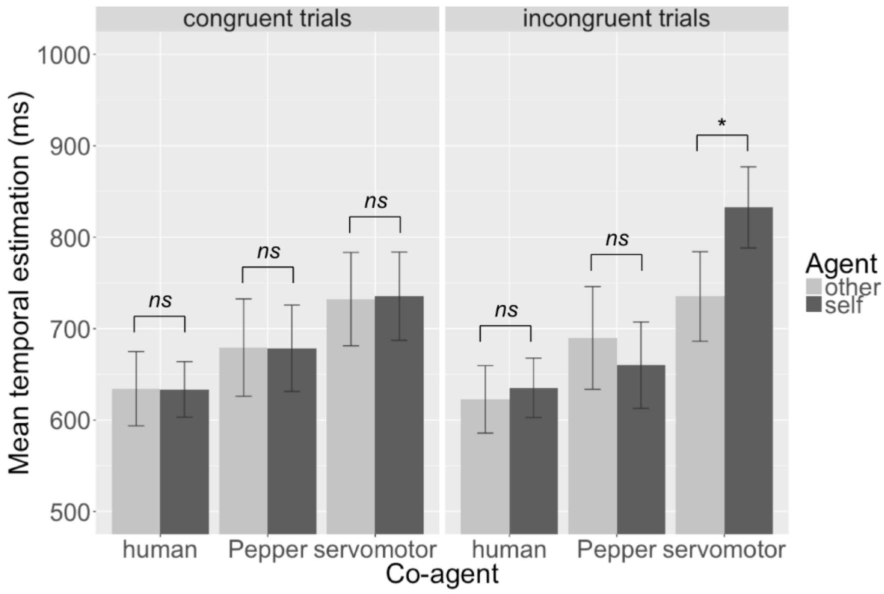

Figure 6. The Congruency x Agent x Co-agent interaction. Error bars represent standard errors. All tests were two-tailed. * corresponds to a p value < .05.

5 Discussion

The aim of this study was to investigate (1) the effect of the robot's physical appearance on the individuals' sense of *self-agency* and *vicarious agency*, and (2) to explore the cognitive mechanisms underlying the sense of shared agency when individuals they are engaged in a joint task with a machine. Participants were requested to perform a joint Simon task with a co-agent that could be either another human, or a machine. The machine appearance was manipulated so that it could be a full humanoid robot or a servomotor. Every accurate target detection (self- or other- generated) triggered an auditory tone after a certain delay. The participants had to report the perceived delay between the onset of the target detection and the onset of the auditory tone, which served as an implicit measure of the participants' sense of agency (Haggard, et al., 2002).

With regards to the effect of the impact of the robot's physical appearance on the individuals' experience of agency, our results revealed that overall, the participants reported shorter mean action-tone temporal intervals during the joint task with the other human compared to the humanoid robot, and shorter temporal estimations during the joint task with the humanoid robot compared to the servomotor. This finding suggested an increased overall experience of agency during the joint task with the humanoid robot relative to the servomotor. Furthermore, our experiment provided additional evidence bearing on a debated issue in the literature, namely, whether one's sense of agency is specific for one's own action or not. In the current research, we demonstrated that the experienced sense of agency was not specific to one's own actions, and that a form of vicarious agency was possible during a joint task. Importantly, our findings also revealed that the humanoid appearance of a machine could impact the development of the individuals' sense of *shared agency* during human-machine interactions. Indeed, participants' overall experience of agency was at its maximum during the human-human interactions and sharply declined during the human-servomotor interactions while an intermediate level was found during the interactions with the humanoid robot. At the same time, we found that the participants' mean temporal estimations for the self-generated actions were not different to the mean temporal estimations for other-generated actions during the joint actions with the other human and with the humanoid robot, suggesting the emergence of a sense of *shared agency* in the both cases. As matters stand, it is difficult to explain why distinct experiences of agency were found for the self-generated actions and the other-generated actions during the joint task with the human co-agent in Sahaï and colleagues (2019)' study whereas no differences were observed in the present study using a similar paradigm. Interestingly, distinct temporal estimations were found for the self-generated actions and for the other-generated actions during the joint task when the co-agent was the servomotor on incongruent trials, meaning that no shared experience of agency (or *shared agency*) occurred with this type of machine when the task difficulty increased. Seemingly, it could be speculated that the similarity with the humanoid robot led participants to treat the machine as a potential social partner (Fogg, 2003), echoing Searle (1983)'s contention that recognizing the other as similar to oneself and as a potential agent is a prerequisite to engaging in a collaborative activity (Searle, 1983). This ability to search for social boundaries has been demonstrated to be present very early in life, which made individuals profoundly social entities (Ciaunica, Constant, Preissl, & Fotopoulou, 2021; Fotopoulou & Tsakiris, 2017). Because human-like robots are known to elicit empathic behaviors in humans as opposed to non-human-like robots (Riek, Rabinowitch, Chakrabarti, & Robinson, 2009; Kwak, Kim, Kim, Shin, & Cho, 2013; Slater, Antley, Davison, Swapp, Guger, et al., 2006; but see also "uncanny valley" phenomenon, Misselhorn, 2009 and empathic behavior with minimal humanity cues, Vaes, Meconi, Sessa, & Olechowski, 2016), it is conceivable to think that participants were more likely to create similarity boundaries with the humanoid robot compared to the servomotor (De Vignemont & Singer, 2006). Indeed, a linear relation has been observed between the degree of anthropomorphism of robots and the activation of brain areas involved in the processing of others' minds (Krach, et al., 2008).

Consistently, the participants' RTs were also modulated by the human-like features of the co-agent. Indeed, we found a Social Simon Effect (SSE) with longer RTs on the incongruent trials compared to the congruent trials only when participants performed the joint Simon task with another human and with the humanoid robot. By contrast, this effect disappeared during the joint task with the non-human like machine (i.e., the servomotor). Supporting this, previous studies using a joint Simon paradigm have shown no SSE on participants' RTs when they partnered with non-biological agents (Sahaï, et al., 2019; Tsai, Kuo, Hung & Tzeng, 2008; Tsai, Knoblich, & Sebanz, 2011). Yet, some studies nevertheless observed a SSE when sharing a Simon task with a non-human agent (Dolk et al., 2011; Dolk, Hommel, Prinz, & Liepelt, 2013; Puffe, Dittrich, & Klauer, 2017; Stenzel & Liepelt, 2016). A possible reason for explaining such a discrepancy may relate to the agents' belief regarding the partner. Indeed, previous work emphasized that agents' beliefs on the origin of the robotic behavior could influence the outcome on a variety of behavioral and neuroimaging tasks (Hortensius & Cross, 2018; Stenzel, et al., 2012; Wykowska, Chaminade, & Cheng, 2016). For example, Stenzel and collaborators showed that the SSE reappeared when an intentional stance towards the machine was encouraged, that is, when the robot was described as an active and intelligent agent (Stenzel, et al., 2012), and suggest that ascribing agency to the co-actor (i.e., perceiving the co-actor as being the initiator of the action effect) is critical to observe the SSE (Stenzel, Dolk, Colzato, Sellaro, Hommel, & Liepelt, 2014). In the current study, even if it was not explicitly pointing out during task instructions, it is possible that having the participants interact with agents of a different nature had unconsciously led them to focus on the intentional aspect of the agents. Interestingly, the participants' RTs revealed that the SSE did not differ in amplitude when the participants performed the task with another human and with the humanoid robot. This suggested that the biological nature of the co-agent *per se* was not what influenced the SSE, but rather the ability to consider the co-agent as a social partner as it is the case with robots such as Pepper.

Finally, our findings support the existence of differences in the processing of the sense of *self-agency* and the processing of the sense of *shared agency* during a joint task. On the one hand, it has been proposed that the individuals' sense of *self-agency* was informed by the dynamic integration of both internal motor cues and contextual cues, with typically more weight given to the motor cues (Moore & Fletcher, 2012; Synofzik, et al., 2013). In addition, some authors pointed out that when action selection was easy (e.g., on congruent trials), the participants' sense of *self-agency* was stronger compared to a conflictual action selection context (e.g., on incongruent trials) (Sidarus & Haggard, 2016). On the other hand, in the current study, no Congruency effect was observed on the participants' mean action-tone temporal estimations neither for the self- or the other-generated actions when participants partnered with another human or with the humanoid robot, that is to say, when a sense of *shared agency* was experienced. By contrast, the Congruency did have an effect on the participants' mean action-tone temporal estimations for the self-generated actions when the sense of *shared agency* was not present anymore, that is to say, when the participants partnered with the servomotor. Taken together, this suggested that the fluency of action selection had a weaker role in the construction of the sense of *shared agency* than in the sense of *self-agency*. Consequently, it could be tentatively suggested that the weight of the egocentric internal cues linked to decision fluency was weakened when the individuals involved in the joint task were not considered as separate entities but holistically within a shared "we-identity". Unlike individual actions, during a joint task, the modulations in the individuals' sense of agency may be prominently dependent on contextual cues, even if internal motor cues are available. This corroborates previous investigations that showed that the experience of agency exhibited during joint actions, which were performed simultaneously by the agents involved in the task, was not based on egocentric predictions but depended on the degree of control exhibited by the whole team (Dewey, et al., 2014; van der Wel, Sebanz, & Knoblich, 2015). Hence, the current study provided additional evidence in the context of a joint task where the actions were performed alternately by the two agents involved in the shared task. Lastly, the question of whether similar results would have been observed with an explicit measurement of one's sense of agency could be raised. However, explicit measurement of the individuals' sense of agency have been shown to be mostly influenced by contextual cues such as prior thoughts for example (Synofzik, Vosgerau, & Newen, 2008). Yet, in the current study, the author of the generated key press was clearly identified given the color of the target so that the participants' self-reported explicit judgment of agency would not have differed from the instructions induced by the joint Simon task.

Eventually, some limitations of our work must nonetheless be acknowledged. Firstly, note should be taken that during the experiment, the co-agent's effector was not hidden by a physical separation. Hence, even though participants were given the explicit instruction to look at the screen and not at the actions performed by the co-agent, peripheral vision might have allowed them to discern their co-agent's actions. Consequently, it was not possible in the current research to distinguish between the contribution of the low-level processing of the social visual cues and the contribution of higher-order socio-cognitive processes. Secondly, because our experiment did not include a baseline condition (e.g., an experimental condition wherein the participants would have to estimate the action-tone temporal delays triggered by a computer program), it is difficult to know whether the lower level of agency found in the joint task with the servomotor was a floor effect or already an increment of agency. Yet, recent findings emphasized variations in the sense of *vicarious agency* exhibited by individuals during a joint task with a non-human-like robot, according to the level of embodiment of the machine-generated action (Roselli, Ciardo, & Wykowska, 2021). More specifically, it has been shown that when participants were performing a joint task with a robot that performed a physically perceivable action (e.g., executing a key press with the help of a limb, which triggered an auditory tone), they were able to experience a sense of *vicarious agency* over the robot's generated outcome. Conversely, when the robot's action was known but unperceivable and digitized (e.g., sending a Bluetooth command made noticeable with a visual signal, which triggered an auditory tone), the participants did not demonstrate a sense of *vicarious agency* for the robot's generated outcome anymore. Hence, it could be possible that the lower level of agency observed during the joint task with the servomotor in the current study did not imply a total absence of agency. Nonetheless, it could be said that the similarity of the humanoid robot appearance with the participants seemed to boost the participants' sense of *shared agency* during the joint task, in comparison to the non-anthropomorphized machine. Finally, another limitation of the study was that the robot co-agents did not have the same physical size. Indeed, while the humanoid robot was the size of a child, the servomotor was only about 10 centimeters tall. Hence, it is difficult to evaluate the part of the exogenous salience of the co-agent related to its size, and that related to the human-like embodiment in our results on the representation of other-generated actions and on the sense of agency.

In conclusion, the findings of this research showed that automation technology design could significantly change the individuals' agentive experience. Remarkably, human-like machines helped to mitigate the reported negative aspects induced by traditional automated systems in the individuals' experience of agency (Obhi & Hall, 2011b; Sahaï, et al., 2019). Indeed, the participants' sense of agency was reinforced during the joint task with the humanoid robot compared to the traditional machine, leading even to the construction of a sense of *shared agency* during the interaction with the human-like automata. Importantly, it must be said that the experience of agency is highly flexible and other factors could also influence how individuals develop a sense of *shared agency* with a robot, such as the duration of collaboration, the participants' intentional stance toward the robots (Barlas, 2019; Ciardo, Beyer, De Tommaso, & Wykowska, 2020), and the robot behavior predictability (Bolt & Loehr, 2017). Considering both the impact of the individuals' sense of agency on their capacity to engage in cooperative joint tasks (Babcock & Loewenstein, 1997; Caruso, et al., 2006) and the inexorable drive toward more automation, such findings must be taken into consideration for the successful design of new automated systems.

6 Funding

The authors disclosed receipt of the following financial support for the research, authorship, and publication of this article: AS and BB were supported by a French National Research Agency grant [grant number ANR-15-CE26-0010-01]. AdB. was supported by the Research Foundation – Flanders (FWO, Belgium). EAC was supported by a Marie Skłodowska-Curie grant [Agent No 743685]. EP was supported by French National Research Agency grants [grant numbers ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL, and ANR-16-CE33-0017-02].

7 Declaration of conflicting interests

The authors declare that there is no conflict of interest.

8 References

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic perspectives*, *11*(1), 109-126. DOI: 10.1257/jep.11.1.109

Barlas, Z. (2019). When robots tell you what to do: Sense of agency in human-and robot-guided actions. *Consciousness and Cognition*, 75, DOI: 10.1016/j.concog.2019.102819

Barlas, Z., Hockley, W. E., & Obhi, S. S. (2018). Effects of free choice and outcome valence on the sense of agency: evidence from measures of intentional binding and feelings of control. *Experimental brain research*, *236*(1), 129-139. DOI: 10.1007/s00221-017-5112-3

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300. DOI: 10.1111/j.2517-6161.1995.tb02031.x

Blakemore, S. J., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia*, *43*(2), 260-267. DOI: 10.1016/j.neuropsychologia.2004.11.012

Bolt, N. K., & Loehr, J. D. (2017). The predictability of a partner's actions modulates the sense of joint agency. *Cognition*, 161, 60-65. DOI: 10.1016/j.cognition.2017.01.004

Bolt, N. K., Poncelet, E. M., Schultz, B. G., & Loehr, J. D. (2016). Mutual coordination strengthens the sense of joint agency in cooperative joint action. *Consciousness and Cognition*, 46, 173-187. DOI: 10.1016/j.concog.2016.10.001

Breazeal, C., Takanishi, A., & Kobayashi, T. (2008). Social robots that interact with people. *Springer handbook of robotics*, 1349-1369. DOI: 10.1007/978-3-540-30301-5_59

Caruso, E. M., Epley, N., & Bazerman, M. H. (2006). The costs and benefits of undoing egocentric responsibility assessments in groups. *Journal of personality and social psychology*, *91*(5), 857. DOI: 10.1037/0022-3514.91.5.857

Ciardo, F., Beyer, F., De Tommaso, D., & Wykowska, A. (2020). Attribution of intentional agency towards robots reduces one's own sense of agency. *Cognition*, 194, 104109. DOI: 10.1016/j.cognition.2019.104109

Ciaunica, A., Constant, A., Preissl, H., & Fotopoulou, K. (2021). The first prior: from co-embodiment to co-homeostasis in early life. *Consciousness and cognition*, 91, 103117. DOI: 10.1016/j.concog.2021.103117

Crivelli, D., & Balconi, M. (2010). Agency and inter-agency, action and joint action: theoretical and neuropsychological evidence. In *Neuropsychology of the Sense of Agency* (pp. 107-122). Springer, Milano. DOI: 10.1007/978-88-470-1587-6_6

De Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why?. *Trends in cognitive sciences*, *10*(10), 435-441. DOI: 10.1016/j.tics.2006.08.008

Dewey, J. A., Pacherie, E., & Knoblich, G. (2014). The phenomenology of controlling a moving object with another person. *Cognition*, *132*(3), 383-397. DOI: 10.1016/j.cognition.2014.05.002

Dolk, T., Hommel, B., Colzato, L. S., Schütz-Bosbach, S., Prinz, W., & Liepelt, R. (2011). How "social" is the social Simon effect?. *Frontiers in Psychology*, 2, 84. DOI: 10.3389/fpsyg.2011.00084

Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2013). The (not so) social Simon effect: a referential coding account. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1248. DOI:10.1037/a0031031

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. DOI: 10.3758/bf03193146

Fogg, B.J. (2003) Computers as persuasive social actors. *Persuasive Technology*, pp. 89–120. DOI: 10.1016/b978-155860643-2/50007-x.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, *42*(3-4), 143-166. DOI: 10.1016/S0921-8890(02)00372-X

Fotopoulou, A., & Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference. *Neuropsychoanalysis*, *19*(1), 3-28. DOI: 10.1080/15294145.2017.1294031

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London*. *Series B: Biological Sciences*, *355*(1404), 1771-1788. DOI: 10.1098/rstb.2000.0734

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, *4*(1), 14-21. DOI: 10.1016/S1364-6613(99)01417-5

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage*, *16*(3), 814-821. DOI: 10.1006/nimg.2002.1117

Glasauer, S., Huber, M., Basili, P., Knoll, A., & Brandt, T. (2010, September). Interacting in time and space: Investigating human-human and human-robot joint action. In *19th International Symposium in Robot and Human Interactive Communication* (pp. 252-257). IEEE. DOI: 10.1109/ROMAN.2010.5598638

Grynszpan, O., Sahaï, A., Hamidi, N., Pacherie, E., Berberian, B., Roche, L., & Saint-Bauzel, L. (2019). The sense of agency in human-human vs human-robot joint action. *Consciousness and Cognition*, 75. DOI: 10.1016/j.concog.2019.102820

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature neuroscience*, *5*(4), 382. DOI: 10.1038/nn827

Hortensius, R. & Cross, E. S. (2018) From automata to animate beings: the scope and limits of attributing socialness to artificial agents. *Annals of the New York Academy of Sciences*, *1426*(1), pp. 93-110. DOI: 10.1111/nyas.13727

Hu, Y., Pan, Y., Shi, X., Cai, Q., Li, X., & Cheng, X. (2018). Inter-brain synchrony and cooperation context in interactive decision making. *Biological psychology*, 133, 54-62. DOI: 10.1016/j.biopsycho.2017.12.005

Jenkins, M., Esemezie, O., Lee, V., Mensingh, M., Nagales, K., & Obhi, S. S. (2021). An investigation of "We" agency in co-operative joint actions. *Psychological Research*, 1-15. DOI: 10.1007/s00426-020-01462-6

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166. DOI: 10.1007/s10339-007-0170-2

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PloS one*, *3*(7), DOI: 10.1371/journal.pone.0002597

Kwak, S. S., Kim, Y., Kim, E., Shin, C., & Cho, K. (2013, August). What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In *2013 IEEE RO-MAN* (pp. 180-185). IEEE. DOI: 10.1109/ROMAN.2013.6628441

Le Bars, S., Devaux, A., Nevidal, T., Chambon, V., & Pacherie, E. (2020). Agents' pivotality and reward fairness modulate sense of agency in cooperative joint action. *Cognition*, 195, 104117. DOI: 10.1016/j.cognition.2019.104117

Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. *Minds and Machines*, 19(3), 345. DOI: 10.1007/s11023-009-9158-2

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: a review of cue integration approaches. *Consciousness and Cognition*, *21*(1), 59-68. DOI: 10.1016/j.concog.2011.08.010

Obhi, S. S., & Hall, P. (2011a). Sense of agency and intentional binding in joint action. *Experimental brain research*, *211*(3-4), 655. DOI: 10.1007/s00221-011-2675-2

Obhi, S. S., & Hall, P. (2011b). Sense of agency in joint action: Influence of human and computer co-actors. *Experimental brain research*, *211*(3-4), 663-670. DOI: 10.1007/s00221-011-2662-7

Pacherie, E. (2007). The sense of control and the sense of agency. *Psyche*, *13*(1), 1-30. ijn_00352565

Pacherie, E. (2012). The Phenomenology of Joint Action: Self-Agency vs. Joint-Agency. In Axel Seemann (ed.), *Joint Attention: New Developments*, Cambridge MA: MIT Press, pp. 343-389.

Picard, F., & Friston, K. (2014). Predictions, perception, and a sense of self. *Neurology*, *83*(12), 1112-1118. DOI: 10.1212/WNL.0000000000000798

Puffe, L., Dittrich, K., & Klauer, K. C. (2017). The influence of the Japanese waving cat on the joint spatial compatibility effect: a replication and extension of Dolk, Hommel, Prinz, and Liepelt (2013). *PloS one*, *12*(9), e0184844.

Renes, R. A., van Haren, N. E., Aarts, H., & Vink, M. (2015). An exploratory fMRI study into inferences of self-agency. *Social cognitive and affective neuroscience*, *10*(5), 708-712. DOI: 10.1093/scan/nsu106

Riek, L. D., Rabinowitch, T. C., Chakrabarti, B., & Robinson, P. (2009, March). How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction* (pp. 245-246). ACM. DOI: 10.1145/1514095.1514158

Roselli, C., Ciardo, F., & Wykowska, A. (2021). Intentions with actions: The role of intentionality attribution on the vicarious sense of agency in Human–Robot interaction. Quarterly Journal of Experimental Psychology, 17470218211042003. DOI: 10.1177/17470218211042003

Rubichi, S., & Nicoleti, R. (2006). The Simon effect and handedness: Evidence for a dominant-hand attentional bias in spatial coding. *Perception & Psychophysics*, *68*(7), 1059-1069. DOI: 10.3758/bf03193709

Sahaï, A., Desantis, A., Grynszpan, O., Pacherie, E., & Berberian, B. (2019). Action co-representation and the sense of agency during a joint Simon task: Comparing human and machine co-agents. *Consciousness and Cognition*, 67, 44-55. DOI: 10.1016/j.concog.2018.11.008

Sahaï, A., Pacherie, E., Grynszpan, O., & Berberian, B. (2017). Predictive mechanisms are not involved the same way during human-human vs. human-machine interactions: a review. *Frontiers in neurorobotics*, 11, 52. DOI: 10.3389/fnbot.2017.00052

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, *94*(3), 241-255. DOI: 10.1016/j.cognition.2004.04.003

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social cognitive and affective neuroscience*, *7*(4), 413-422. DOI: 10.1093/scan/nsr025

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and brain sciences*, *36*(4), 393-414. DOI: 10.1017/S0140525X12000660

Searle, J. R. (1983). The word turned upside down. *The New York Review of Books*, 30(16), 74-79.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, *10*(2), 70-76. DOI:10.1016/j.tics.2005.12.009

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353-367. DOI: 10.1111/j.1756-8765.2009.01024.x

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, *88*(3), B11-B21. DOI: 10.1016/s0010-0277(03)00043-x

Seibold, J. C., Chen, J., & Proctor, R. W. (2016). Exploring handedness asymmetries in the Simon effect. *Attention, Perception, & Psychophysics*, *78*(2), 437-451. DOI:10.3758/s13414-015-1037-8

Shiraishi, M., & Shimada, S. (2021). Inter-brain synchronization during a cooperative task reflects the sense of joint agency. *Neuropsychologia*, *154*, 107770. DOI: 10.1016/j.neuropsychologia.2021.107770

Sidarus, N., & Haggard, P. (2016). Difficult action decisions reduce the sense of agency: A study using the Eriksen flanker task. *Acta Psychologica*, 166, 1-11. DOI: 10.1016/j.actpsy.2016.03.003

Silver, C. A., Tatler, B. W., Chakravarthi, R., & Timmermans, B. (2020). Social Agency as a continuum. *Psychonomic Bulletin & Review*, 1-20. DOI: 10.3758/s13423-020-01845-1

Simon, J. R., & Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics*, *6*(1), 99-105. DOI: 10.1080/00140136308930679

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., ... & Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PloS one*, *1*(1), DOI: 10.1371/journal.pone.0000039

Stenzel, A., Chinellato, E., Bou, M. A. T., del Pobil, Á. P., Lappe, M., & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1073, DOI: 10.1037/a0029493

Stenzel, A., Dolk, T., Colzato, L. S., Sellaro, R., Hommel, B., & Liepelt, R. (2014). The joint Simon effect depends on perceived agency, but not intentionality, of the alternative action. *Frontiers in Human Neuroscience*, 8, 595. DOI: 10.3389/fnhum.2014.00595

Stenzel, A., & Liepelt, R. (2016). Joint Simon effects for non-human co-actors. Attention, *Perception, & Psychophysics*, *78*(1), 143-158. DOI: 10.3758/s13414-015-0994-2

Strother, L., House, K. A., & Obhi, S. S. (2010). Subjective agency and awareness of shared actions. *Consciousness and Cognition*, *19*(1), 12-20. DOI: 10.1016/j.concog.2009.12.007

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and cognition*, *17*(1), 219-239. DOI: 10.1016/j.concog.2007.03.010

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: an interplay between prediction and postdiction. *Frontiers in psychology*, 4, 127. DOI: 10.3389/fpsyg.2013.00127

Takahashi, H., Terada, K., Morita, T., Suzuki, S., Haji, T., Kozima, H., ... & Naito, E. (2014). Different impressions of other agents obtained through social interaction uniquely modulate dorsal and ventral pathway activities in the social human brain. *Cortex*, 58, 289-300. DOI: 10.1016/j.cortex.2014.03.011

Tsai, C. C., Kuo, W. J., Hung, D. L., & Tzeng, O. J. (2008). Action co-representation is tuned to other humans. *Journal of Cognitive Neuroscience*, *20*(11), 2015-2024. DOI: 10.1162/jocn.2008.20144

Tsai, J. C. C., Knoblich, G., & Sebanz, N. (2011). On the inclusion of externally controlled actions in action planning. *Journal of experimental psychology: human perception and performance*, *37*(5), 1407. DOI: 10.1037/a0024672

Vaes, J., Meconi, F., Sessa, P., & Olechowski, M. (2016). Minimal humanity cues induce neural empathic reactions towards non-human entities. *Neuropsychologia*, 89, 132-140. DOI: 10.1016/j.neuropsychologia.2016.06.004

van der Wel, R. P., Sebanz, N., & Knoblich, G. (2012). The sense of agency during skill learning in individuals and dyads. *Consciousness and Cognition*, 21(3), 1267-1279. DOI: 10.1016/j.concog.2012.04.001

van der Wel, R. P., Sebanz, N., & Knoblich, G. (2015). A joint action perspective on embodiment. *Conceptual and interactive embodiment: Foundations of embodied cognition*, 2, 165.

Wang, Y., & Quadflieg, S. (2015). In our own image? Emotional and neural processing differences when observing human–human vs human–robot interactions. *Social cognitive and affective neuroscience*, *10*(11), 1515-1524. DOI: 10.1093/scan/nsv043

Wen, W., & Haggard, P. (2020). Prediction error and regularity detection underlie two dissociable mechanisms for computing the sense of agency. *Cognition*, 195, 104074. DOI: 10.1016/j.cognition.2019.104074

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Frontiers in psychology*, 8, 1663. DOI: 10.3389/fpsyg.2017.01663

Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), DOI: 10.1098/rstb.2015.0375

Wylie, S. A., Ridderinkhof, K. R., Elias, W. J., Frysinger, R. C., Bashore, T. R., Downs, K. E., ... & van den Wildenberg, W. P. (2010). Subthalamic nucleus stimulation influences expression and suppression of impulsive behaviour in Parkinson's disease. *Brain*, *133*(12), 3611-3624. DOI: 10.1093/brain/awq239