Cognition 194 (2020) 104109

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Attribution of intentional agency towards robots reduces one's own sense of agency

Francesca Ciardo[a,*], Frederike Beyer[b], Davide De Tommaso[a], Agnieszka Wykowska[a]

[a] Italian Institute of Technology, Genoa, Italy
[b] Queen Mary University, London, UK

ARTICLE INFO

Keywords:
Human-robot interaction
Sense of agency
Diffusion of responsibility
Intentional agency

ABSTRACT

In the presence of others, sense of agency (SoA), i.e. the perceived relationship between our own actions and external events, is reduced. The present study aimed at investigating whether the phenomenon of reduced SoA is observed in human-robot interaction, similarly to human-human interaction. To this end, we tested SoA when people interacted with a robot (Experiment 1), with a passive, non-agentic air pump (Experiment 2), or when they interacted with both a robot and a human being (Experiment 3). Participants were asked to rate the perceived control they felt on the outcome of their action while performing a diffusion of responsibility task. Results showed that the intentional agency attributed to the artificial entity differently affect the performance and the perceived SoA on the outcome of the task. Experiment 1 showed that, when participants successfully performed an action, they rated SoA over the outcome as lower in trials in which the robot was also able to act (but did not), compared to when they were performing the task alone. However, this did not occur in Experiment 2, where the artificial entity was an air pump, which had the same influence on the task as the robot, but in a passive manner and thus lacked intentional agency. Results of Experiment 3 showed that SoA was reduced similarly for the human and robot agents, threby indicating that attribution of intentional agency plays a crucial role in reduction of SoA. Together, our results suggest that interacting with robotic agents affects SoA, similarly to interacting with other humans, but differently from interacting with non-agentic mechanical devices. This has important implications for the applied of social robotics, where a subjective decrease in SoA could have negative consequences, such as in robot-assisted care in hospitals.

1. Introduction

In everyday life, humans interact with various and often complex systems. We interact not only with natural social agents (humans) but also with artificial and computerized systems like the Google assistant. In the near future, robots will become a completely new category of social beings among us. As *artefacts*, they are only mechanical and programmable entities. However, through their embodiment[1] they become artificial *agents*, as they can have an impact on the surrounding environment through their capability to move around and perform actions on objects. As agents, robots receive a completely different status in our psychological niche. This is because attribution of intentional agency to others has an impact on how we interact with them. For example, in interaction with other agents, we try to predict and explain their action goals and what they are planning to do next. This is crucial for our efficient functioning in natural environment. Past research has shown that attribution of intentional agency to others influenced fundamental mechanisms of cognition, such as attention, spatial coding and perspective taking (Wiese, Wykowska, Zwickel, & Müller, 2012; Wykowska, Wiese, Prosser, & Müller, 2014; Stenzel et al., 2012; Zwickel, 2009; Ward, Ganis, & Bach, 2019). An open question is whether the attribution of intentional agency to others also affects *our own* sense of agency, while interacting with them.

1.1. Attribution of intentional agency and other processes of cognition

Wiese et al. (2012) investigated the impact of attributed intentional agency on gaze-mediated orienting of attention. The authors found that orienting of attention (measured as gaze cueing effects) was engaged to a larger extent when participants believed that the agent they were observing was controlled by a human, as compared to the belief that it merely represented a computer program. In Wykowska et al. (2014),

* Corresponding author at: Istituto Italiano di Tecnologia, Via E. Melen, 83, 16152 Genova, Italy.
E-mail address: francesca.ciardo@iit.it (F. Ciardo).

[1] It should be noted that when using the term *embodiment*, we refer at the physical presence of artificial intelligent agent in the environment (Duffy & Joue, 2000).

https://doi.org/10.1016/j.cognition.2019.104109
Received 18 March 2019; Received in revised form 10 October 2019; Accepted 11 October 2019
Available online 29 October 2019

the authors investigated the electrophysiological correlates of such differential effect. Results showed that early attention-related ERP components of the EEG signal were observed only when participants believed that the agent was controlled by a human, but not when they believed the agent's behaviour was pre programed. This was independent of the agent's identity (human or robot face). Similarly, Stenzel et al. (2012) tested whether the believed humanness of the robot modulates stimulus-response compatibility in a shared go/no-go Simon task (e.g., Sebanz, Knoblich, & Prinz, 2003; Ciardo & Wykowska, 2018). Results showed that the Simon effect emerged only in the human-like but not in the machine-like condition.

Another cognitive mechanism that has been investigated as a function of the attributed intentional agency is visual perspective taking (VPT), i.e. the ability to consider what another agent can see and how the scene looks from a different point of view. Zwickel (2009) tested whether processes that evoke agency interpretations and mental state attributions also lead to the adoption of the actor's visuospatial perspective by the observer. Zwickel used film clips in which actors were represented by triangles (the Heider & Simmel, 1944) and showed that allocentric perspective taking increased as a function of the likelihood of agentic interpretations. Specifically, for video clips in which the movement of the triangles was characterized by agency, i.e. goal-directed and theory-of-mind behaviours, participants showed longer reaction times when their perspective and those of the triangles differed, compared to when they did not (Zwickel, 2009). In a more recent study, Ward, Gianis, and Bach (2019), by using a mental rotation task, compared the VPT effect elicited by an intentional agent (a human) and non-agentic entities (an object). Results showed slower reaction times for letters oriented away from rather than toward participants (i.e. the toward/away bias) when a human was present in the visual scene. However, the bias decreased when the human agent was replaced with an object (i.e. a lamp), even when the object had the same directionality and faced the items as the persons did. Together, the aforementioned evidence highlights the pivotal role played by attribution of intentional agency in cognition.

Neuroimaging studies (Chaminade et al., 2012; Gallagher, Jack, Roepstorff, & Frith, 2002; Krach et al., 2008; Takahashi et al., 2014; Wang & Quadflieg, 2015) showed that, under certain constraints (i.e., human appearance and motor kinematics), activation of social brain areas involved in lower-level social cognitive processing, like action understanding (i.e., the anterior intraparietal sulcus, aIPS) is comparable for human-robot and human-human interaction. Although brain areas constituting the mentalizing network were found to respond to both human-human and human-robot interactions, systematic variations in the activity across sites (i.e., medial prefrontal cortex, MPFC; temporal–parietal junction, TPJ; and insula) suggest that high-level social cognitive processes depend on mind perception. Specifically, Wang and Quadflieg (2015) showed that increased activity in the left TPJ indicative of situation-specific mental state attributions occurred for human-human interaction, whereas the observation of human–robot recruited areas associated with script-based social reasoning, as the precuneus and the ventromedial prefrontal cortex (VMPFC). Thus, it has been proposed that humans are able to automatically perceive robots as agents, i.e. entities that are able to plan and act. However, the representation of robots' actions might sometimes lack attribution of intentional agency, that is, robots might sometimes be perceived as agents whose actions are not guided by a specific intent (Gray, Gray, & Wegner, 2007; but see Marchesi et al., 2019 for evidence showing that humans can use mental states to explain robots' behaviour).

1.2. Sense of agency

Sense of agency constitutes a crucial aspect of human cognition. The sense of agency (SoA) describes the feeling that one is in control over one's actions and their consequences. Importantly, it has been shown that the human SoA is affected by social action contexts, that is, by the presence of other potential intentional agents during a task (Bandura, 1991; but see also Khalighinejad, Bahrami, Caspar, & Haggard, 2016). In two recent studies, Beyer, Sidarus, Bonicalzi, and Haggard, 2017; Beyer, Sidarus, Fleming, & Haggard, 2018) showed that SoA is reduced when participants believe they are playing with another human, even if they are actually playing with a computer (Beyer et al., 2017; Beyer et al., 2018). In Beyer et al.s' work, the reduction in agency ratings was also mirrored at the electrophysiological level with a reduction of the feedback-related negativity amplitude related to outcome monitoring (Beyer et al., 2017). In a subsequent neuroimaging study, Beyer et al. (2018) found that the reduction of SoA in the social condition was associated with an increased activity of the precuneus, an area involved in mentalizing processes (Beyer et al., 2018). Importantly, these studies used immediate action feedback to avoid ambiguity of authorship: even when they knew that they had caused a given outcome, participants felt less in control over the consequences of their own actions, if they performed those actions in the presence of another potential agent. These effects are distinct from explicitly cooperative tasks which may have opposing effects on sense of agency, such as increased sense of agency during joint action (Obhi & Hall, 2011).

Based on these findings, Beyer et al. proposed a model of how social context influences SoA. Specifically, they argued, that in the presence of other potential agents, representing their potential actions interferes with action selection, increasing interference in our own action planning processes (Beyer et al., 2017, 2018). Increased interference (or action disfluency), in turn, has been shown to decrease SoA over action outcomes (Chambon, Sidarus, & Haggard, 2014). As such, the model predicts that SoA should be more strongly reduced in the presence of active intentional agents that elicits representation of their potential actions, than in the presence of a non-agentic and unintentional source of alternative task outcomes. The findings of Beyer et al. inspire an interesting question in terms of human-robot interaction: Does interaction with an embodied robot also induce a reduced SoA? According to Beyer et al.'s model (Beyer et al., 2017), and based on previous literature showing that robots are in fact perceived as agentic entities (Wang & Quadflieg, 2015), interaction with a robot should induce reduced SoA, if the robot is perceived as an intentional agent. This is because representation of the robot's potential actions should interfere with one's own action planning, similarly to interaction with another human. Along the same line of reasoning, a non-agentic entity should not induce such an effect.

1.3. Aim of the study

The present study investigated whether interacting with an embodied robot produces an effect of reduced SoA in a similar way as interactions with an alleged human agent, and whether it is indeed attribution of intentional agency that is critical for evoking this effect. In Experiment 1 and 2, we asked participants to perform a diffusion of responsibility task (Beyer et al., 2017, 2018) alone or within the presence of another artificial entity: either a robot (Experiment 1), or an air pump (Experiment 2) which can make changes in the environment, but in a non-agentic passive manner. In Experiment 3, we tested wether the nature of the agent, human or a robot, differetly affect SoA in a social context. Since the model of Beyer et al. (Beyer et al., 2017, 2018) postulates sub-mentalising as a process that interferes with action planning, and thereby reduces SoA, we reasoned that if we observe a reduction of SoA in the robot condition, similar to the human condition, and no such effect in the passive pump condition, this would indicate attribution of intentional agency to the robot. Therefore reduced SoA might serve as a marker of adoption of intentional stance towards artificial agents (Dennett, 1971, 1981).

SoA is a multi-faceted concept, with a resulting diversity in measures used to assess it (Wen, 2019). On the one hand, SoA is related to a feeling of authorship, which is commonly measured using intentional binding (Engbert, Wohlschläger, & Haggard, 2008). On the other hand,

SoA is understood as a sense of control over one's own actions and their consequences. This aspect of SoA is sensitive to manipulations of decision-making complexity, and is commonly assessed using explicit ratings of control (Sidarus, Vuorre, & Haggard, 2017). As the focus of the present experiments lies on decreased SoA in the absence of ambiguity of authorship, we used explicit ratings, in line with previous studies (Beyer, Sidarus et al., 2017; Beyer et al., 2018).

2. Experiment 1

Experiment 1 aimed at examining SoA when interacting with a robot. To this end, we asked participants to perform the diffusion of responsibility task developed by Beyer et al. (2017; 2018) while interacting with the Cozmo robot (Anki Robotics). Participants were asked to perform costly actions (i.e. losing various amounts of points) to stop an inflating balloon from bursting in Individual vs. Joint contexts. We hypothesized that if the robot is perceived as an intentional agent, then we should observe the effect of reduced SoA in Experiment 1. According to Beyer et al.,'s model (Beyer et al., 2017) this would be because of automatic representation of robot agent's potential actions, which might result in an increased action disfluency and a mentalising process that might interference at the action planning stage. On the contrary, if the robot is perceived as a mere artefact, i.e. a non-intentional agent, then no interference should occur, since participants wouldn't necessarily represent the events related to the robotic entity as action plans. Thus, in this latter case, perceived SoA should not vary across conditions. In contrast to previous studies finding socio-cognitive effects in human-robot interactions (Wiese et al., 2012; Wykowska et al., 2014. Stenzel et al., 2012), we did not tell participants explicitly that the robot was controlled by a human. Thus, any intentional agency attributed to the robot would be a result of a spontaneous process, rather than an explicitly induced belief.

2.1. Method and materials

2.1.1. Participants

Thirty participants (10 males; 5 left-handed; Mean age: 27.6 ± 7.5 years) took part in the study. All participants had normal or corrected-to-normal vision and were not informed with respect to the purpose of the experiment. Participants received a reimbursement of 10€ for their participation. All gave their written informed consent before participating. All experiments were conducted in accordance with the ethical standards laid down in the 2013 Declaration of Helsinki and were approved by the local ethical committee (Comitato Etico Regione Liguria). Sample size was defined according to previous experiments (Beyer et al., 2017, 2018), and by a priori power analysis indicating a sample N = 28 to detect a large effect size [Cohen's $F^2 = 0.35$, alpha (one-tailed) = .05 and power = 0.85].

2.1.2. Apparatus and stimuli

The experimental setup consisted of: 1) A Cozmo robot and two Cozmo cubes (91.125 cm^3), 2) a mobile Android device in which the standard Cozmo application with 'SDK enabled option', 3) a laptop connected with Cozmo through the Android Debug Bridge (adb) as described in (http://cozmosdk.anki.com/docs/adb.html), 4) a 21' inches screen (1920 × 1080) to display the task. The participants were seated facing Cozmo. The screen laid horizontally on the table between the participant and Cozmo. The cubes were located on both sides of the screen (Fig. 1). Stimuli consisted of pictures of a pin and a red balloon (113 × 135 pixels). Responses during the game were executed by tapping the assigned cube with the full hand. SoA ratings were collected using a standard computer mouse. Stimulus presentation, response timing, and data collection were controlled by Opensesame software (Mathôt, Schreij, & Theeuwes, 2012) version 3.2.4 for Windows. The Cozmo SDK was integrated in Opensesame, see Ciardo, De Tommaso, Beyer, & Wykowska (2018) for the procedure.

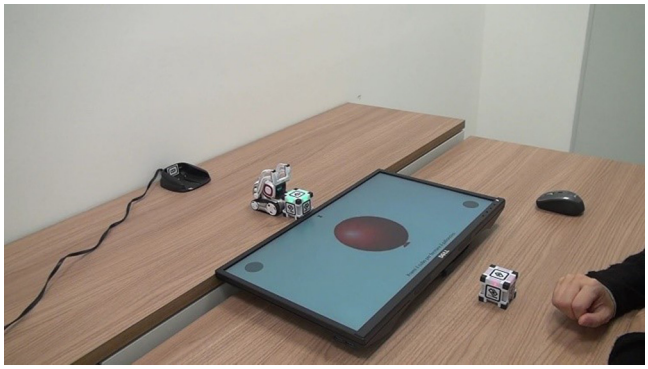

Fig. 1. Experimental setup of Experiment 1.

2.1.3. Procedure

The task consisted of a game in which participants had to stop inflation of a balloon before it would reach a pin and burst (see Beyer et al., 2018 for a similar task). They could stop the balloon by tapping a cube that was assigned to them. Each action resulted in the loss of a variable number of points, which depended on the size of the balloon. Participants were instructed that the later they stopped the balloon, the fewer points they would lose. However, they were told that if the balloon burst they would lose the maximum amount of points. As a result, the action (i.e. stopping the balloon) resulted to be costly, but less costly than not acting.

Participants were instructed that at the beginning of the game, they and Cozmo would receive 2500 points each, and in each trial, they and Cozmo could lose up to 100 of these points. Before the task began, participants were shown a fictional ranking indicating the best eight scores obtained in the task by other players and Cozmo. They were instructed to try and perform as the best player of all (fictional, unbeknownst to participants) players. They were also told that at every game Cozmo was trying to improve its score. The exact number of points lost could not be fully predicted from the size of the balloon at the moment it was stopped. Indeed, the inflating sequence was divided into four different payoff sections of equal length. Within each payoff section, the actual number of points lost was varied randomly from trial to trial.

The payoff structure of the task is reported in Table 1. The task consisted of 12 blocks of 10 trials each. Blocks were randomly assigned to either the 'Individual' or the 'Joint' condition, with the constraint that 6 blocks were performed per condition. A short practice session including 3 trials per condition was administered before starting the experiment. At the beginning of 'Individual' blocks, Cozmo moved away from its cube and entered into the sleep mode (see Video 1 in the Supplementary materials). Participants were instructed that in the Individual trials they were the only agent in charge to avoid the balloon bursting; if they did not act, the balloon would burst and they would lose the maximum amount of points. At the beginning of 'Joint' blocks, Cozmo woke up and took up its position close to the respective cube. Participants were instructed that, in these trials, both they, and Cozmo would be playing, and they could use their respective cube to stop the inflation of the balloon. If neither the participant nor Cozmo acted, the balloon would burst and both would lose the same number of points. If Cozmo stopped the inflation of the balloon, the participant would not lose any points. If the participant stopped the balloon, they would lose a number of points according to the size at which they stopped it, and Cozmo would not lose any points. Cozmo was programmed to act only in 60% of the Joint trials (i.e. 36 out of 60 Joint trials). In the Joint condition, Cozmo's tap was triggered when the 90% of the inflating sequence was run and no action was executed by the participant. To avoid the ambiguity of simultaneous tapping, two grey dots were presented on the upper-left and lower-right corner of the screen. When participants or Cozmo successfully stopped the balloon, the dots turned

Table 1
Payoff structure of the task for trials in which the balloon was stopped or burst. The table shows the payoff sections in which the inflating sequence was divided, the corresponding balloon size, and the amount of points lost for each payoff section. Within each payoff section, the actual number of points lost was varied randomly from trial to trial.

Payoff section	Stop Balloon size	Outcome
4	The balloon was stopped over the 50% of the maximum size.	15-1
3	Participant stopped the balloon at a size between the 49% and 33% of the maximum size.	29-16
2	Balloon stop size was between the 33% and the 17% of the maximum size.	45-31
1	Participant stopped the balloon at a size lower or equal to < the 17% of the maximum size.	60-46
0	The balloon burst.	100-80

from grey to blue, indicating who responded first: the participant (lower-right dot) or Cozmo (upper-left).

Each trial started with a frame presented for 1000 ms indicating the condition: Individual or Joint. At the same time, the cubes lit up indicating that they were ready to detect a response. If the condition was Individual, only the cube assigned to the participant lit up. Then a frame indicating the start of a new trial was presented for 1500 ms, followed by a fixation point for a random 800–1000 ms time. Subsequently, the balloon inflation sequence started. The balloon was presented at its starting size (i.e., 0.05 rate of the image size) for 500 ms before starting to inflate toward the pin. At any point of the balloon inflation sequence, participants could tap their cube to stop the balloon. If they did so, the balloon was displayed at its stop size for 1000 ms. If participants did not react in time, the balloon reached the pin and burst. When the balloon burst the word "Pop" was flashed for 1000 ms. Subsequently, a fixation dot was presented for a time random between 800 and 1000 ms. Afterwards, a feedback frame indicating how many points participants lost, i.e. the action outcome, was displayed for 2000 ms. Then an 8-point Likert scale with the question 'How much control did you feel over the outcome?' was presented. The endpoints of the scale labels were 1 ='No control' and 8 ='Full control'. Participants used the mouse to indicate how much control they felt they had over the number of points lost during that trial (i.e. the outcome) (Fig. 2).

In order to make it difficult to always stop the balloon near to the pin, the speed with which the balloon inflated varied from trial to trial. Also, at some point along the trial, the inflating sequence sped up, and this point varied from trial to trial.

2.2. Data analysis

To fully characterize the risk-taking behaviour in the task, we estimated the number of trials in which the balloon was stopped by the participants (Valid trials), the balloon burst (Missed trials), and in which Cozmo acted (Cozmo trials). Frequencies of Valid, Missed, and Cozmo trials were compared through paired sample t-tests. Then, we analyzed Valid and Missed trials separately. For Valid trials, i.e. when the participant acted and successfully stopped the balloon, we estimated for each participant the size at which the balloon was stopped (Stop size), the number of points lost in each trial (Outcome), and agency ratings. Stop size and Outcome were standardized for each participant. Behavioural data were analyzed using linear mixed-effects models. Stop size was modelled as a function of Condition (Individual, Joint). We modelled the Outcome of each trial as a function of the Condition (Individual, Joint) and Stop size. Agency ratings were modelled using Condition (Individual, Joint) and Outcome, plus their interactions. For Missed trials, we addressed if SoA ratings were predicted by Condition (Individual, Joint), the number of points lost (i.e. Outcome), or their interaction. Fixed effects were modelled as participant random effects (random intercepts and slopes). Analyses were conducted using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R. Parameter estimates (β) and their associated t-tests (t, p), calculated using the Satterthwaite approximation for degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2015) are presented to show the magnitude of the effects, with bootstrapped 95% confidence intervals (Efron & Tibshirani, 1994).

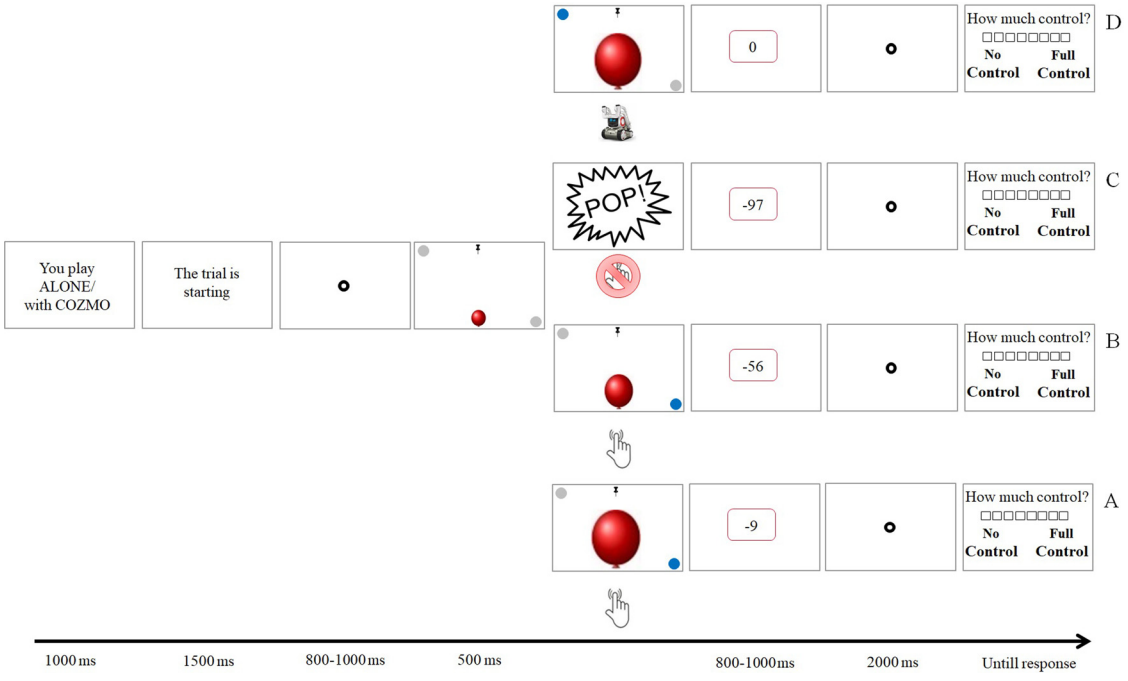

Fig. 2. Task procedure and outline of a high-risk trial (A), a low-risk trial (B), a missed trial (C), and a Cozmo trial (D).

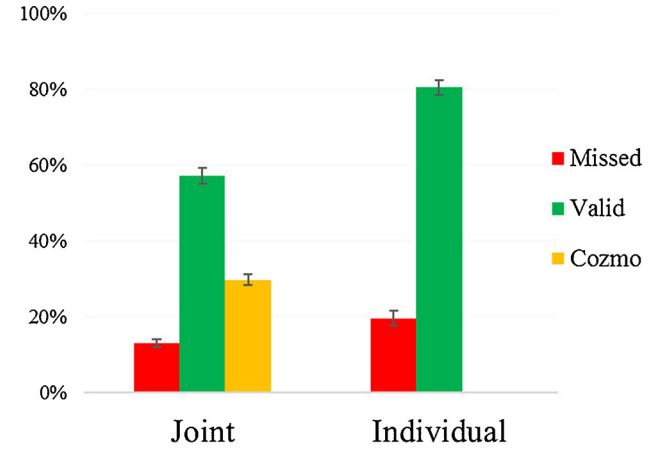

Fig. 3. Frequencies of responses for Experiment 1 plotted as function of Missed (red), Valid (green), and Cozmo trials (yellow) across Joint and Individual condition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

2.3. Results and discussion

The balloon burst significantly more frequently when participants performed the task alone than when playing with Cozmo (Fig. 3), as indicated by higher percentage of Missed trials in the Individual (M = 20.8%, SE = 1.4) than in the Joint condition (M = 14.3%, SE = 0.8) [t_{29} =5.89 $p < .001$, d = 1.08]. In the Joint condition, Cozmo acted more often than the balloon burst, as Missed trials were less frequent than Cozmo trials (M = 31.2%, SE = 1.3) [$t_{29} = 13.89$, $p < .001$, d = 2.54].

2.3.1. Valid Trials

2.3.1.1. Stop size. The social context did not predict the size at which the balloon was stopped [$\beta = 0.03$, $t_{27.47} < 1$, $p > .05$, 95% CI = (-0.06, 0.12)] (Individual: M = 3.26, SE = 0.02, Joint: M = 3.22, SE = 0.02).

2.3.1.2. Outcome. As intended by the task setup, the outcome was predicted by the size at which participants stopped the balloon [$\beta = -0.35$, $t_{31.03} = -5.12$, $p < .001$, 95% CI = (-0.48, -0.22)]. In valid trials, participants lost a significantly smaller amount of points in the Joint (M = 9.46, SE = 0.24) compared to the Individual condition (M = 10.14, SE = 0.21) [$\beta = 0.10$, $t_{186.88} = 2.69$ $p = .008$, 95% CI = (0.03, 0.17)].

2.3.1.3. Sense of agency (SoA) ratings. Results showed a significant reduction in agency ratings in the Joint (M = 6.23, SE = 0.05), compared to the Individual (M = 6.50, SE = 0.04) condition [$\beta = 0.29$, $t_{28.68} = 3.34$, $p = .002$, 95% CI = (0.12, 0.45)]. Agency ratings were also predicted by the Outcome [$\beta = -0.47$, $t_{28.99} = -6.40$, $p < .001$, 95% CI = (-0.61, -0.32)], with smaller losses being associated with higher SoA ratings. There was no significant interaction between condition and outcome (Fig. 4).

2.3.2. Missed trials

When the balloon burst, agency ratings were not predicted by Condition or Outcome (all ps > .15)

2.3.3. Discussion

Experiment 1 aimed at examining SoA in interaction with an embodied robot. To this end, we asked participants to perform the diffusion of responsibility task with the Cozmo robot (Anki Robotics). Results showed a lower percentage of missed trials in the Joint compared to the Individual condition. Moreover, in the Joint condition, the balloon was stopped by the robot more frequently than it burst. When participants successfully stopped the balloon, SoA was rated as lower in the Joint than in the Individual condition, independently of the number of lost points. This result suggests that interacting with a robot reduces SoA, similarly to the case of human-human interaction (Beyer et al., 2017, 2018). Moreover, in accordance with previous studies, SoA was reduced for more negative outcomes, confirming that participants followed the instructions and rated their perceived control over the outcome, rather than over the success of the trial. Importantly, as there was no Condition*Outcome interaction, and participants obtained better outcomes in the Joint condition, the reduced SoA in the Joint condition cannot be explained by condition effects on outcomes.

3. Experiment 2

Results of Experiment 1 showed reduced SoA in the Joint (Cozmo) condition as compared to Individual condition. This can be interpreted in line with Beyer et al.s' model, which suggests that representing a robot as an intentional agent may interfere with one's own action planning, thereby producing the reduced SoA effect. If this is true, then this effect should not occur for non-agentic entities. Thus, it is important to have a control experiment in which a mechanical non-agentic device has a comparable physical presence as the robot.

In previous experiments (Beyer et al., 2017, 2018), reduced SoA was only found for the presence of a human co-player (represented by an avatar picture), however, there are important differences between previous studies and the above Cozmo experiment. While in Beyer and colleagues' studies pictures on the computer screen indicated the experimental condition, here we used a physically present robot. This may be more evocative than a mere picture (with respect to representing the other's action plans) and it also introduces physical embodied presence.

In order to test whether the reduction in SoA reported in Experiment 1 is indeed due to attribution of intentional agency, we conducted Experiment 2 in which the robot was replaced by a non-agentic passive entity. In Experiment 2, participants performed the same task of Experiment 1 with the only exception that the alternative trial outcome was associated with a faulty air pump which could stop the balloon by *non-action* (breaking down), thus it was unlikely to induce attribution of intentional agency.

3.1. Method and materials

3.1.1. Participants

Thirty new participants (11 males; 2 left-handed; Mean age: 24 ± 4.2 years), selected according to the same criteria as in the previous experiment, took part in Experiment 2. All participants gave their written informed consent and the study was conducted in accordance with the ethical protocol applied also in Experiment 1.

3.1.2. Apparatus and stimuli

The experimental setup consisted of: 1) Two air pumps, 2) A QWERTY keyboard, and 3) a 21' inches screen (1920 × 1080) to display the task. The participants were seated facing the air pumps. The screen laid horizontally on the table between the participant and the pumps (see Fig. 5). Stimuli were the same used in Experiment 1. Responses during the game were executed pressing the spacebar. SoA ratings were collected using a mouse. Stimulus presentation, response timing, and data collection were controlled as in Experiment 1.

3.1.3. Procedure

The task was the same as used in Experiment 1. Participants were instructed that, at the beginning of the game, they would receive 2500 points each, and in each trial, they could lose up to 100 of these points. Participants were instructed to perform as the best player of all (fictional) players. The task consisted of 12 blocks of 10 trials each, preceded by a short practice of 6 trials. Blocks were randomly assigned to

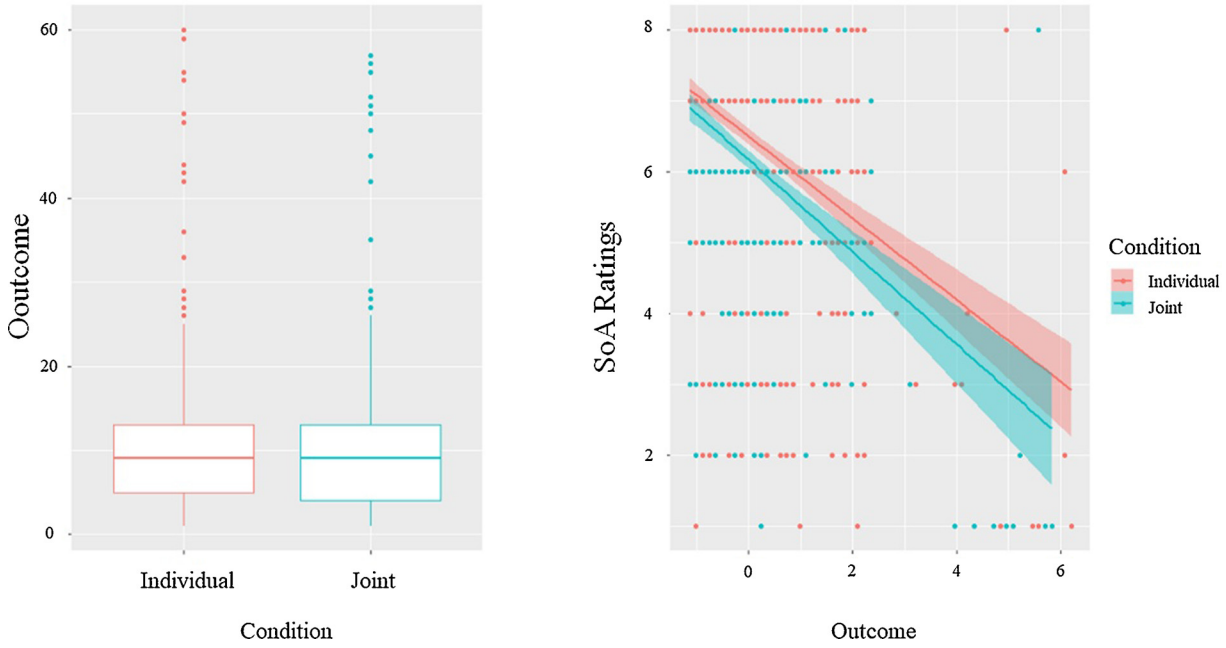

Fig. 4. Left Panel: Outcome plotted as a function of Condition (Individual vs. Joint). Right Panel: Sense of agency ratings plotted as a function of standardized outcome across Individual (red dots) and Joint (blue dots) conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Fig. 5. Experimental setup of Experiment 2. The Old pump (on the left) was marked with a blue sticker. The New pump (on the right) was marked with a green sticker. The New Pump had also a price tag attached to it, to make the impression that it was brand new from a store. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

either the 'New Pump' or the 'Old Pump' condition. The "New Pump" condition was designed to resemble the "Individual" condition of Experiment 1. Participants were instructed that in the New Pump trials the pump was working properly and it would inflate the balloon until the balloon bursts. Thus, if they did not act, the balloon would always burst and they would lose the maximum amount of points. For the Old pump trials (designed to resemble the "joint" Cozmo condition of Experiment 1), participants were told that the pump was broken and sometimes it would stop inflating the balloon before it reached the pin. If the pump stopped inflating the balloon before it reached the pin, the participant would not lose any points. If the participant stopped the balloon, s/he would lose a number of points according to the size at which they stopped it. As in Experiment 1, participants were instructed that the later they stopped the balloon, the fewer points they would lose. However, if the balloon burst, they would lose the maximum amount of points. As a result, as in Experiment 1, the action (i.e. stopping the balloon) resulted to be costly, but less costly than not acting. As in Experiment 1 (joint Cozmo trials) Old pump condition was programmed in order to stop the inflation sequence in the 60% of the trials only. (i.e. 36 out of 60 Old pump trials). In the Old pump condition, the inflating sequence was stopped when 90% of the sequence was run and no action was executed by the participant.

The sequence of events was the same as in Experiment 1, with only two exceptions. At the beginning of each trial, the starting frame indicated which pump was inflating the balloon: the New or the Old pump. In addition, during the inflation sequence two grey dots were presented at the upper corners of the screen, in order to appear in the proximity of the two pumps. When the inflation started, the dot close to the pump in charge to inflate the balloon turned from grey to blue or green accordingly to which pump was inflating the balloon (see Fig. 5). The speed of inflation was the same of Experiment 1 and it varied across and within trials. The exact number of points lost in each trial (i.e. the outcome) was computed as in Experiment 1, see Table1. To match the salience of the presence of the old pump with the presence of Cozmo, we presented two sounds. The sound of a pump inflating a balloon (1536kbps) was presented during the entire inflation sequence. The second sound was a 'beep' sound (1411kbps) presented when the Old pump broke down. The sounds were played through computer speakers.

3.2. Data analysis

Data analysis were performed as in Experiment 1, with the Pump condition factor (New vs. Old) instead of the Condition factor (Joint vs. Individual). For Valid trials we modelled balloon's Stop size as a function of Pump condition (New, Old), and the Outcome of each trial as a function of the Pump condition (New, Old) and Stop size. Finally, agency ratings for Valid trials were modelled using Pump condition (New, Old) and Outcome, plus their interactions. For Missed trials we addressed if SoA ratings were predicted by Pump condition (New, Old), the number of points lost (i.e. Outcome), or their interaction.

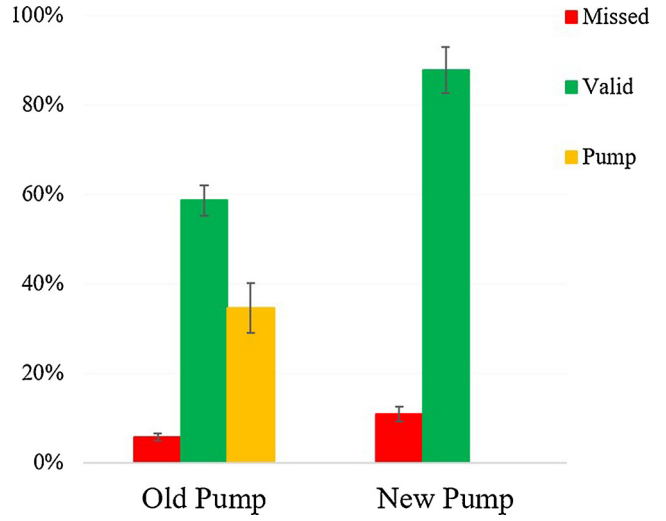

Fig. 6. Frequencies of responses for Experiment 2 plotted as function of Missed (red), Valid (green), and Pump trials (yellow) across Old Pump (corresponding to "Joint" condition of Experiment 1) and New Pump condition (corresponding to "Individual" condition of Experiment 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

3.3. Results and discussion

The balloon burst significantly more frequently when participants performed the task alone than when playing with the Old Pump, as indicated by higher percentage of Missed trials in the New Pump (corresponding to the "Individual" condition of Experiment 1) (M = 10.9%, SE = 1.7) than in the Old Pump condition (corresponding to the "Joint" condition of Experiment 1) (M = 5.7%, SE = 0.9) [t_{29} = 4.62 p < .001, d = .84]. In the Old Pump condition ("Joint"), the balloon burst less often than the pump broke, as Missed trials were less frequent than Pump trials (M = 34.6%, SE = 5.5) [t_{29} = 5.68, p < .001, d = 1.04] (Fig. 6).

3.3.1. Valid Trials

3.3.1.1. Stop size. The pump condition context predicted the size at which the balloon was stopped [β = 0.22, $t_{28.13}$ = 4.66, p < .001, 95% CI = (0.12, 0.31)]. In the Old pump condition, the stop size of the balloon was smaller (M = 2.61, SE = 0.02) compared to the New pump condition (M = 2.84, SE = 0.02).

3.3.1.2. Outcome. The outcome was predicted by the size at which participants stopped the balloon [β = -0.18, $t_{24.08}$ = 2.22, p = .036, 95% CI = (-0.34, -0.02)]. No effect of the Pump condition was found [β = 0.03, $t_{56.12}$ < 1].

3.3.1.3. Sense of agency (SoA) ratings. Results showed that agency ratings were predicted by the Outcome [β= -0.68, $t_{24.33}$ = -6.40, p < .001, 95% CI = (-0.89, -0.47)], with smaller losses being associated with higher SoA ratings. No differences between the Old pump condition (corresponding to "Joint" condition of Experiment 1) (M = 5.72, SE = 0.05) and the New pump condition (corresponding to "Individual" condition of Experiment 2) (M = 5.82, SE = 0.04) were found in the agency ratings [β = 0.09, $t_{27.98}$ < 1]. There was no significant interaction.

3.3.2. Missed Trials

When the balloon burst, agency ratings were predicted neither by the Pump condition or the Outcome (all ps > .11).

3.3.3. Discussion

Experiment 2 aimed at investigating whether the subjective reduced SoA reported is indeed due to attribution of intentional agency towards the artificial entity. To this end, we asked participants to perform the same task as in Experiment 1, but with a passive non-agentic entity (broken air pump) instead of a robot which potentially induces attribution of intentional agency. Similarly to Experiment 1, results showed higher percentage of missed trials when participants were alone in charge of stopping the balloon (New Pump = "Individual" condition) compared to when an external event it was potentially causing it (Old Pump = "Joint" condition). When participants successfully stopped the balloon, the Pump condition predicted the size at which the balloon was stopped, suggesting that when the Old pump was in charge participants adopted overall a lower-risk strategy. The outcome was predicted by the size at which participants stopped the balloon. This latter result may be a consequence of the lower rate of missed trials for the Old Pump condition resulting in a lower rate of late-action trials. This might be due to that participants "trusted" less a "broken" pump and preferred to take less risks, as compared to the analogous "Joint" condition in Experiment 1, where such effect was not found. Finally, and most importantly for the purposes of our study, the perceived SoA over the outcome was not affected by the type of pump that was inflating the balloon, as indicated by the lack of the main effect of Pump condition. This result suggests that, in contrast to the presence of the Cozmo robot, the faulty passive, non-agentic mechanical device did not affect participants' SoA. This is in line with our hypothesis, that the reduced SoA observed in the presence of Cozmo is due to attribution of intentional agency to the robot.

4. Experiment 3

Results of Experiment 2 showed that when the alternative trial outcome is the result of a non-agentic mechanical device no effect of reduced SoA was observed. This latter results suggest that the reduced SoA reported in the Joint condition of Experiment 1 can be interpreted as the consequence of representing the robot as an intentional agent, and the mentalising process related to its action plan, which might have interfered with participants' own action planning (Beyer et al., 2017, 2018). However, one might argue that the manipulation of Experiment 1 differed from Experiment 2 in terms of attribution of mere agency only, not necessarily intentional agency. In order to confirm that the observed effect of reduced SoA in the robot condition is due to attribution of intentional agency, we compare the effect of reduced SoA in interaction with a robot to the reduced SoA in human-human interaction. To this end, we conducted Experiment 3 in which participants performed the task with the Cozmo robot in one session and with another human (confederate) in another session (within-participants design). We hypothesised that if the reduced SoA observed in Experiment 1 is due to the fact that the robot is perceived as an intentional agent, then, firstly, we should replicate the reduced SoA in Joint compare to Individual condition of Experiment 1, secondarily, no difference across the interactive agents (Human vs. Robot) should be observed.

4.1. Method and materials

4.1.1. Participants

Thirty new participants (10 males; 4 left-handed; Mean age: 26 ± 5.0 years), selected according to the same criteria as in the previous experiments, took part in Experiment 3. All participants gave their written informed consent and the study was conducted in accordance with the ethical protocol applied also in Experiment 1 and 2. Four participants were excluded from data analysis because during the session with Cozmo the Android application crushed and the robot stopped working.

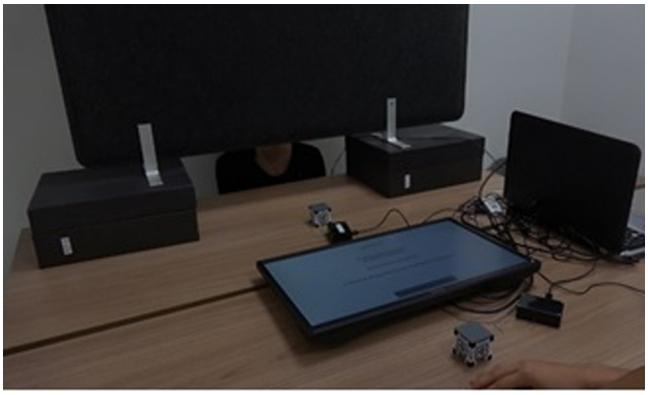

Fig. 7. Experimental setup of the Human agent session in Experiment 3.

4.1.2. Apparatus and stimuli

The experimental setup was exactly the same as in Experiment 1 with only one exception: responses were executed pressing a one-key keyboard attached on Cozmo cubes (see Fig. 7). SoA ratings were collected using a mouse. Stimulus presentation, response timing, and data collection were controlled as in Experiment 1.

4.1.3. Procedure

The experiment consisted of two sessions run in two separate days. In both sessions, the task was the same as in Experiment 1. Across sessions, we manipulated the interactive agent. In one session, the interactive agent was Cozmo, whereas in the other session the interactive agent was a human confederate. The order of the two sessions was counterbalanced across participants. In the Human Agent session, the confederate was seating facing the participant. However, the confederate's face was occluded, thus participants were able to see only the hand of the confederate (see Fig. 7). This was done in order to control for any implicit signals that participants could have picked up from the confederate's face (mutual gaze, facial micro-expressions, perhaps emotional expressions such as smiles, etc.). Participants were instructed that, at the beginning of the game, they would receive 2500 points each, and in each trial, they could lose up to 100 of these points. As in Experiment 1 and 2, participants were instructed to perform as the best player of all (fictional) players. The task consisted of 12 blocks of 10 trials each, preceded by a short practice of 6 trials. Blocks were randomly assigned to either the Joint or the Individual condition. In both sessions, the Individual condition was the same as Experiment 1 and 2. For the Joint trials, in the Cozmo session, the robot was programmed to act exactly as in Experiment 1. In the Human Agent session the confederate was "triggered" via headphones to play with the same risk-taking behaviour we programmed for the robot (i.e. acting only in the 60% of joint trials and only when the 90% of the inflation sequence was reached). As in Experiment 1 and 2, participants were instructed that the later they stopped the balloon, the fewer points they would lose. However, if the balloon burst, they would lose the maximum amount of points. As a result, as in Experiment 1 and 2, the action (i.e. stopping the balloon) resulted to be costly, but less costly than not acting. The sequence of events was the same as in Experiment 1 and 2. The speed of inflation was the same as Experiment 1 and 2 and it varied across and within trials. The exact number of points lost in each trial (i.e. the outcome) was computed as in Experiment 1 and 2, see Table 1. After the Cozmo Agent session, participants filled out the Godspeed Questionnaire (Bartneck, Kulić, Croft, & Zoghbi, 2009) in order to evaluate their perception of the robot. The questionnaire consists of twenty-four semantic differential scales, on which participants need to rate their positioning between two opposing adjectives describing the robot on a 5 point scale. The ratings are summed to five subscales: "Anthropomorphism", "Animacy", "Likeability", "Perceived Safety", and "Perceived Intelligence". Each of the subscales has been psychometrically validated (Bartneck et al., 2009).

4.2. Data analysis

Data analysis were performed as in Experiment 1, with the only addition of Agent (Robot, Human) as fixed factor. For Valid trials, we modelled balloon's Stop size as a function of Condition (Individual, Joint) and Agent (Robot, Human), plus their interactions. The Outcome of each trial was analysed as a function of Condition (Individual, Joint), Agent (Robot, Human), and Stop size, plus their interactions. Finally, agency ratings for Valid trials were modelled using Condition (Individual, Joint), Agent (Robot, Human), and Outcome, plus their interaction. Sperman tests were run to test the relation between average agency ratings in valid trials of the Joint condition for the Cozmo session and scores of the Goodspeed subscales. For Missed trials we addressed if SoA ratings were predicted by Condition (Individual, Joint), Agent (Robot, Human), and the number of points lost (i.e. Outcome), or their interaction.

4.3. Results and discussion

4.3.1. Cozmo agent session

Balloon bursts were equally frequent across conditions (Fig. 8), as indicated by a similar percentage of Missed trials in the Individual (M = 9.1%, SE = 0.01) and Joint condition (M = 7.4%, SE = 0.01) [t_{25} = 1.76, p = .091, d = 0.35]. In the Joint condition, Cozmo acted more often than the balloon burst, as Missed trials were less frequent than Cozmo trials (M = 43.9%, SE = 0.02) [t_{25} = 21.12, p < .001, d = 4.14].

4.3.2. Human agent session

Balloon bursts were equally frequent across conditions (Fig. 8) as indicated by similar percentage of Missed trials in the Individual (M = 8.3%, SE = 0.01) and Joint condition (M = 8.1%, SE = 0.01) [t < 1]. In the Joint condition, the confederate acted more often than the balloon burst, as Missed trials were less frequent than other agent trials (M = 36.5%, SE = 0.02) [t_{25} = 18.02, p < .001, d = 3.53].

4.3.3. Comparison across agents

A paired-sample *t*-test was run to compare whether the percetages of trials in which participants let the other agent act differed across the type of agent (Human vs. Robot). Results showed that the percetages of other agent acting trials were lower when the participants interacted with a Human (M = 36.5%; SE = 0.02) than with the robot (M = 43.9%; SE = 0.02) [t_{25} = 4.76, p < .001, d = 0.93].

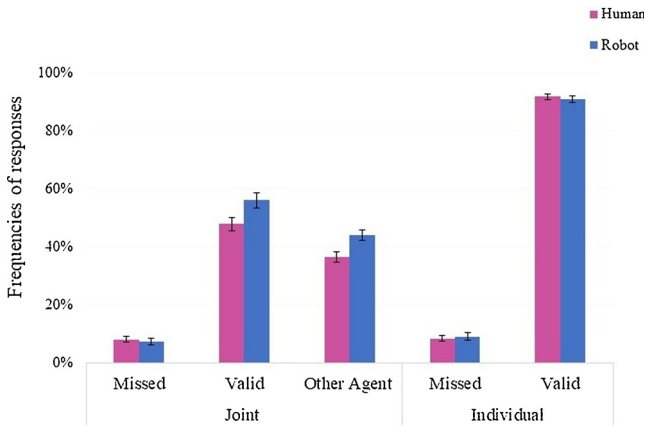

Fig. 8. Frequencies of responses for Experiment 3 plotted as function of Missed, Valid and other agent (Human: Pink or Robot: Purple) agent trials across Joint and Individual condition. (For interpretetion of the references to colours in this figure legend, the reader is refereed to the web version of this article).

Percetanges of Missed and Valid trials were submitted to two separate ANOVAs with Condition (Individual vs Joint) and Agent (Human vs Robot) as within-subjects factors. No significant main effects or interaction were found for Missed trials (all ps > .11). The analysis on Valid trials revealed a main effect of Condition [F_{25} = 435.24, p < .001, η_p^2 = 0.95], Agent [F_{25} = 8.48, p = .007, η_p^2 = 0.25], and a significant interaction [F_{25} = 26.10, p = .007, η_p^2 = 0.51]. Paired-samples *t*-test showed that for the Joint Condition, the parcentages of Valid trials was lower when the interactive agent was a Human (M = 47.9%, SE = 0.02) than a Robot (M = 56.0%, SE = 0.03) [t_{25} = 4.16, p < .001, d = 0.82], whereas no differences across agents was found for the Individual condition [t < 1].

4.3.4. Goodspeed questionnaire

The mean total score of the Godspeed questionnaire was 82.96 (SD = 17.04). Average scores and standard deviations of each subscale are reported in Table 3.

4.3.5. Valid Trials

4.3.5.1. Stop size. The social context predict the size at which the balloon was stopped [β = 0.12, $t_{25.9}$ = 2.92, p = .005, 95% CI = (0.04,0 .20)] as well the interaction with the Agent [β = 0.12, $t_{25.9}$ = -2.26, p = 0.02, 95% CI = (-0.23, -0.02)]. When the agent was the robot, the balloon was stopped at smaller sizes in the Joint (M = 3.33, SE = 0.02) compared to the Individual condition (M = 3.42, SE = 0.01) [$p_{bonferroni\ corrected}$ = .021]. No differences across conditions were found in stop balloon size when the confederate was a human agent (Joint: M = 3.37, SE = 0.02; Individual: M = 3.40; SE = 0.01, see Fig. 9, left panel).

4.3.5.2. Outcome. As in Experiment 1 and intended by the experimental design, the outcome was predicted by the size at which participants stopped the balloon [β = -0.24, $t_{20.1}$ = -5.99, p < .001, 95% CI = (-0.32, -0.16)].

4.3.5.3. Sense of agency (SoA) ratings. Results showed a significant reduction in agency ratings in the Joint (M = 6.28, SE = 0.03) compared to the Individual (M = 6.58, SE = 0.03) condition [β = 0.33, t_{25} = 3.12, p = .004, 95% CI = (0.12, 0.54)]. Agency ratings were also predicted by the Outcome [β = -0.48, $t_{25.2}$ = -6.22, p < .001, 95% CI = (-0.64, -0.33)], with smaller losses being associated with higher SoA ratings (Fig. 9, right panel). There was no significant interaction between Condition and Agent [β = 0.08, t_{25} = 1.48, p = .140, 95% CI = (-0.18, 0.03)] or between Condition and Outcome [t < 1]. No significant correlation was found between average SoA rating in the joint condition for the robot agent and mean scores of the Goodspeed subscales (all ps > .35).

4.3.6. Missed Trials

When the balloon burst, agency ratings were predicted neither by the Condition, the Outcome or the Agent (all ps > .14).

4.3.7. Discussion

Experiment 3 aimed to test whether the observed reduced SoA in Experiment 1 was due to the attribution of intentional agency. To this end, in a within-participants design, we compare the effect of reduced SoA when participants interacted with the robot to the reduced SoA in human-human interaction. Results showed that in the Joint condition, the balloon was stopped by the other agent more frequently than it burst. Across agents, participants let the robot act more often than the human confederate. This difference between agents was also present in Valid trials, with participants acting more often when they were interacting with a human than with Cozmo. Such a difference can be due to the fact that, even if the action was triggered at the very same point of the inflation sequence, the human confederate may react faster then Cozmo. Indeed, once the command is sent through the Android Debug Bridge to the Cozmo Application, then it has to be transmitted to the robot via wi-fi. The longer time needed to the robot to act may have prompt participants to act earlier in the Joint condition compare to when they were playing alone, explaining the smaller average size at which the ballon was stopped. However, such differences in the performance did not result in a different amount of lost points in each trial, as indicated by the lack of the Agent main effect over the outcome. Thus, we can assume that they did not affect agency ratings, given that participants were explicitly instructed to rate the control they perceived over the number of lost points.

As for Experiment 1, when participants successfully stopped the balloon, SoA was rated as lower in the Joint than in the Individual condition, independently of the number of lost points. Moreover, in accordance with results of Experiment 1 and 2, SoA was reduced for more negative outcomes, confirming that participants followed the instructions and rated their perceived control over the outcome, rather than over the success of the trial. Importantly, there was no Condition*Agent interaction, indicating that interacting with a robot reduces SoA, similarly to the case of human-human interaction (Beyer et al., 2017, 2018).

5. General discussion

The present study investigated whether attribution of intentional

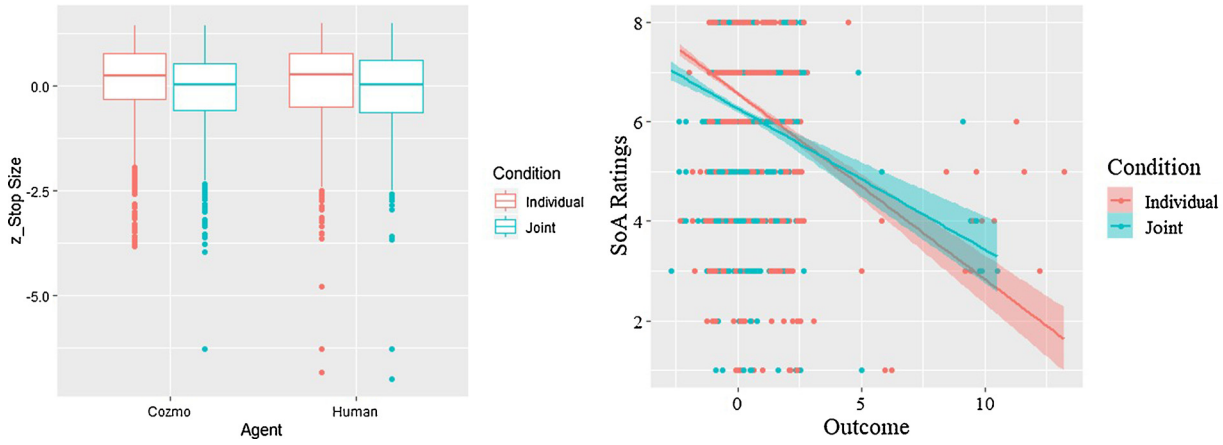

Fig. 9. Left Panel: Stop size z-scores plotted as a function of Condition (Individual vs. Joint) and Agent (Human vs. Robot). Right Panel: Sense of agency ratings plotted as a function of standardized outcome across Individual (red dots) and Joint (blue dots) conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

agency to an artificial agent (an embodied robot) evokes the phenomenon of reduced SoA as it has been observed in human-human interaction. To this end, we tested SoA when people interacted with a robot (Experiment 1 and 3), with a passive, non-agentic air pump (Experiment 2) and with a human (Experiment 3). Participants were asked to rate the perceived control they felt on the outcome of their action while performing a diffusion of responsibility task (Beyer et al., 2017, 2018). We hypothesized that if humans perceive robots as intentional agents (or agentic artefacts), then we would expect a similar decrease in SoA as observed in the presence of human interaction partners (Beyer et al., 2017, 2018). On the contrary, if robots are perceived as non-agentic artefacts, then the SoA experienced in interaction with robots should be comparable to acting alone. Results of Experiment 1 showed that, when participants successfully stopped the balloon (i.e. they performed an action), they rated SoA on the outcome as lower in trials in which the robot was also in charge of stopping the balloon (Joint condition), compared to when they were performing the task alone (Individual condition). Experiment 2 aimed at investigating whether the ascription of intentional agency is the key factor for the reduction in SoA to occur. In Experiment 2, the robot of the "Joint" condition was replaced by a faulty air pump that could "passively" stop the balloon from inflating by simply breaking down. Results showed that when participants performed an action (i.e. they successfully stopped the balloon), perceived SoA over the outcome was not affected by the type of pump (functioning or not, thereby comparable to the Individual or Joint condition), as indicated by the lack of the main effect of Pump condition. This shows that when in interaction with a non-agentic passive device, SoA is not affected. Thus, the reduction in SoA reported in Experiment 1 cannot be interpreted as a consequence of mere physical presence of an entity. This speaks in favour of the interpretation that attribution of intentional agency to others is crucial for the effect of reduced SoA. However, in order to test if it is indeed attribution of intentional agency, and not mere agency, we conducted Experiment 3, in which an interaction with a human agent was introduced. In a within-participants design, participants performed the task in two sessions: once with another human agent (confederate) and once with Cozmo. Results showed that SoA was reduced similarly in the human and Cozmo conditions, threby indicating that attribution of intentional agency plays a crucial role in reduction of SoA, in line with the model of Beyer et al. (2017,2018). It should be noted that compared with Experiments 1 and 3, agency ratings in Experiment 2 were overall lower (see Table 2). It could be possible that in Experiment 2, introducing one of the pumps as "faulty" might biased participants, who perceived the entire experimental setting as unreliable. Thus, they acted faster than in Experiment 1 and 3 (see Table 2), and as a result the balloon was stopped earlier, resulting in worst outcomes on average. This would then be reflected in lower agency ratings (Table 2). It could be argued that instructing participants to save points without any monetary reward associated to it may have not been sufficiently motivating. However, our findings mirror those found previously with using monetary reward, both in terms of task performance, and sensitivity of agency ratings to outcome magnitude and social context (Beyer et al., 2017; 2018).

Taken together, these results suggest that interacting with robots reduces SoA, similarly to when we interact, or we believe to interact, with another human (Beyer et al., 2017, 2018). As sense of agency was measured after the outcome was presented, the current paradigm does not allow us to distinguish, whether the presence of another agent affected the online experience of sense of agency, or the post-hoc evaluation of it. However, this does not limit the relevance of our findings for our theoretical framework and our understanding of human-robot interactions, as the only difference between conditions occurred during the action phase of the task. Thus, cognitive processes during the planning and execution of an action affected the subsequent judgement of control, which is in line with previous findings showing a correlation between neural processes during the action phase and post-hoc judgements of agency (Beyer et al., 2018).

Our results give new hints regarding SoA during joint action with embodied artificial agents. Indeed, previous studies investigating SoA with artificial agents mainly focused on vicarious SoA during human-computer interaction (e.g., Obhi & Hall, 2011). For instance, Sahaï et al. (2019) systematically compared SoA when participants were performing a task alone, together with another human being, or with a desktop computer. In line with literature (see Haggard, 2017 for a review), Sahaï et al. (2019) reported a lack of vicarious SoA during human-computer interaction, as compared to when participants were performing the task with a human co-agent (see also Obhi & Hall, 2011; Wohlschläger, Haggard, Gesierich, & Prinz, 2003). It has been proposed that in human-computer settings, humans do not attribute agency to the artificial system, since they cannot use their sensorimotor system to simulate and understand the machine-generated action (Obhi & Hall, 2011; Wohlschläger et al., 2003). As a consequence, they are unable to develop a vicarious SoA (Berberian, Sarrazin, Le Blaye, & Haggard, 2012; Sahaï, Desantis, Grynszpan, Pacherie, & Berberian, 2019). Accordingly, when the artificial agent is an embodied anthropomorphic hand, vicarious SoA occurs in a similar way as during observation of another person performing the same action (Khalighinejad et al., 2016). In line with previous literature, our results suggest that when studying SoA during social interaction with artificial agents considering the embodiment may play a crucial role[2] . This is quite plausible, given that intentional agency is defined as the ability to plan and act. While an embodied robot can act in the environment, manipulate objects and provide tangible physical presence, a disembodied computer program cannot. Interestingly, in the previous studies (Beyer et al., 2017, 2018) agency has been attributed also to non-embodied avatars that were believed to be controlled by a human. In this case, however, participants were presumably aware that the avatar is just a "representation" of a human agent, and the intentional agency was attributed to the human controlling the avatar, and not to the avatar itself. When an "agent" is artificial, as in the case of a computer, its "action" and its consequences belong to two different frames of reference. Specifically, while the "action", i.e. the command in a computer programme, belongs to a digital /virtual environment, the effect that it generates

Table 2
Means and standard deviations (in brackets) of Stop size, Outcome, and Agency ratings across the three experiments.

	Experiment 1	Experiment 2	Experiment 3
Stop size	3.25 (0.68)	2.75 (0.73)	3.39 (0.49)
Outcome	9.86 (7.84)	10.84 (10.85)	9.44 (7.85)
SoA rating	6.39 (1.64)	5.78 (1.65)	6.47 (1.39)

Table 3
Mean scores and standard deviations for the subscales of Goodspeed questionnaire in Experiment 3.

Subscale	M	SD
Animacy	19.39	3.83
Anthropomorphism	13.64	4.53
Likeability	25.18	3.83
Perceived Intelligence	17.81	3.29
Perceived Safety	6.94	1.55

[2] Note however, that obviously embodiment is not enough for attribution of agency to the other entity, as additionally, and by definition, acting in the real world needs to be an active process. This is why our "embodied" objects of Experiment 2 – the air pumps – did not evoke reduced SoA: in that case, stopping the balloon from inflating was not defined as an "action" of the broken pump but rather a consequence of its passive state of "inaction".

occurs in a physical environment. Thus, reduced (or eliminated) attribution of intentional agency to computer programs might be driven by a difficulty in representing the cause-effect link between the action and its consequences. In these terms, the embodied nature of robots allow to overcome the gap between frames of reference, since robots "act" in the physical environment through "physical" action events.

Our results are important for the development of robots that are to co-exist with humans in daily living. It has been proposed that decreased SoA plays a critical role in the diffusion of responsibility (feeling less responsible for the consequences of one's actions, especially when those consequences are negative, Bandura, 1991). The result of diffusion of responsibility in social situations is that humans tend to decrease the likelihood of performing an action in the presence of others. For instance, the likelihood that someone will intervene in an emergency situation decreases in a crowd (Chekroun & Brauer, 2002; Latane & Darley, 1968). Given that in future robots will be present in our social environments, as our houses or public spaces (e.g., hospitals, schools, and airports), it is important to be aware that robots might evoke similar diffusion of responsibility as other humans do. We propose that the design of robots' behaviour in social contexts should consider the impact that the presence of an embodied artificial agent exerts on humans' decision-making. Therefore, in emergency situations, it would be best if robots are able to efficiently detect an emergency signal and act upon it, as the human counterparts may not be efficient and fast enough. Interestingly, we did not find a correlation between how participants perceived the robot and SoA ratings, however, it is possible that individual differences in adoption of intentional stance towards robots (Marchesi et al., 2019) may also affect the diffusion of responsibility. Indeed, it could be that differences based on the degree of familiarity with robots can be reflected also in the stance (designed or intentional Dennett, 1971; 1981) used to explain and represent robots' behaviour. Future research should address this possibility by evaluating, for example, how SoA is reduced when interacting with robots across groups with different level of expertise about artificial agents (e.g. engineers in robotics vs historians).

In summary, our study showed that interaction with a robot induces reduced SoA, similarly to interaction with another human. Reduced SoA is not observed in the presence of a passive non-agentic device. Reduced SoA in interaction with robots might have consequences not only for theoretical understanding of mechanisms of cognition but can also have practical implications for how (and when) we act in the "social" presence of robots – and possibly other artificial agents - in real life. To fully understand the mechanisms behind such phenomena as reduced SoA in HRI, future studies should combine behavioural and electrophysiological measures in order to investigate if the reduction in agency ratings when interacting with a robot is also mirrored in the processing of action-outcome at the neural level. Furthermore, systematic manipulation of human-like appearance of the robot might be examined to understand the role of physical appearance in attribution of intentional agency and reduction of SoA.

Author contribution

F.C. conceived, designed and performed the study; analyzed the data, discussed and interpreted the results; wrote the manuscript. F.B. conceived and designed the study, discussed and interpreted the results; wrote the manuscript. D.D.T. Programmed the Cozmo robot and wrote the "Apparatus and Materials" section. A.W. conceived and designed the study; discussed and interpreted the results; wrote the manuscript. All authors reviewed the manuscript.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant awarded to AW, titled "InStance: Intentional Stance for Social Attunement". Grant agreement No: 715058)

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

The authors are grateful to N.A.H. for having taking part as a confederate in Experiment 3. Preliminary results of a pilot experiment related to this study appeared in Proceedings of the 10th International Conference on Social Robotics, Qingdao, 28-30 November 2018: Ciardo, De Tommaso, Beyer, & Wykowska, A. (2018). Reduced in Human-Robot interaction. In Ge, Cabibihan, Salichs, Broadbent, He, Wagner, & Castro-González (Eds.), *Social Robotics* (pp. 441-450). Qingdao, CHN: Springer.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.cognition.2019.104109.

References

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*(2), 248–287.
Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71–81.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(7), 1–23.
Berberian, B., Sarrazin, J. C., Le Blaye, P., & Haggard, P. (2012). Automation technology and sense of control: A window on human agency. *PloS One, 7*(3), e34075.
Beyer, F., Sidarus, N., Bonicalzi, S., & Haggard, P. (2017). Beyond self-serving bias: Diffusion of responsibility reduces sense of agency and outcome monitoring. *Social Cognitive and Affective Neuroscience, 12*(1), 138–145.
Beyer, F., Sidarus, N., Fleming, S., & Haggard, P. (2018). Losing control in social situations: How the presence of others affects neural processes related to sense of agency. *eNeuro* ENEURO-0336.
Chambon, V., Sidarus, N., & Haggard, P. (2014). From action intentions to action effects: How does the sense of agency come about? *Frontiers in Human Neuroscience, 8*, 320.
Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutscher, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience, 6*, 103.
Chekroun, P., & Brauer, M. (2002). The bystander effect and social control behavior: The effect of the presence of others on people's reactions to norm violations. *European Journal of Social Psychology, 32*(6), 853–867.
Ciardo, F., & Wykowska, A. (2018). Response coordination emerges in cooperative but not competitive joint task. *Frontiers in Psychology, 9*, 1919.
Ciardo, F., De Tommaso, D., Beyer, F., & Wykowska, A. (2018). *Reduced sense of agency in human-robot interaction.* November*International conference on social robotics*. Cham: Springer441–450.
Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy, 68*, 87–106. https://doi.org/10.2307/2025382.
Dennett, D. C. (1981). *True believers: The intentional strategy and why it works.* Cambridge,MA: MITpress.
Duffy, B. R., & Joue, G. (2000). Intelligent robots: The question of embodiment. *Proceedings of the Brain-Machine Workshop*.
Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
Engbert, K., Wohlschläger, A., & Haggard, P. (2008). Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition, 107*(2), 693–704.
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.
Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619.
Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience, 18*(4), 196.
Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243–259.
Khalighinejad, N., Bahrami, B., Caspar, E. A., & Haggard, P. (2016). Social transmission of experience of agency: An experimental study. *Frontiers in Psychology, 7*, 1315.
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), e2597.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest'. *R package version, 2*(0).

Latane, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology, 10*(3), 215.

Marchesi, S., Ghiglino, D., Ciardo, F., Perez-Osorio, J., Baykara, E., & Wykowska, A. (2019). Do we adopt the intentional stance toward humanoid robots? *Frontiers in Psychology, 10*, 450. https://doi.org/10.3389/fpsyg.2019.00450.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324.

Obhi, S. S., & Hall, P. (2011). Sense of agency in joint action: Influence of human and computer co-actors. *Experimental Brain Research, 211*(3–4), 663–670.

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.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*(3), B11–B21.

Sidarus, N., Vuorre, M., & Haggard, P. (2017). How action selection influences the sense of agency: An ERP study. *NeuroImage, 150*, 1–13.

Stenzel, A., Chinellato, E., Bou, M. A. T., del Pobil, Á., 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.

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.

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.

Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*. https://doi.org/10.1016/j.cub.2019.01.046.

Wen, W. (2019). Does delay in feedback diminish sense of agency? A review. *Consciousness and Cognition, 73*, 102759.

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I see what you mean: How attentional selection is shaped by ascribing intentions to others. *PloS One, 7*(9), e45391.

Wohlschläger, A., Haggard, P., Gesierich, B., & Prinz, W. (2003). The perceived onset time of self-and other-generated actions. *Psychological Science, 14*(6), 586–591.

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PloS One, 9*(4), e94339.

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychonomic Bulletin & Review, 16*(6), 1089–1093.