# Joint agency in Human-Machine Interactions: How to design more cooperative agents ?

Sahai Aisha

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PSL

Préparée à l'Ecole Normale Supérieure (ENS)
et à l'Office National d'Etudes et de Recherches Aérospatiales (ONERA)

# Agentivité conjointe lors des Interactions Homme-Machine: comment concevoir des agents plus coopératifs?


Soutenue par

**Aïsha Sahaï**

Le 17 Juillet 2019


Ecole doctorale n° 158

**Cerveau, Cognition, Comportement**

Spécialité

**Sciences cognitives**

ENS | PSL

ONERA
THE FRENCH AEROSPACE LAB

Composition du jury :

| | |
|---|---|
| Mme. Magalie OCHS<br>Maître de conférences - HDR<br>*LSIS, Aix-Marseille Université* | *Rapporteur* |
| M. Patrick HAGGARD<br>Professeur des Universités<br>*ICN, University college of London* | *Rapporteur* |
| Mme. Aurélie CLODIC<br>Ingénieure de recherche<br>*LAAS, CNRS* | *Examinateur* |
| M. Bert TIMMERMANS<br>Maître de conférences<br>*Sinclab, University of Aberdeen* | *Examinateur* |
| M. Jean-Louis VERCHER<br>DR<br>*ISM, Aix-Marseille Université* | *Examinateur* |
| Mme. Elisabeth PACHERIE<br>DR<br>*Institut Jean Nicod - ENS* | *Directrice de thèse* |
| M. Bruno BERBERIAN<br>Ingénieur de recherche<br>*ONERA* | *Examinateur (co-encadrant)* |
| M. Ouriel GRYNSZPAN<br>Professeur des Universités<br>*LIMSI- Université Paris Sud* | *Examinateur (co-encadrant)* |

# Joint agency in Human-Machine Interactions: How to design more cooperative agents?

“Perhaps the major human factors concern of pilots in regard to introduction of automation is that, in some circumstances, operations with such aids may leave the critical question, who is in control now, the human or the machine?”

(*Baron, 1988*)

# Abstract

System automation has steadily created a gap between the human operators and the loop of control (i.e., “out-of-the-loop” (OOTL) problem), disconnecting them from the machines’ actions and outcomes (Kaber, Onal, & Endsley, 2000). In this thesis, we aimed at investigating how to keep the human operators in the loop of control. We based our investigations on the theoretical framework of the science of agency. The sense of agency can be defined as the experience of controlling one’s own actions, and, through them, the events in the outside world (Haggard & Chambon, 2012). Interestingly, it has been shown that during human-human interactions, individuals could exhibit a sense of agency for other-generated actions and outcomes (or sense of “we-agency”) while such ability was impaired for machine-generated actions and outcomes (Obhi & Hall, 2011b). The first stage of the thesis sought to finely examine the cognitive processes underlying individuals’ loss of agency during joint tasks with automated artificial systems, both at the behavioral (Experiment 1) and at the cerebral (Experiment 2) levels. Using a joint Simon task coupled with an intentional binding task, we found that the ability to develop a sense of agency for other-generated action outcomes was in pace with the ability to represent other-generated actions into one own cognitive system, which were both impaired during human-machine interactions. The second stage of the thesis sought to investigate on which characteristics of the machine it was possible to act in order to regain the human operator’s sense of agency. Firstly, using a top-down approach, we found that humanoid machines could mitigate the loss of agency during joint tasks (Experiment 3). Secondly, using a bottom-up approach, we found that individuals’ sense of agency was sensitive to kinesthetic feedbacks (Experiment 4). Implications of our findings are discussed in regard with the literature on the sense of agency and the operational OOTL issue.

# Résumé de la thèse

A travers l'automatisation des systèmes, les mutations technologiques ont progressivement éloigné l'opérateur humain des actions et des effets générés par les machines. Ce phénomène de « sortie de boucle » (Kaber, Onal, & Endsley, 2000) a généré un ensemble de difficultés pour l'opérateur humain qui restent encore aujourd'hui difficile à appréhender et compenser. Au cours de cette thèse, nous avons proposé un cadre théorique, celui de l'Agentivité, permettant d'expliquer et de compenser ces difficultés. Le sens d'agentivité peut se définir comme le sentiment de contrôle de ses propres actions, et à travers elles, des évènements de l'environnement extérieur (Haggard & Chambon, 2012). De manière intéressante, il a été montré qu'au cours des interactions purement humaines, les individus pouvaient ressentir un sens d'agentivité pour des actions et effets générés par autrui (ou « sens d'agentivité conjoint»). En revanche, la capacité à développer un sens d'agentivité conjoint et à se représenter les actions générées par un système artificiel automatisé semble altérée lors des interactions homme-machine (Obhi & Hall, 2011b). La première partie de la thèse a consisté à examiner finement les mécanismes sous-jacents de cette perte d'agentivité lors de tâches conjointes avec les systèmes automatisés, à la fois au niveau comportemental (Expérience 1) et au niveau cérébral (Expérience 2). En utilisant une tâche de Simon conjoint couplée à une tâche de liage intentionnel, nous avons montré que la capacité à développer un sens d'agentivité conjoint allait de pair avec la capacité de se représenter dans son propre système cognitif les actions générées par un partenaire. Nous avons également montré que ces deux processus étaient altérés lors de nos interactions avec une machine. La deuxième partie de la thèse visait à déterminer sur quelles propriétés de la machine il était possible d'agir pour créer un sens d'agentivité conjoint avec des systèmes artificiels automatisés. Premièrement, en utilisant une approche top-down, nous avons montré que l'apparence physique humanoïde d'une machine atténuait la perte d'agentivité classiquement reportée lors des interactions homme-machine (Expérience 3). Deuxièmement, en utilisant une approche bottom-up, nous avons montré que le sens d'agentivité conjoint était sensible aux informations sensorimotrices générées par des retours haptiques (Expérience 4). Les implications de nos résultats sont discutées au regard de la littérature sur l'Agentivité et de la problématique opérationnelle du phénomène de « sortie de boucle ».

# Acknowledgements

My first thanks go to my three co-supervisors, Prof. Elisabeth Pacherie, Prof. Ouriel Grynszpan and Ph. D. Bruno Berberian. How lucky I felt to have each one of you during this journey, which is the doctorate. Thank you for your benevolence.

I also express my thanks and appreciation to Prof. Magalie Ochs and Prof. Patrick Haggard to examine this dissertation.

I would like to thank Prof. Liliane Borel who gave me the working conditions to run the EEG study at the Sensory and Cognitive Neurosciences laboratory (Aix-Marseille University, France). I also express my gratitude to Prof. Axel Cleeremans who made possible the experiment with the servomotor and Pepper at the Research center of Cognition and Neurosciences (Free University of Brussels, Belgium). A special thank goes to Ph. D. Emilie Caspar and research engineer Albert De Beir who contributed to this work and had kindly accepted to lend us the robots.

I must concede that I have a particular thought for Pr. Yann Coello who sparked my interest in cognitive sciences when I was a master student. I am grateful for it.

I would also like to say some words to my colleagues from ONERA :

Zaynab, my first words are for you. Thank you for EVERYTHING. I have so many good memories with you.

Sala, thank you for making me laugh every single day! And by the way, thank you for the help with the statistics. Mick, Anne-Laure, Kelly, Nico P., Alexis, Xavier B., Vala, Joseph, Nico C., Nico M., Patrick L.B., Armin, PMB, Stéphane, Jean, Xavier D., Christelle, Christian C., thank you for your kindness and your indefectible support. Raphaël, thank you for all the good times we spent.

Alex, I could write many pages about you but the manuscript length is limited so I will be brief: you are the most beautiful person I've met over the past three years.

Laurent, I spend my entire doctorate with you as the director of the center. Thank you for your nice words and for the daily good vibes as well.

To the trio Greg, J-C, Chri., thank you for your encouragement (and Greg, also for all the chocolates in the white and blue packaging).

Last but not least, Andrea, thank you for all you have taught me, from the EEG to the linear models. You rock!

Finally, my deepest gratitude goes to my family, Papa, Ali, Hari, Theepa, Shayla, Sohan, Sharvin, Karim, my aunts Lydie and Françoise, my uncle Claude. I've found my motivation in you.

# Content

## Content

# List of Abbreviations

| | |
|---|---|
| **ANOVA** | Analysis of Variance |
| **CM** | Comparator Model |
| **EEG** | Electroencephalographic |
| **ERP** | Event Related Potential |
| **IB** | Intentional Binding |
| **JoA** | Judgment of Agency |
| **MNS** | Mirror Neuron System |
| **OOTL** | Out-of-the-loop |
| **RT** | Response Time |
| **SoA** | Stimulus onset Asynchronies |
| **SSE** | Social Simon Effect |
| **TMS** | Transcrananial Magnetic Stimulation |

# List of Figures

# List of Tables

# Chapter 1: Preamble

## 1.1. Technical progress and system automation

For the last decades, the world surrounding us has become increasingly technological. Indeed, technical progress has made it possible to automate physical labor as well as mental labor that were originally performed by humans in various sectors such as aeronautics, the automobile industry, farming and economy for instance. As a matter of fact, automation has been defined by Parasuraman and colleagues (2000) as “a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator” (*Parasuraman, Sheridan, & Wickens, 2000*). At the beginning, designers were focused on optimizing the capabilities of the machines in terms of power, speed and intelligence, ignoring the effects of application on a human operator. The main concern initially consisted in automatizing as efficiently as possible human actions at a lower cost for the purpose of economic benefits (*Kaber, Onal, & Endsley, 2000; Parasuraman & Riley, 1997*). Interestingly, several authors have progressively addressed the issue of automation through a human-centered approach, taking into account the relationship between the human and the machine (*Billings, 1997; Norman,* 1986; *Parasuraman, Sheridan, & Wickens, 2000*).

In this respect, human-machine interactions can be characterized according to a continuum of different levels of automation starting from fully manual control (i.e., the human takes all the decisions, the machine offers no assistance) to fully automated (i.e., the machine acts autonomously, overriding the human decisions) (*Endsley, 1999; Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Verplank, 1978*). Within the intermediate levels of automation, an increased and progressive autonomy is given to the machine, at four specific stages of action decision and implementation namely, (a) the acquisition of information, (b) the analysis of that information, (c) the decision about the actions to take based on that information, and (d) the implementation of that action (*Endsley, 1999; Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Parasuraman, 2006;* see **Table 1** for an example of levels of automation taxonomy by Parasuraman, Sheridan, & Wickens, 2000).

| | |
|---|---|
| HIGH | 10. The computer decides everything, acts autonomously, ignoring the human. |
| | 9. informs the human only if it, the computer, decides to |
| | 8. informs the human only if asked, or |
| | 7. executes automatically, then necessarily informs the human, and |
| | 6. allows the human a restricted time to veto before automatic execution, or |
| | 5. executes that suggestion if the human approves, or |
| | 4. suggests one alternative |
| | 3. narrows the selection down to a few, or |
| | 2. The computer offers a complete set of decision/action alternatives, or |
| LOW | 1. The computer offers no assistance: human must take all decisions and actions. |

**Table 1.** Levels of automation of decision and action selection from Parasuraman, Sheridan, & Wickens, (2000).

## 1.2. Advantages of automation: a better life

On the one hand, humans have undeniably benefited from the hegemony of machine automation. It made some aspects of life easier, safer and faster (*Kaber, Onal, & Endsley, 2000*). Easier, in the sense that technological developments made possible for individuals to go beyond their limits by allowing people with disabilities to move and communicate for example (*Grynszpan, Weiss, Perez-Diaz, & Gal, 2014*; *de Almeida Ribeiro, et al. 2013*). Safer, by liberating individuals from uncertainty and risky situations as, for instance, with the flight envelope protection function of the Airbus 320 aircraft which is a control automated system that prevents the pilot from making control commands that would force the aircraft to exceed its structural and aerodynamic limits (*Niedermeier & Lambregts, 2012*). And faster, given that automation allowed the processing a voluminous amount of data withdrawing individuals from time-consuming and thus enhancing productivity (*Kaber, Onal, & Endsley, 2000; Sarter, Woods, & Billings, 1997*).

One the other hand, however, it must be said that automation has also brought a set of negative aspects.

## 1.3. Disadvantages of automation: distancing the human

Indeed, for a long time, developers have assumed that adding automation was a simple substitution of a machine activity for human activity. This was the *substitution myth* (*Woods & Tinapple, 1999*). Rather, expanding the machine role modified the architecture of the human-machine cooperative relation, changing profoundly and unintendedly the human thinking (*Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000*; *Sarter, Woods, and Billings, 1997*). This shift from active to passive information processing led individuals to disengage themselves from the control process of the ongoing action resulting in the establishment of what Norman (1986) called the "gulf of execution". At the same time, Norman (1986) reported that automation made humans less sensitive to the feedbacks given by the machine, what he has called the "gulf of evaluation" (*Limerick, Coyle, & Moore, 2014; Norman, 1986;* see **Figure 1**). According to Norman (1986), these discrepancies between the humans and the machines may arise from divergences in form and content between the two agents. Indeed, understanding of the machine's state implies interpreting its feedbacks that are expressed in physical terms and only indirectly refer to the intention of the human operator, which is expressed in psychological terms (*Norman, 1986*).

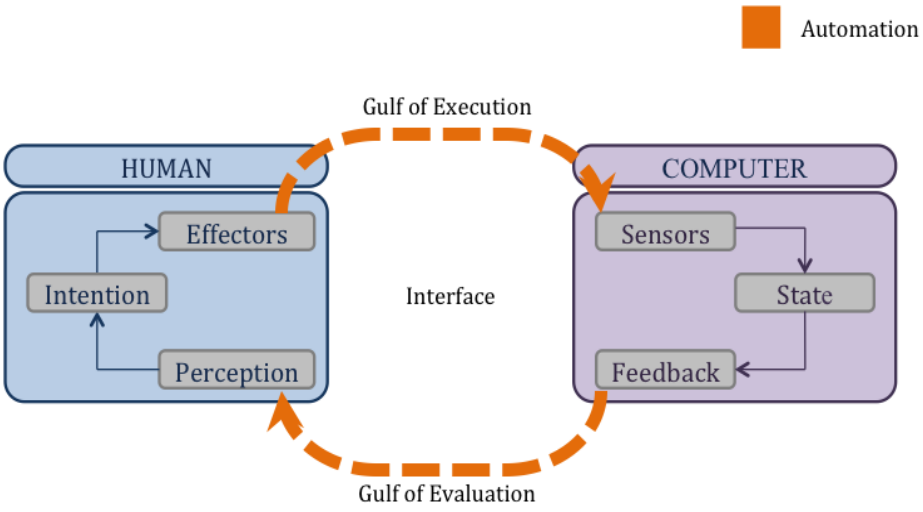


**Figure 1.** A human-machine interaction loop in regard with automation from Limerick, Coyle, & Moore (2014) and Norman (1986).

## 1.4. The out-of-the-loop (OOTL) performance problems

Specifically, allocation of higher cognitive functions to machines brought a set of human factors issues referenced under the label “out-of-the-loop (OOTL) performance problems” that are well documented. The OOTL phenomenon corresponds to a deterioration of the operator’s performance when interacting with highly automated systems. The terms “total confusion”(*Bureau d’Enquête et d’ Analyse, 2002, p.167*; *National Transport Safety Board, 1975, p.17*), “surprise effect” (*Bureau d’Enquête et d’Analyse, 2012a, p.10, 2016, p.44*) or “no awareness of the current mode of the system”(*Bureau d’Enquête et d’Analyse, 2012b, p.178*) indicate a similar process: a mental state in which the human operator has lost her/his situation awareness and is not able to monitor the system efficiently.

At an operational level, the OOTL phenomenon induces a clear performance decrease whenever trying to transfer manual control over the automated artificial system. Amongst other problems, an operator who is OOTL might take longer or be completely unable to detect an automation failure, to decide if an intervention is needed, and to find the most adequate response (*Kaber, Onal, & Endsley, 2000; Parasuraman, Molloy, & Singh, 1993*; *Parasuraman & Riley, 1997*). For example, in the automotive field, it has been showed that during a driving simulation task, individuals’ responses to critical events were faster in the manual driving condition compared to the automatic driving condition (*Merat & Jamson, 2008*).

It has been ventured that the OOTL performance problems classically observed during human-machine interactions could be attributed to underlying factors such as vigilance decrement, complacency, system opacity and loss of situation awareness (*Kaber, Onal, & Endsley, 2000*). The vigilance decrement referred to the absence of sustained attention during interaction with highly automated systems, which is however mandatory in order to detect and response to critical events (*Warm, Parasuraman, & Matthews, 2008*). Because highly automated systems led humans to endorse the role of supervisors, they shifted from an active stance to a passive one. In that respect, researches on sustained attention showed that individuals had increased detection times of about 50% after one hour on task when they had to manually detect conflicts in a simplified version of an air traffic control task (*Thackray & Touchstone, 1989*).

Together with this difficulty to maintain high levels of vigilance in time, a complacency phenomenon toward the machine has been highlighted (*Parasuraman, Molloy, & Singh, 1993*). This phenomenon is referred to overreliance on automation. For example, Lee (2006) showed that operators maintain their beliefs that automated artificial system are more competent than they actually are (*Lee, 2006*). This overreliance on automation has been showed to represent a considerable aspect of misuse that can result in decision biases and failures of supervision (*Parasuraman, Molloy, & Singh, 1993*; *Parasuraman & Riley, 1997*).

On the side of system opacity, as automation steadily advanced in complexity and sophistication, artificial systems tended to exhibit cascades and runaway chains of automatic reactions that attenuate, or even eliminate predictability and cause outsized and unpredicted events (*Taleb, 2012*). Hence, human operators faced difficulties in having a clear idea of the machine's intentions and to predict the occurrence of subsequent events.

Finally, the loss of situation awareness refers to the mental state whereby the human operators have lost an internalized model of the current state of their environment (*Endsley, 2016*). For instance, it has been pointed out that individuals' performance and understanding of the state of a navigation system during a failure period following a fully automated mode was significantly degraded compared to a failure period following a fully manual navigation task mode (*Endsley & Kiris, 1995*). In sum, several contributors to the OOTL performance problems have been put forward, explaining how automation deeply changed human thinking and behavior.

## 1.5. Recovering the human in the control loop: a key challenge

Automation technology has dramatically changed the way we interact with our environment. As mentioned earlier, automation has created an increasing distance between the human operators and the loop of control, disconnecting them from the machines. Indeed, human operators are isolated from most of the physical structures of the system, and reciprocally, automation isolates humans from the moment-to-moment operations of the machine. This interplay of relative physical and mental isolations tends to distance human operators from the results of their actions (*Obhi & Hall, 2011b*). Replacing the humans by automated tools in the control loop gives rise to several critical issues.

A first concern relates to system acceptability issues. Improving acceptance of new technologies by human operators is an important area of concern to equipment suppliers (*Horberry, Stevens, & Regan, 2014*). To be acceptable, new technologies must be reliable, efficient and useful. However, such qualities do not guarantee acceptability from the human operator. As pointed by Shneiderman and Plaisant, (2004), users "strongly desire the sense that they are in charge of the system and that the system responds to their actions" (*Shneiderman & Plaisant, 2004*). Increases in automation can seriously threat this sense of control in human operators.

A second concern raises ethical issues and in particular, the attribution of responsibility. Yet, the sense of agency is known to underpin this concept of attribution of responsibility (*Frith 2014; Haggard, 2017; Moretto, Walsh, & Haggard 2011*). Indeed, it has been argued that to be involved in the consequence of an action is a necessary condition on acting with ethics and with moral judgment (*Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006*). For example, it has been shown that decrements in individuals' sense of agency and in responsibility attribution increased antisocial behaviors (*Caspar, Christensen, Cleeremans, & Haggard, 2016*). Hence, people's choice to act ethically or not, that is, to carry out actions that are judged to be "right" or "wrong'" is shaped by their own beliefs about our involvement in the results of the action (i.e., their sense of agency). The loss of agency might therefore constitute a form of moral disengagement regarding these actions and disturb the mechanism classically used to regulate human behavior (*Bandura, 1999*). Military robots provide a vivid illustration of this ethical issue. The autonomy of these robots increases as technology progresses. Even the final decision is left to the human operator's hands, the physical and cognitive distance between the human operators and their actions is a problem, for the soldiers themselves but also for the society.

Finally, legal issues should also be taken into consideration. Indeed, interacting with highly automated systems makes the attribution of responsibility a thorny issue, as there is a certain ambiguity about who is in control (*Courgeon, Rautureau, Martin, & Grynszpan, 2014*). Therefore, the question of the human operator responsibility in case of incidents remains unclear. This is however particularly important in safety critical artificial systems and in semi-automated systems where humans in charge of supervising the tasks are held responsible in case of task failures. With the next generation of full-automated cars (e.g., the Google car project), this penal issue will become a major concern.

Consequently, keeping the human operators in the control loop, as intentional agents, appears as a fundamental challenge for the coming years. Indeed, the major question raised here was how to make human-machine interactions more collaborative?

## 1.6. The PhD thesis

Creating partially autonomous machine agents is somehow equal to adding new team members. One repercussion is the emergence of new coordination demands and the manifestation of new classes of problems (i.e., the so-called "OOTL performance problems"), which results from failures in the human-machine relationship. Indeed, it is now well established that individuals experience a loss of agency, that is to say, a reduced sense of control over their environment when interacting with artificial automated systems (*Berberian, Sarrazin, Le Blaye, & Haggard, 2012; Coyle, Moore, Kristensson, Fletcher, & Blackwell, 2012*; *Obhi & Hall, 2011b*). For example, Beberian and colleagues (2012) showed that human operators experienced a very strong decrease of their sense of agency when interacting with highly automated autopilot interfaces (*Berberian, Sarrazin, Le Blaye, & Haggard, 2012*).

The overall aim of the PhD thesis was to better characterize individuals' sense of agency during human-machine interactions, as well as to investigate how to improve such kind of interactions. With respect with the taxonomy proposed by Parasuraman and colleagues (2000), we used a level of automation corresponding to the 7$^{th}$ rank (*Parasuraman, Sheridan, & Wickens, 2000*). Indeed, in this thesis, human-machine interactions were investigated through a target detection task (Simon task, *Simon & Small, 1969*) equally distributed across a human agent and an artificial automated system. Specifically, the human agent was in charge of one specific type of target, whereas the machine was in charge of another type of target.

The first stage of the thesis sought to finely examine the cognitive processes underlying the individuals' loss of agency during joint tasks with automated artificial systems, both at the behavioral (**Experiment 1**) and at the cerebral (**Experiment 2**) levels. In **Experiment 1**, participants ran a Simon task coupled together with an intentional binding task. More specifically, participants performed the Simon task alone or with a co-agent that could be either another human or an algorithm implemented in a desktop computer; or passively observed another agent (human or desktop computer) doing the task. Accurate

target detections triggered an auditory tone after a certain delay. Participants were required to estimate the time interval between their own action or the action of their co-agent and the subsequent auditory outcome. The response times observed in the Simon task served as an index of participants' ability to represent their co-agent's action into their own motor system (*Sebanz, Knowblich, & Prinz, 2003*). Underestimation of time intervals served as an implicit measure of participants' sense of agency (*Haggard, Clark, & Kalogeras, 2002*). The results suggested that the ability to form co-representations of a co-agent's actions was in pace with the ability to develop a sense of agency for other-generated actions. Critically, while participants were able to co-represent the human co-agent's actions and develop a sense of agency for self- and other-generated action outcomes, such abilities were impaired when cooperating with the artificial system's actions. In **Experiment 2**, the previous experimental paradigm has been replicated in a simplified version. The electroencephalographic activity of the participants was recorded to analyze event-related potentials. The analysis of the auditory N1 component served as an implicit measure of participants' sense of agency for other-generated action outcomes (*Poonian, Mcfadyen, Ogden, & Cunnington, 2015*). The analysis of the P3 component during the vision of the co-agent's target served as an indicator of action co-representation phenomenon (*Sebanz, Knoblich, Prinz, & Wascher, 2006*). We did not succeed in showing N1 amplitude modulation according to the nature of the participants' co-agent (i.e., human versus machine). However, a reliable modulation of P3 component amplitude has been highlighted, suggesting that computer-generated actions were not represented into the participants' brain during the joint task, unlike the actions generated by the human co-agent.

The second stage of the thesis sought to examine on which characteristics of the machine we should act to regain the human operator's sense of agency during human-machine interactions. We began to study the influence of high-level properties such as the physical appearance of artificial systems on the individuals' sense of agency during a joint task (**Experiment 3**). Finally, we focused on the influence of low-level properties such as sensorimotor information on the individuals' sense of agency during a joint haptic task (**Experiment 4**). In **Experiment 3**, participants ran a behavioral Simon task jointly with a human or machine co-agent. We manipulated the physical appearance of the machine co-agent so that participants could cooperate either with a humanoid robot or a servomotor. Accurate target detection triggered an auditory tone after a certain delay. Participant temporal interval estimations served as an implicit measure of their sense of agency

(*Haggard, Clark, & Kalogeras, 2002*). We found that participants' sense of agency for self- and other-generated action outcomes declined during interaction with the servomotor compared to the human interactions. In addition, our findings revealed that participants' sense of agency for self- and other-generated outcomes was improved during the interactions with the human-like machine. In **Experiment 4**, participants manipulated a pair of haptic interfaces jointly with another human co-agent. Participants were told that they could receive haptic feedbacks from each other. Unknown to them, during half of the experiment, the sensorimotor feedbacks they received were artificially generated by a robotic system. An auditory tone followed participants' actions. Participants' sense of we-agency was assessed using temporal interval estimations (*Haggard, Clark, & Kalogeras, 2002*) and explicit judgments of agency. Our results showed that participants reported higher judgment of agency when they cooperated with the robot compared to the human co-agent. On the contrary, at the implicit level, participants demonstrated a sense of we-agency only during the human-human joint actions in comparison with the human-machine joint actions. Our findings suggest that participants' sense of agency was sensitive to modulations in tactile reafferences about co-agent's actions.

The findings of the four experiments are discussed in regard with the existing scientific literature on the sense of agency, and also with the operational problematic of the thesis, namely, the OOTL problem.

# Chapter 2: State of the art

## 2.1. The sense of self-agency: definition and measures

The sense of agency can be defined as the experience of controlling one's own actions, and, through them, the events in the outside world (*Haggard & Chambon, 2012*). Individuals' experience of agency for their own actions is a crucial part of normal mental life and is what allows societies to hold individuals legally and morally responsible for what they do. The human species is also characterized by its ultra-sociality and many of the most significant human achievements result from our capacity to engage in cooperative joint actions (*Richerson & Boyd, 1997*). The sense of agency experienced in joint action is thus a central subjective dimension of human sociality and also an essential aspect of human cooperativeness.

Individuals' sense of agency can be empirically assessed using explicit (*Sato & Yasuda, 2005*) or implicit measures (*Haggard, Clark, & Kalogeras, 2002; Poonian & Cunnington, 2013; Poonian, Mcfadyen, Ogden, & Cunnington, 2015*). On the one hand, the most obvious way to estimate individuals' sense of agency is to explicitly ask participants to report on Likert scales their degree of agreement with statements about their actions or their subsequent sensory consequences (*Sato & Yasuda, 2005; van der Wel, Sebanz, & Knoblich, 2012*). For example, participants can be requested to rate from 0 ("absolutely not") to 100 ("absolutely") how much they had experienced to be in control of a given performance (*van der Wel, Sebanz, & Knoblich, 2012*) or how much they felt they caused a given sensory outcome (*Sato & Yasuda, 2005*).

On the other hand, to overcome conformity biases, intentional binding (IB) and sensory attenuation phenomena have been proposed as implicit measures of individuals' sense of agency. Intentional binding phenomenon is the method the most used in the experimental area of the sense of agency. This perceptive temporal attraction between the onset of a generated action (e.g. a key press) and the onset of its sensory consequence (e.g. an auditory tone) does occur when the action has been intentionally triggered but not when the action has been passively performed (*Haggard, Clark, & Kalogeras, 2002*). Indeed, this illusory bias has been initially showed by Haggard and colleagues (2002) when they asked participants to perform keypresses. In the single-event voluntary condition, participants had to press the key at the moment of their own choosing. In the single-event involuntary condition, a transcranial magnetic stimulation (TMS) was delivered over the motor cortex of participants which produced an involuntary twitch of their hand toward the key. In the single-event sham condition, a TMS was delivered over the left parietal cortex of

the participants which generated an audible click but neither muscular activity nor abnormal perceptual experiences. In three further pairs of operant conditions, participants' keypresses triggered an auditory tone after a delay of 250 ms. Hence, in the operant voluntary condition, the sensory consequence of the action could be deliberately triggered by the participants whereas the two operant TMS conditions did not involve agency at all. Participants were asked to judge either the onset time of their keypress (or the onset of the audible click for the sham conditions), or the onset time of the tone. Participants' temporal estimations in the single-event conditions were subtracted from participants' temporal estimations in the corresponding operant conditions. The resulting time served as a measure of the shift of the action toward the tone. The results showed that the onset of the intentional action (voluntary condition) was perceived later and the onset of the subsequent auditory tone sooner compared to the corresponding events in the single-event conditions. In other words, participants perceived the voluntary keypress and its sensory consequence closer in time than they actually were. Such a binding has been considered as a cognitive signature of the brain for voluntary goal-directed action perception. Unlikely, unintentional actions (involuntary condition) were perceived pushed away from the tone and this temporal repulsion has been considered as a marker of non-intentional agentive experience (*Haggard, 2005; Haggard, Clark, & Kalogeras, 2002; see* **Figure 2**). Various methods have been used to compute IB phenomenon in studies investigating individuals' sense of agency so far. The first method consisted in presenting a rotating hand clock to participants who had to report the perceived location of the hand clock at the onset of their action or the subsequent sensory consequences (*Wohlschläger, Engbert, & Haggard, 2003*). The main advantage of this method is that it permits to distinguish the *action* binding effect from the *outcome* binding effect as each might rely on distinct underlying mechanisms (*Wolpe, Haggard, Sibener, & Rowe, 2013*). However, this method has been thrown out progressively considering that it was inducing an attentional shift toward the clock. Another method consisted in reporting the perceived delay between the action and its following sensory consequence by writing (*Moore, Teufel, Subramaniam, Davis, & Fletcher, 2013*) or by verbal judgment (*Wohlschläger, Haggard, Gesierich, & Prinz, 2003*), or by reproducing the duration of the perceived interval (e.g. pressing the space bar as long as the perceived temporal delay between the two events, *Poonian & Cunnington, 2013*).

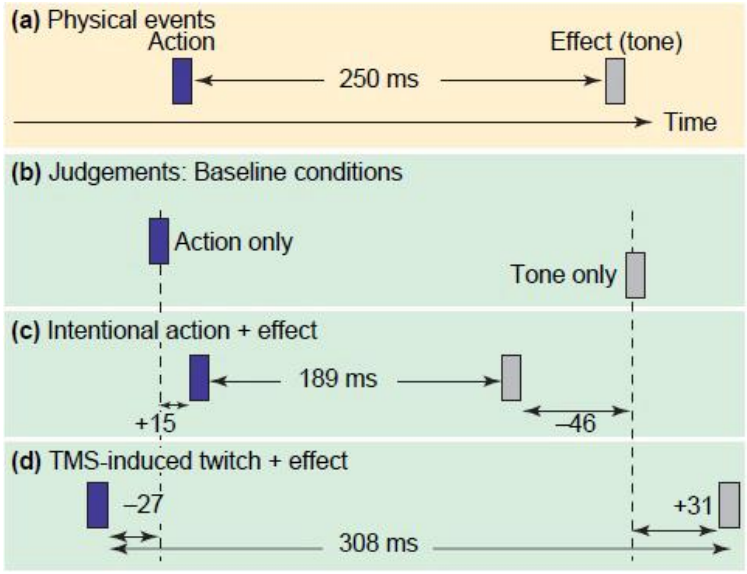


**Figure 2.** Intentional binding phenomenon between self-generated actions and their following sensory effects, from Haggard (2005).

Another well-known implicit measure of individuals' sense of agency is the sensory attenuation phenomenon (*Poonian, Mcfadyen, Ogden, & Cunnington, 2015; Weiss, Herwig, & Schütz-Bosbach, 2011*). The theory behind is that because the outcome of a voluntary self-generated action can be better predicted (through the internal forward model of the agent's motor system) compared to an external generated effect, the salience of self-generated sensory effects is diminished in the former case (*Blakemore, Wolpert, & Frith, 2000*). For example, Weiss and colleagues (2011) asked participants to perform keypresses that triggered an auditory tone after a certain variable delay or to observe the experimenter making the same actions that triggered the same sensory consequences, or to listen to the very same tone that has been automatically triggered by a computer. Participants were asked to compare the loudness of the tones. The results showed that self-generated auditory tones induced a reduced perception of loudness intensity in comparison with external generated auditory tones (*Weiss, Herwig, & Schütz-Bosbach, 2011*). Likewise, at the cerebral level, it has been showed that the amplitude of the N1 auditory cerebral component was diminished during the perception of self-generated auditory tones compared to automatic generated tones (*Poonian, Mcfadyen, Ogden, & Cunnington, 2015*). Enthusiastically, many authors from philosophers to cognitive scientists have tried to understand how one's sense of agency was built when individuals performed self-generated actions.

## 2.2. How does individuals' sense of agency is build?

### 2.2.1. Internal cues for the sense of agency

Considering the *predictive* framework, individuals' sense of agency is a personal experience that stems from sensorimotor cues (*Chambon, Sidarus, & Haggard, 2014; Frith, 2005; Frith, Blakemore, & Wolpert, 2000; Gallagher, 2000; Sato & Yasuda, 2005*). Indeed, strong links have been made between (a) the comparison processes advanced by computational model of action control and the sense of agency (*Frith, Blakemore, & Wolpert, 2000; Gallagher, 2000; Sato & Yasuda, 2005*) and (b) the fluency of action selection and the sense of agency (*Chambon & Haggard, 2012*; *Sidarus & Haggard, 2016; Sidarus, Vuorre, Haggard, 2017*).

According to the Comparator Model (**Figure 3**, CM, *Frith, Blakemore, & Wolpert, 2000*), once a goal-directed action has been selected by an agent, controllers receive information about the estimated actual state of the body and generate the appropriate motor command in accordance with the desired state of the body. Before the motor command triggers the action, a copy of the motor command, called an efference copy, is made and sent to a set of predictors through a forward model that compute the predicted consequences of the action. Then, this up-coming predicted state is compared to the estimated actual state and the quality of the matching between these two states has been proposed to retrospectively (that is to say, after action execution) informs the agent's sense of agency (*Frith, 2005, Sato & Yasuda, 2005*). More specifically, little spatio-temporal discrepancies between the predicted state of the motor system and its estimated actual state have been showed to generate a strong sense of agency whereas mismatches will damage this sense of agency (*Frith, 2005; Sato & Yasuda, 2005*). For example, Sato & Yasuda (2005) showed that participants exhibited a lower sense of agency when a self-triggered effect did not match with a given intentional action. In addition, participants misattributed self-generated sensory effects to an external source when an unexpected longer temporal delay followed their actions (*Sato and Yasuda, 2005*). In the first experiment of their study, Sato and Yasuda (2005) asked participants to press a right key that triggered a specific auditory tone (e.g., 1000 Hz) after being pressed on, or a left key that triggered a different auditory tone (e.g., 600 Hz) after being pressed on. This served as a learning session to make the participants associate one specific keypress with one given

sensory consequence. In the testing phase, participants where induced the false thought that the auditory tones could be generated by themselves or by the experimenter, but in fact, the tones were always participant-generated. In the congruent tone condition, participants had to perform right or left keypresses and their actions evoked the same tones as during the learning session. In the incongruent tone condition, participants had to perform right or left keypresses but the action/tone mapping was the opposite of the mapping used in the learning session. In this case, the sensory consequences of their actions were different from prediction. The authors also induced temporal disruptions in the testing phase so that the tones could be generated either immediately, 200 ms, 400 ms or 600 ms after the participants' keypresses. In the control condition, participants had to passively listen to the two tones. After each trial, participants had to rate from 0 ("totally disagree") to 100 ("totally agree") two items. The first item was "I was the one who produced the tone" and aimed at investigate the participants' sense of agency. The second item was "I was the one who was listening to the tone" and aimed at assess participants' sense of ownership. The results showed that in the congruent tone condition, when the predicted sensory consequences matched the estimated actual sensory consequence, participants' sense of agency ratings were stronger than in the incongruent tone condition, when there was a mismatch. In addition, participants' sense of agency ratings were stronger when the tone occurred immediately after the key press, as during the learning session, compared to when it appeared after an unpredicted delay. The participants' sense of ownership ratings were always at its maximum. Hence, the results of this first experiment showed that inconsistencies in space or in time between a predicted sensory consequence and an estimated actual sensory feedback impaired explicit participants' sense of agency. In a second experiment, the authors asked participants to perform a similar paradigm as the first experiment except that they analyzed separately the "self-" and the "other-" generated auditory tones. Interestingly, the authors made the "other-" generated effects corresponds to participants' predictions based on their own performance during the learning session. After each trial, participants had to rate from 0 ("totally disagree") to 100 ("totally agree") the same two items as in experiment 1. The results revealed that participants reported a strong explicit sense of agency when the other-generated estimated real state fitted in space or in time with their predicted state that has been computed by their forward model. Taken together, these findings show that individuals misattribute self-generated sensory effects to an external source when an unexpected longer temporal delay follows their action and are able to experience a sense of self-agency when other-generated sensory effects matched

their predictions made by their internal model. Finally, in a third experiment, Sato and Yasuda (2005) asked participants to perform an Eriksen flanker task wherein a five-letter array was presented on a screen and participants had to respond according to the identity of the central letter. When the central letter was an "H", participants had to press a specific key (e.g. a right key) and when the central letter was a "N", participants had to press another specific key (e.g. a left key). The surrounding distractor letters could be similar to (e.g. "HHHHH") or different from (e.g. "NNHNN") the central target. The aim of this task was to induce errors in participants' behavior due to the potential activation of an inconsistent action elicited by the presence of the surrounding distractors. Participants' responses triggered an auditory tone that could be congruent or incongruent with a previous action/tone learning session, and that could occur immediately or 400 ms after the keypress. After each trial, participants were asked to report whether their preceding response was correct or not and to rate from 0 ("totally disagree") to 100 ("totally agree") the same two items as in experiment 1. The authors analyzed the correct trials and those that participants correctly judged as errors (unnoticed errors were not taken into consideration). The authors found that the sense of self-agency was reduced when there were spatio-temporal discrepancies between the predicted and the actual sensory consequence, irrespective of whether participants had given a correct response or made an error. Indeed, in both trial types, they reported a stronger explicit sense of agency when the tone was congruent to the learning session compared to when it was incongruent, and when the tone was presented immediately after the keypress compared to when the occurrence of the tone was delayed. Hence, even when participants performed an unintended action, the better the prediction made on their motor command matched the actual sensory consequence, the stronger the sense of agency they experienced (*Sato & Yasuda, 2005*).

This suggests that individuals' sense of agency is crucially dependent on the retrospective comparison between a predicted state from an internal motor forward model and the actual consequence of this action, reinforcing the involvement of the CM in the building of one's sense of agency.

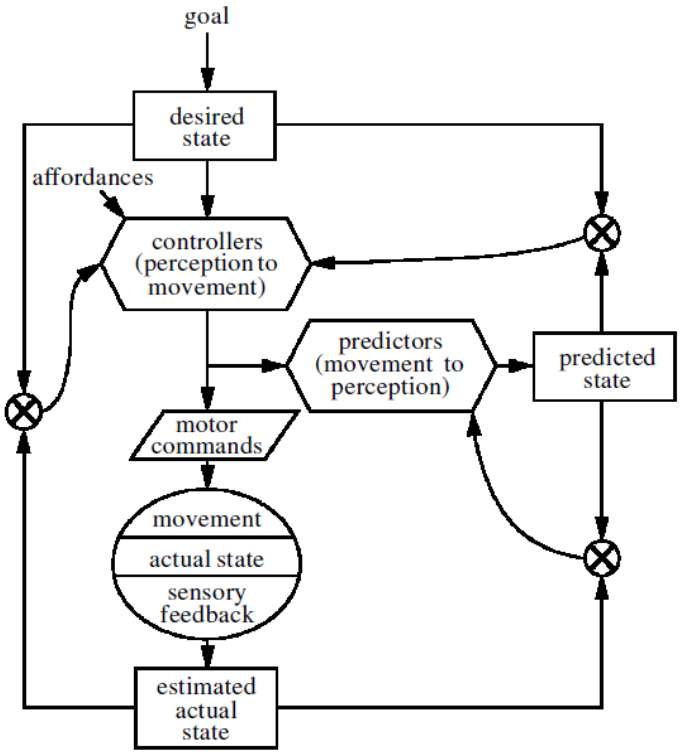


**Figure 3.** The Comparator Model (CM) from Frith and colleagues (2000).

Furthermore, the fluency of action selection has also been proposed to inform the sense of agency, but, unlike the CM, in a prospective manner given that this process was assumed to occur before action execution (*Chambon & Haggard, 2012*; *Sidarus & Haggard, 2016; Sidarus, Vuorre, Haggard, 2017*). Indeed, there was evidence showing that when action selection was easy, individuals' sense of agency became stronger compared to conflictual action selection (*Chambon & Haggard, 2012*; *Sidarus & Haggard, 2016; Sidarus, Vuorre, Haggard, 2017*). For instance, Sidarus and Haggard (2016) asked participants to perform an Eriksen flanker task wherein a five-letter array was presented on a screen and participants had to respond to the identity of the central letter. The surrounded distractors could be congruent (e.g. "HHHHH"), incongruent ("SSHSS") or neutral ("OOHOO") from the central target. When the central letter was an "H", participants had to press a specific key (e.g. a right key) and when the central letter was an "S", participants had to press another specific key (e.g. a left key). Participants' keypresses triggered the apparition of a colored dot after a certain delay. At the end of each trial, participants were asked to rate on a 9-point Likert scale how much control they felt over the colored circles that were triggered by their actions. The results showed that when the participants' responses were facilitated by supraliminal congruent cues, participants reported a stronger explicit judgment of control

compared to when their responses were disturbed by supraliminal incongruent cues (*Sidarus & Haggard, 2016*). In the same vein, it has been showed that subliminal congruent priming allowing easy action selection for a subsequent target detection enhances individuals' sense of agency in comparison with a subliminal incongruent priming (*Sidarus, Vuorre, & Haggard, 2017*). Hence, when action selection is facilitated by congruent cues, individuals' sense of agency becomes stronger compared to when action selection is damaged by incongruent cues (*Chambon, Sidarus & Haggard, 2014; Sidarus, Vuorre, & Haggard, 2017*).

In sum, there is a certain amount of empirical data showing that intrinsic predictive sensorimotor information helps to shape individuals' sense of agency. Conversely, when considering the *inferential* account, individuals' sense of agency rather originates from extrinsic circumstantial cues (*Aarts, Custer, & Wegner, 2005; Caspar, Christensen, Cleeremans, & Haggard, 2016; Desantis, Roussel, & Waszak, 2011*; *Wegner & Wheatley, 1999*).

### 2.2.2. External cues for the sense of agency

Consistent with the *inferential* account, it has been shown that the induction of prior false thoughts about a performed action or its subsequent sensory consequence could change individuals' sense of agency (*Desantis, Roussel, & Waszak, 2011*). For example, Desantis and colleagues (2011) asked participants to perform keypresses together with a confederate. The keypresses triggered an auditory tone after a variable delay of 350 ms, 550 ms or 750 ms. The co-agents were separated from each other with the help a card board to avoid participants seeing what the confederate was doing. In the subject condition, the names of the participants were displayed on a screen in front of them to make them believe that the tone was generated by their own keypress. In the confederate condition, the name of the confederate was displayed on the screen to make them believe that the tone was generated by the confederate keypress. In the ambiguous condition, a sequence of "######" was displayed on the screen so that they believed they may or not may trigger the tone. In fact, in all experimental conditions, the tones were always triggered by the participants' keypresses. For some trials of the operant conditions, participants were asked to report the position of a clock-hand at the onset time of their keypress of the believed confederate keypress, and for the other trials, the position of the clock-hand at the onset

time of the tone. In the baseline conditions, participants had to make a keypress that did not trigger any subsequent tone, or to listen to an auditory tone externally generated via a loudspeaker. Participants had to judge the position of a clock-hand at the onset time of the given event occurred. The results indicated that when participants had to judge the onset time of the action, they made anticipatory temporal estimations compared to the baseline condition in a similar fashion for the believed self-generated actions (subject condition) and the believed other-generated action (confederate condition). However, the authors found a temporal anticipation in the participants' time reports for the tone onset when they believed they self-triggered the tone (subject condition) compared to when the tone was externally generated without no intentional action (baseline condition). When participants believed the tone was generated by the confederate (confederate condition), there was no such anticipatory temporal estimation. In the ambiguous condition, both the onset time of the action and the onset time of the tone were anticipated compared to the baselines as for the subject condition, suggesting an egocentric bias in attribution of action in ambiguous context (*Desantis, Roussel, & Waszak, 2011*). Hence, induced belief about an action can influence the temporal perception of its subsequent sensory consequences. Indeed, when individuals are convinced they produced a sensory effect, they exhibit an implicit sense of agency as indicated by IB phenomenon. Unlikely, they do not manifest such a temporal attraction when they are convinced the sensory effect has been triggered by another agent even though they were the effective generator of this effect.

In addition, it has been showed that the social context of voluntary actions can modulate individuals' sense of agency (*Caspar, Christensen, Cleeremans, & Haggard, 2016; Obhi & Hall, 2011b*). For example, Caspar and colleagues (2016) showed that voluntary actions such as painful electric shock administrations that have been executed under coercion were experienced in the same weak way that passive actions in term of agency (*Caspar, Christensen, Cleeremans, & Haggard, 2016*). In this experiment, female participants were paired, one was attributed the role of the "agent" and the other the role of the "victim". The agent was asked to perform a keypress under various experimental conditions, which triggered an auditory tone after a delay of 200 ms, 500 ms or 800 ms. In the active control condition, the "agent" has to press a given key whenever she wanted. In the passive control condition, the "agent" has her finger laid down on a given key and the experimenter pressed the "agent" 's finger on the key. In the two following experimental conditions, the "agent" had to press one between two specific keys that lead to different consequences. Indeed, one given keypress (e.g. the F key) made her earn money while it

made the "victim" lose the same amount money or it delivered a painful electric shock to the "victim". On the contrary, the other given keypress (e.g. the H key) made the "agent" refrain from tacking money or deliver an electric shock but she did not earn money. In the coercive condition, the experiment was next to the "agent" and ordered her to press one specific key. In the free-choice condition, the "agent" could freely press the key of her own-choosing. In all the four experimental conditions, both the "agent" and the "victim" had to report their perceived duration between the onset of the keypress and the onset of the auditory tone. The authors found two main results. Firstly, the participants' action-tone interval estimations were shorter in the free-choice condition compared to the coercion condition. Secondly, there were no differences in the participants' temporal estimations in the passive control condition and the coercion condition. Consistently, when the authors recorded the cerebral activity of the participants, they found that N1 amplitude was weaker during the perception of the tone in the coercion condition than in the free-choice condition, interpreted as a distancing in the processing of outcomes self-generated under authority (*Caspar, Christensen, Cleeremans, & Haggard, 2016*).

Accordingly, Obhi and Hall (2011b) showed that individuals' sense of agency for self-generated actions performed in a joint task was dependent of prior thoughts about the nature of a co-agent. In their study, participants were separated by a curtain from either a confederate (presented as a genuine participant) or a desktop computer. Hence, there was no direct visual information but the participants had in mind that they were partnered with another human or a machine. In the baseline action alone condition, participants had to tap a touchpad at the time of their choice and estimate the onset time of their action with the help of a clock. In the baseline tone alone condition, they had to estimate the onset time of an externally generated auditory tone. Then, in the operant condition, participants were asked to tap the touchpad at the time of their choice, which triggered an auditory tone after 200 ms. In the operant condition participants thought that the confederate and the computer could also cause the tone if they had tapped the touchpad before them. In fact, the tone was always generated by the participants' actions. After each trial, a randomized false feedback about the initiator's identity was given to the participants (self, other or indistinguishable). Afterward, they had to judge the onset time of their own action or the onset time of the tone with the help of a clock. In addition, participants had to explicitly report their belief about who had caused the tone. The results revealed that the onset time of the action was perceived later in the operant condition (i.e., closer to its following sensory consequence) than it was when the action did not cause any effect but only when the

participants thought they were running the experiment with another human participant. Indeed, the authors did not found such a difference when the participants thought they were running the experiment with a computer. Similarly, the perceived onset time of the tone was perceived earlier in the operant condition (i.e., closer to the action) compared to baseline tone alone condition, but only when the participants performed the experiment with the confederate and not with the computer. Interestingly, the participants' belief about the initiator of the action always matched the given false feedback. Thus regardless participants were convinced that the tone was generated by their own action of their human co-agent, they manifested in both cases a temporal attraction between their action and its effect. It was not the case when their co-agent was a desktop computer: they exhibited no intentional binding both for their own action and the machine-generated action. Hence, individuals can experience a reduced agency for self-generated action when they are induced the though they are cooperating with an artificial system.

In the same vein, Wegner and Wheatley (1999) showed that participants could report an illusory sense of agency for a given action outcome after prior thought induction even though they did not actually cause the outcome (*Wegner & Wheatley, 1999*). In the authors' study, participants were paired with a confederate presented as a naïve participant. Participants were asked to move a mouse jointly with the confederate in order to displace a cursor displayed on a screen where pictures of several objects were displayed. Participants were wearing headphones and could hear the name of a target object at a certain time. The confederate has as instructions to stop the moving of the mouse (thereby forcing the participants' moving) when participants heard the name of the target object. Participants were not aware about the trick. Participants had to rate from 0 ("not at all") to 100 ("totally") how much they intended to make the stop. Wegner & Wheatley (1999) found that the more the temporal delay between the target object participants heard and the stop of the participant was short, the more participants reported a strong will to stop the cursor even though, in facts, they always followed passively the confederate intentions and actions (*Wegner & Wheatley, 1999*).

There is thus evidence that high-level inferences and prior thoughts contribute to individuals' sense of agency.

### 2.2.3. Multifactorial integration for the sense of agency

At the end, it is now well accepted that individuals' sense of agency can benefit from the dual contribution of sensorimotor cues (feed-forward cues, proprioception and sensory feedback) and circumstantial cues (intentions, thoughts and contextual cues) (*Moore & Fletcher, 2012; Synofzik, Vosgerau, & Newen, 2008; Synofzik, Vosgerau, & Newen, 2008*). Indeed, by considering a Bayesian approach, Moore & Fletcher (2012) claimed that individuals' sense of agency emerged from the dynamic integration of multifactorial internal and external cues (*Moore & Fletcher, 2012*). Moore and Fletcher (2012) proposed that the influence of each cue in the construction of the sense of agency was dependent of its reliability, with a higher weighting for internal motor cues by default. However, the authors admitted that external contextual cues could gain ascendancy when sensorimotor information became less reliable (*Moore & Fletcher, 2012*).

This multifactorial account is supported by empirical evidence (*Moore, Wegner & Haggard, 2009*). For example, Moore and colleagues (2009) asked participants to listen to an auditory tone (e.g., 600 Hz), which served as a prime to induce prior thought about a forthcoming action. Thereafter, in the voluntary condition, participants had to perform a voluntary keypress. In the involuntary condition, participants experienced an equivalent involuntary movement that was induced by a mechatronic device or by the experimenter's hand. All keypresses, voluntary or not, were followed by a second auditory tone after a variable delay of 100 ms, 400 ms or 700 ms. This second tone could be congruent (e.g., 600 Hz) or incongruent (e.g., 1000 Hz) with the tone presented at the beginning of the trial. Participants were asked to verbally report their perceive duration between the onset of the keypress and the onset of the subsequent tone. The authors found that participants' temporal estimations were shorter when the second tone was congruent with the prime compared to incongruent priming, suggesting that congruent prior though about an action outcome enhances the sense of agency. Interestingly, the type of keypress participants performed modulated this effect of prime congruency. Indeed, the benefit of the congruent priming was stronger when participants experienced involuntary keypresses compared to when they performed intentional keypresses (*Moore, Wegner & Haggard, 2009*). These findings support the idea that when sensorimotor information is available, contextual cueing is negligible. However, in the absence of motor command, when internal cueing is weak, contextual cues such as induced primes play a role in individuals' sense of agency. Hence,

congruent priming of sensory effect generated by involuntary movements can make individuals to exhibit an illusory sense of agency.

## 2.3. Intentionality as the keystone of the sense of agency

Critically, when considering individuals' sense of agency, some components of action have to be taken into consideration and in particular, the intentional stance of the performed action. Indeed, it has been claimed that the integrity of the intention-action-effect chain should not be compromised to insure individuals experienced a sense of agency. Supporting this view, Pacherie (2007) proposed that individuals' sense of agency involves three main components: the experience of intentional causation, the sense of initiation and the sense of control. These three components are for their parts sustained by three levels of action specification, namely the future-directed intention (F-intention), the present-directed intention (P-intention) and the motor intention (M-intention) levels. The F-intention is conceptual and determined before the action onset and refers to the action as a whole unit (e.g. to have a drink). Therefore, it informs individuals' that they are the agent of a forthcoming action, given them a conscious experience of intentional causation. The P-intention is more pragmatic as it immediately defines action implementation in time and in space considering the context of the action (e.g. to grasp the glass of water put on the table in front of me). Therefore, it allows individuals to have a conscious sense of initiation of the action. Finally, the M-intention is very short-lasting and aims at specify the sensorimotor information of the selected motor program. The automatic adjustments and corrections that operate at this level, together with the intentional corrections made at the level of M-intentions support the individuals' sense of control over the generated-action (*Pacherie, 2007*). In sum, the experience of intentional causation, the sense of initiation and the sense of control are supported by three levels of intention that participate in the genesis of individuals' sense of agency.

Furthermore, humans rarely act in isolation. As social agents, we often interact with our conspecifics. Indeed, we are engaged in a wide range of cooperative tasks wherein actions are intentionally produced together with one or more other actors. Unlike individuals actions, joint tasks require individuals to make additional motor plans as they need to take into consideration the co-agent' intention and to build a shared intention to accomplish the joint goal (*Sebanz, Bekkering, & Knoblich, 2006; Tomasello, 2009*). In this respect, a large

body of studies suggests that during human interactions, individuals are able to co-represent their co-agent's actions into their own motor system (*Sebanz, Knoblich & Prinz, 2003*).

## 2.4. Shared representation for self- and other- generated human actions

### 2.4.1. Co-representation of human-generated actions

Action co-representation of human-generated actions has been nicely demonstrated by the *Social* Simon Effect (SSE, *Sebanz, Knoblich & Prinz, 2003*). In the classical Simon effect (*Simon & Wolf, 1963*), participants' response time to the presentation of a target stimulus decreases when the target is presented in the same relative spatial location as the response. For instance, in one version of this paradigm where participants were required to execute a left hand action as soon as a red target was presented and a right hand action as soon as a green target was presented. Red/green targets were displayed in either the same relative location as the participant's response (e.g., the target was presented to the left of a central fixation and participants have to perform a left key press) or in the opposite location (e.g., participants had to press the right key but the target was presented on the left). It has been shown that people were faster and more accurate when the target was on the same side as the response compared to when it was presented on the opposite side. In the go/no-go version of the task, that is to say when participants had to respond to only one target (e.g., respond only to the green target with the right hand), the interference disappeared. This result suggested that the Stimulus-Response congruency effect observed in the classical Simon task derived from the cognitive interference caused by two different action representations that were concurrently activated (*Simon & Wolf, 1963*). However, interestingly, Sebanz and colleagues (2003) showed that when participants performed the same go/no-go task with another agent (joint Simon task), and each of them was required to respond to only one target (i.e., the participant responded only to the green target, and the partner only to the red target), the interference effect for the incongruent Stimulus-Response key mapping reappeared. This suggested that the co-agent's actions were functionally represented in the participants' motor system and interfere with their own performance. Hence, it has been proposed that during a joint task, individuals integrate the behavior of their partner into their own motor scheme even though it is not task relevant (*Sebanz, Knoblich & Prinz, 2003*). Note should be taken that the "social" component of the

SSE has been criticized (*Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz, et al., 2011*; *Dolk, Hommel, Prinz, & Liepelt, 2013*). Instead, a referential-coding approach has been proposed according to which salient events, such as a physical presence for example – whatever social or not – served as a spatial referential for action coding (*Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz, et al., 2011*; *Dolk, Hommel, Prinz, & Liepelt, 2013*).

Action co-representation might be sustained by the human "mirror neuron system" (MNS) - mainly composed by the superior temporal sulcus and fronto-central connections - (*Keysers & Gazzola, 2009; Rizzolatti & Craighero, 2004*). Consistently, neuroimaging data showed that when participants observed or imagined a conspecific performing an action, their premotor cortex was activated – though to a lesser extent than during action execution – allowing action understanding and prediction with the help of the participants' own motor expertise (*Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005*; *Kilner, Friston, & Frith, 2007*; *Pineda, Allison, & Vankov, 2000*).

The ability to co-represent other-generated actions into one own motor system could be thought to allow individuals to extract relevant information (e.g. action intention) that would shape their sense of vicarious agency for other-generated action outcomes. According to the predictive coding account, internal model involved during action execution are also involved during observation of others acting and permit understanding of other-generated actions at different levels (see **Figure 4**, *Kilner, Friston, & Frith, 2007; Picard & Friston, 2014*). Indeed, during observation of other-generated actions, low-level visual information is translated into high-level representation of the goal underpinning the observed action through the superior temporal sulcus and fronto-central areas (*prediction 1*). Then, based on this prior inference about the goal of the observed action, one can predict the motor command of the observed action (*prediction 2*). Based on this motor command prediction, one can predict the kinematics of the observed person thanks to one's own motor system (*prediction 3*). Finally, the comparison between the predicted kinematics and the actual observed kinematics will produce a prediction error improving the initial inferred representation of the other's motor command (*prediction 2*) yielding a more accurate motor command estimate (*prediction 4*). Likewise, the comparison between the prior predicted motor command (*prediction 2*) and the new improved motor command estimate (*prediction 4*) would improve the inferred goal estimate. Thus, one can infer the causes of an observed other-generated action by minimizing the prediction errors at the intention, goal, motor and kinematic levels.

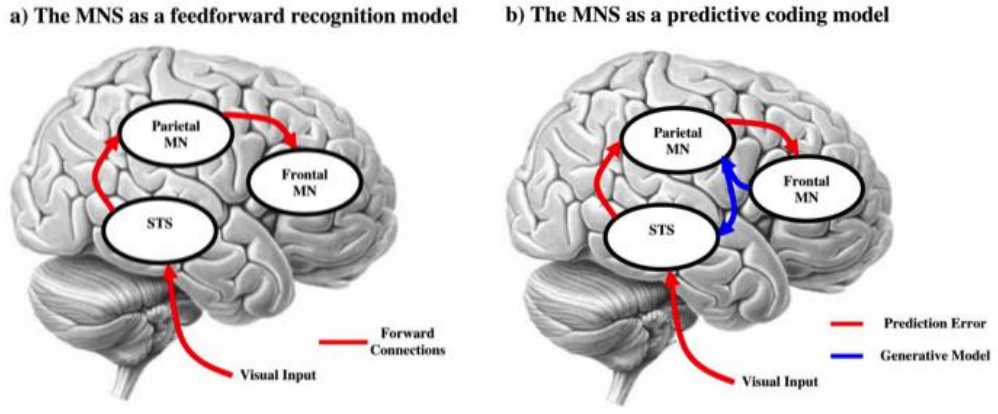


**Figure 4.** Individuals' "mirror neuron system" (MNS) for (a) feedforward action recognition and (b) action inference or predictive coding, from Kilner, Friston, and Frith, (2007).

Such action simulation might sustain individuals' sense of vicarious agency by allowing intention and action understanding. Indeed, it has been proposed that the motor mechanisms that are involved in the sense of agency during self-generated actions are of the same kind as those underlying the sense of agency during joint action with human peers (*Obhi & Hall, 2011b; Pacherie, 2012*).

### 2.4.2. A sense of we-agency for human-generated actions

Interestingly, there are more and more studies suggesting that sense of agency experienced by individual agents was not self-specific and could also occur for other-generated actions in the context of joint tasks (*Dewey, Pacherie & Knoblich, 2014; Obhi & Hall, 2011a*). In Obhi and Hall (2011a)'s study, paired participants were asked to act jointly by pressing a spacebar at will. In the first experiment, whenever one person pressed the spacebar first, the other had to also press the spacebar as soon as possible. The first key press triggered an auditory tone after a delay of 200 ms. In this setting, both participants co-intended to trigger the sensory consequence. In the second experiment, one participant was instructed to press a spacebar at the time of his or her choice while the other had to press the spacebar as soon as possible after the initiator's key press. In this context, the sensory consequence was triggered by a personal intention. In both experiments, participants were asked to report their feeling of causal responsibility (i.e., a subjective

experience of agency) using a percentage scale and also to judge the onset time of the first key press or the onset of the tone. Interestingly, in both experiments, although only the initiator reported a reliable feeling of causal responsibility, both individuals demonstrated IB. This finding has been explained by the spontaneous and pre-reflexive "we-identity" formation occurring when two humans cooperate (*Searle, 1983*; *Crivelli and Balconi, 2010*). This new agentive identity has been proposed to allow individuals to experience agency as soon as one of the two had performed an action directed at their joint goal.

Similarly, Dewey and colleagues (2004) reported the existence of a vicarious agency when participants performed a tracking task jointly with another human. The authors asked paired participants to track a target moving on a horizontal axis on a screen with a joystick (each participant had her or his own joystick). Each participant could control only one specific direction of the tracker, that is to say, to the right or to the left, so that the contributions of both agents were required to succeed on the task. On some trials, either one or both of the participants' joysticks were turned off and/or noisy perturbations were induced on the joystick motion. After each trial, participants had to give their explicit judgment about how much their own action contributed to the outcome with the help of a 9-point Likert scale from 1 ("no control") to 9 ("complete control"). The authors found that the participants' judgments of control where highest when both joysticks were activated and the noise was turned off. This could be interpreted either as evidence that in these joint tasks participants evaluate their control from the perspective of the team rather than from their own egocentric perspective (joint control hypothesis) or as evidence of a self-serving bias, where participants attribute more control to themselves when the action is successful. To adjudicate between these two possibilities, Dewey and colleagues (2014) repeated the experiment, but asked participants to rate their own control on half of the trials, and their co-agent's control on the other half of the trials. Results indicated that both self and other judgments were highest in the condition where the two joysticks were activated and the noise was off. This indicates that collaborative actions, where co-agents make complementary contributions and the action effects produced by the co-agent are predictable, contribute to a sense of "we-ness," where participants experience a shared sense of control over the joint task (*Dewey, Pacherie & Knoblich, 2014*).

Finally, van der Wel and colleagues (2012) found that participants had a general bias to claim more control over the joint action than they objectively had. In their study, the authors asked participants to pull cords attached on each side of a pole to move it back and

forth, either individually or jointly with another participant. Participants were also asked to rate from 0 (“no control at all”) to 100 (“complete control”) how strongly they had experienced to be in control during the preceding performance. The authors found that there was no relation between the force exerted by the participants themselves and their explicit rating of agency, regardless of whether they performed the task individually or jointly with another person. More interestingly, the degree of control reported by the participants was similar whether they performed the task alone or with a co-agent (*van der Wel, Sebanz, & Knoblich, 2012*). Consistently, other investigations have shown that during joint actions where participants had asymmetric roles, being a leader or a follower did not modulate the individuals’ explicit judgments of control provided that the final goal was equally shared and not imposed by the leader (*van der Wel, 2015*). Taken together, these different studies seem indicate a modulation of the self-agency toward a new agentive identity, a “we-agency”, during social joint action context. However, nowadays, interactions with others do not only involve human peers but also automated systems. Therefore, cognitivists have been progressively interested in better understanding how humans behave during their interactions with machines.

## 2.5. Cognitive experience of machine-generated actions

So far, studies have involved a large variety of different automated systems with varying complexity, from algorithms in desktop computers to human-like machines. Yet, while some researchers took an interest in humanizing the external appearance of the machine (*Wohlschläger, Haggard, Gesierich, & Prinz, 2003*; *Riek, Rabinowitch, Chakrabarti, & Robinson, 2009*), others have focused on its motor characteristics (*Kajikawa and Ishikawa, 2000*; *Glasauer, Huber, Basili, Knoll, & Brandt, 2010*). Consequently, in this part of the review of the literature, we will distinct non-human like artificial systems from human-like machines.

### 2.5.1. Representation of machine-generated actions

a) Representation of traditional (non-human-like) machine-generated actions

Several authors investigated whether action co-representation occurred during cooperative tasks with artificial automated systems using the joint Simon task (*Wen &*

*Hsieh, 2015*). Recall that the presence of a congruency effect on participants' response times would suggest that the machine co-agent's actions were co-represented into the participants' motor system during the joint task. Hence, Wen and Hsieh (2015) showed a congruency effect during a joint Simon task when participants thought they were interacting with an unseen desktop computer located in another room (*Wen & Hsieh, 2015*). This contrasted with other investigations demonstrating that action co-representation seems to be impaired during joint tasks with traditional machines. Indeed, using a similar paradigm, Tsai and colleagues (2008) found no congruency effect during a joint Simon task with an unseen computer, whether at the behavioral or the cerebral level, suggesting that action co-representation is linked with biological actions (*Tsai, Kuo, Hung, & Tzeng, 2008*).

Consistently, investigations on motor expertise revealed that cerebral "mirroring" seemed not to occur during the observation of actions generated by non-human-like machines (*Mann, Williams, Ward, & Janelle, 2007*). For example, Mann and colleagues (2007) showed that expert tennis players had faster reaction times than novice tennis players when they were playing with another human but they did not benefit from their experience anymore when they were playing against a cloaked ball machine. Consequently, this result has been explained in regard with participant's inability to simulate artificial system's behaviors in order to anticipate the final trajectory of the ball. On the contrary, participants were able to do so when partnered with another human being (*Mann, Williams, Ward, & Janelle, 2007*).

However, the issue of whether machine-generated actions could elicit individual action/observation matching system is this is still in debate. For example, Gozzola and colleagues (2007) showed that the human action/observation matching system could be activated during the observation of robotic actions that did not follow biological motion laws (*Gazzola, Rizzolatti, Wicker & Keyser, 2007*). In the authors' fMRI study, participants were shown videos of an agent's arm reaching and grasping familiar objects or just performing non-goal-directed movements without any object. The agent's arm could be either a human arm with biological motion laws or a non-humanoid robotic arm with non-biological robotic motion laws. The authors found that the participants' brain was similarly solicited during the human and the robotic goal-directed actions suggesting that the kinematic properties of the agent's action did not matter and that robotic actions could also stimulate the action/observation matching system of the observer. Similarly, a TMS study by Craighero and colleagues (2016) showed the involvement of the motor system during the passive

observation of simplistic hand action representations (using a point light display), both when the kinematics were biological and non-biological, with no difference between the two conditions, and even though the stimuli were not perceived as hands (*Craighero, Jacono, & Mele, 2016*). Hence, at the cerebral level, there is evidence that the human action/observation matching system seems to be sensitive to artificial human-like motions. Interestingly, current research tends to show that action generated by human-like machines could involve the same cognitive processes as during human-human interactions.

b) Representation of human-like machine-generated actions

Action co-representation has also been investigated at the behavioral level during joint task with humanoid robots (*Stenzel, Chinellato, Tirado Bou, & del Pobil, 2012*). Indeed, Stenzel and colleagues (2012) asked participants to perform a joint Simon task or with a robot with high level of human-like physical appearance (i.e., a full humanoid robot with a torso, an anthropomorphic head, eyes, and two arms with fingers). In the intentional condition, the robot was introduced as an intelligent and active agent who can explore the environment and act deliberately. In the unintentional condition, the robot was introduced as a passive machine acting in a deterministic way. The results showed a congruency effect on participants' response times only in the intentional condition, suggesting that co-representation of other-generated actions can also occur during joint actions with automated artificial agents, and not only with human peers (*Stenzel, Chinellato, Tirado Bou, & del Pobil, 2012*). Note should be taken that the referential-coding approach (*Dolk et al., 2011*; *Dolk, Hommel, Prinz, & Liepelt, 2013*; for a review see *Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz, & Liepelt, 2014*) however hardly explains why Stenzel and colleagues (2012) failed to observe this congruency effect when the robot was described as an unintentional agent whereas the SSE was present when participants were induced to believe that the very same robot was an intentional human-like agent (*Stenzel, Chinellato, Tirado Bou, & del Pobil, 2012*).

Note should be taken that a TMS study by Tai and colleagues (2004) showed that the observation of a non-biological reach-to-grasp action executed by a humanoid robotic arm did not elicit participants' action/observation matching system (*Tai, Scherfler, Brooks, Sawamoto, & Castiello, 2004*). Hence, probably due to the divergences in the methodology used, it is unclear whether human beings are sensitive to machine-generated actions.

### 2.5.2. The sense of agency during human-machine interactions

a) A loss of agency during interactions with traditional machines

A clear loss of agency has been reported concerning individuals' sense of self-agency and vicarious agency when interacting with non-human-like artificial agents (*Berberian, Sarrazin, Le Blaye, & Haggard, 2012; Obhi & Hall, 2011b*). For example, in Obhi and Hall (2011b)'s study, participants had to tap a touchpad at the time of their choice, which triggered a tone after a certain delay. In the human-machine interaction condition, participants were performing the task with a non-visible desktop computer. They thought that either their keypresses or their co-agent actions could trigger an auditory tone but, in fact, the sensory effect was always triggered by the participants' actions. After each trial, a false feedback about the initiator of the tone was given to participants. Participants were asked to judge the onset time of their own action or the onset time of the tone with the help of a clock. Participants' beliefs about the initiator of the tone always fitted with the false feedback they were given. Interestingly, the authors found that participants did not demonstrate IB when they were convinced that the tone was computer-generated, and most intriguingly, also when they were convinced that they were responsible for the tone. That is, regardless of participants' beliefs no experience of agency took place despite the fact that it was indeed the participant who caused the sensory effect (*Obhi & Hall, 2011b*).

Furthermore, Berberian and colleagues (2012) showed that human operators experienced a very strong decrease of their sense of agency when interacting with highly automated autopilot interfaces (*Berberian, Sarrazin, Le Blaye, & Haggard, 2012*). In their study, Berberian and colleagues (2012) asked participants to perform an aircraft supervision task using a flight simulator under different levels of automation, from no level of automation to full automation. The task required participants to observe a flight plan and after a random time interval, a conflict occurred due to the presence of another plane. Participants were required to decide and implement an appropriate command using a button-based interface. A visual feedback was given to them after a variable delay (750 ms, 1500 ms or 2250 ms) that they had to estimate. In addition, participants were asked to report on a 4-point Likert scale their explicit judgment of agency by answering how strongly they felt that they caused the maneuver to avoid conflict. The results showed that participants exhibited stronger IB and explicit agency rating when they were interacting with low or no automation compared to highly automated autopilot interfaces. Hence,

participants reported a very strong decrease of their sense of agency when interacting with highly automated artificial systems. In the same time, they reported a lack of control over the action outcomes generated by the automated system.

b) An improvement in the sense of agency during interactions with human-like machines

Human-like automata could more easily be considered as intentional agents and may facilitate action co-representation and in turn ensure optimal human-machine interactions. In line with this notion, it has been shown that humanized machines could enhance the individuals' sense of agency compared to traditional automata during the passive observation (*Wohlschläger, Haggard, Gesierich, & Prinz, 2003*).

During interactive tasks, Caspar and colleagues (2016) have suggested that a humanized artificial system can improve participants' self-agency when they were engaged in a tracking task with this artificial agent (*Caspar, Desantis, Dienes, Cleeremans, & Haggard, 2016*). In their study, the participants were wearing a glove with sensors on their right hand hidden from vision so that they could control a human-like robotic right hand that was placed in full view in front of them. Participants learnt during an association phase a given keypress (e.g., H or F key) would trigger a specific auditory tone (e.g., a 400 Hz tone or a 600 Hz tone). In the robot homologous-tone congruent condition, participants had to press either the H or F key whenever they wanted and the robotic hand immediately did the same action. The robot's keypress triggered a tone that was congruent with the tone learnt in the association phase. In the robot homologous-tone incongruent condition, participants had to press one of the two keys whenever they wanted and the robotic hand immediately did the same action. However, the robot's keypress triggered a tone that was incongruent with the tone learnt in the association phase. In the robot non-homologous-tone congruent condition, participants had to press either the H or F key whenever they wanted but the robotic hand immediately did the opposite action. The robot's keypress triggered a tone that was congruent with the tone learnt in the association phase. In the robot non-homologous-tone incongruent condition, participants had to press one of the two keys whenever they wanted but the robotic hand immediately did the opposite action. In addition, the robot's key press triggered a tone that was incongruent with the tone learnt in the association phase. Participants had to estimate the temporal interval between the keypresses and the subsequent tones. The EEG activity of the participants was also recorded during the

experiment. The authors found that when the robot's action was homologous to the participant's action, the participants made more IB when the tone was congruent compared to when it was incongruent. This effect of congruency did not exist anymore when the robot's action was incongruent with the participant's action. In this respect, participants demonstrated a stronger sense of agency during their interactions with the robotic hand when it followed their own action compared to when it performed an action different from the participant's. In addition, the authors found that N1 amplitude was reduced during the perception of congruent tone compared to incongruent tone only when the robot action was homologous (*Caspar, Desantis, Dienes, Cleeremans, & Haggard, 2016*). Such a sensory attenuation has been considered as an implicit measure of participants' sense of agency (*Poonian, Mcfadyen, Ogden, & Cunnington, 2015; Weiss, Herwig, & Schütz-Bosbach, 2011*). Taken together, these results indicated that the association between a given action and its sensory consequence is not the only thing that matters in individuals' sense of agency construction. In addition, the means used to realize the desired outcome is important. The authors proposed that individuals' sense of agency was mainly informed by an online tracking control process that can predict the intermediate steps along the causal chain. When a disruption occurred, the sense of control over the other-generated action was hence reduced. More importantly, these results suggest that individuals can exhibit an implicit sense of agency during human-machine interaction, under certain circumstances.

## 2.6. Summary and scientific objectives

To summarize, predictive motor mechanisms that are involved in action control have also been shown to intervene in individuals' sense of self-agency (*Frith, 2005; Frith, Blakemore, & Wolpert, 2000; Sato & Yasuda, 2005*). It has been proposed that individuals' sense of self-agency did not only come from such sensorimotor cues but rather from the dynamic integration of both internal motor cues and external contextual cues (*Moore & Fletcher, 2012*). Interestingly, as suggested by Pacherie (2012), the cognitive mechanisms that are involved in the sense of agency during individual actions are of the same kind as those underlying the sense of agency during joint actions with other humans (*Pacherie, 2012*).

Hence, it has been shown that individuals exhibited a sense of we-agency for actions and outcomes generated by another human during joint action contexts (*Dewey, Pacherie*

*& Knoblich, 2014; Obhi & Hall, 2011a*). Such ability is in pace with the cognitive co-representation of other-generated human actions (*Keysers & Gazzola,* 2009; *Rizzolatti & Craighero,* 2004; Sebanz*, Knoblich, & Prinz, 2003*).

On the contrary, it has been shown that individuals experienced difficulties in building a sense of agency both for self- and other- generated actions when interacting with traditional machines (*Berberian, Sarrazin, Le Blaye, & Haggard, 2012*; *Obhi & Hall, 2011b*). In addition, individuals' ability to co-represent machine-generated actions into their cognitive system is still in debate (*Gazzola, Rizzolatti, Wicker & Keyser, 2007; Mann, Williams, Ward, & Janelle, 2007; Tai, Scherfler, Brooks, Sawamoto, & Castiello, 2004; Tsai, Kuo, Hung, & Tzeng, 2008; Wen & Hsieh, 2015*).

However, to our knowledge, no direct links have been established between the ability to co-represent other-generated actions into one's cognitive system and the ability to build a sense of agency for these actions during collaborative tasks. Given these elements, we sought to investigate the foreseeable link between these two phenomena in this thesis work. Further, we aimed at finely understanding the underlying processes involved in the construction of the sense of we-agency. Finally, we aimed at proposing some means to mitigate the negative impact of automation on individuals' sense of control during human-machine interactions.

# Chapter 3: Experimental contribution

# 3.1. Experiment 1: Action co-representation and the sense of vicarious agency: comparing human and machine co-agents

## 3.1.1 Context of the experiment

The first objective of this experiment was to examine individuals' sense of self-agency and sense of vicarious agency during a joint task with another human versus a desktop computer. We choose a desktop computer as automated artificial system considering that it is the artificial system that people most commonly work with. Obhi and Hall (2011b) previously showed that individuals experienced a loss of sense of agency when they were interacting with an automated artificial system such as a desktop computer. Indeed, they found that, in a joint task, participants exhibited IB for their own actions and the actions generated by another human co-agent, whereas, when acting jointly with a machine, IB for their own actions and those generated by the machine disappeared (*Obhi & Hall, 2011b*).

However, there was one major limitation in the author's study. Indeed, in Obhi and Hall (2011b)'s study, it was always the naïve participant who actually caused the sensory consequence (i.e., the auditory tone). Participants were induced the false belief that they could also trigger the sensory consequence if they had tapped a touchpad before them. But in facts, the sensory consequence was always triggered by the participants' keypresses. Hence, Obhi and Hall (2011b)'s findings can be explained in light of the predictive model of action (**Figure 3**, CM, *Frith, Blakemore, & Wolpert, 2000*). Indeed, regardless of participants' thoughts about the initiator of the sensory effect, participants always performed the keypress so that motor information (e.g., the motor command, the forward model and the sensorimotor feedbacks) was in any case available. In addition, participants were always asked to judge the onset of their own action and not the supposed onset of the action of the co-agent's action. Thus, the IB effect found by the authors might reflect the sense of self- agency of the participant himself or herself, which arises from the match between the predicted state (through the forward model) and the estimate actual state. Hence, in the current experiment, we sought to examine individuals' sense of agency in a joint task for effective other-generated actions (human or machine).

The second objective of this experiment aimed at exploring the foreseeable link between individuals' sense of agency during actions performed by another agent and action

co-representation mechanisms during a joint task. Indeed, it has been suggested that the experience of agency for the actions performed by another individual was based on the ability to simulate or co-represent that individual's motor plans and intentions (*Obhi & Hall, 2011b*). Based on this assumption, people would not experience a sense of vicarious agency when interacting with computers because they failed to simulate computer-generated actions (*Obhi & Hall, 2011b*).

We ran a behavioral study combining a Social Simon response time task and an intentional binding task. More specifically, participants performed the Simon task alone (standard Simon condition), or with another agent (joint go/no-go conditions) that could be another human or a desktop computer. Accurate target detections triggered an auditory tone after a certain delay. Participants were required to estimate the time interval between their action or the action of the co-agent and the subsequent auditory outcome. The response times observed in the Simon task served as an index of action co-representation while IB was used as an implicit measure of the sense of agency (*Haggard, Clark, & Kalogeras, 2002*). We hypothesized that participants would co-represent the human-generated actions but not the computer-generated actions. In the same time, they would manifest a sense of agency for the human-generated actions but not the computer-generated actions.

### 3.1.2. Method

a) Ethic statement

This study was approved by the institutional ethical research committee of the Paris Descartes University (France). The investigation was carried out in accordance with the Declaration of Helsinki and 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. Participants were not paid for their participation in the experiment.

b) Participants

Sixteen healthy adults volunteered to take part in the experiment (6 women, mean age 24.75 years, SD of age 3.15 years). One participant was excluded from the sample due to his poor time discrimination performances observed during the training session. All

participants were right-handed and had normal or corrected-to-normal vision. None of them had prior knowledge about the purpose of the experiment. Participants were not paid for their participation in the experiment.

c) Material and stimuli

Participants were sitting about 46 cm from a computer screen with a refresh rate of 120 Hz.

Stimulus presentation was controlled using E-prime software (2.0 version). Visual stimuli consisted of two dots of 0.5 cm diameter, one green and the other red. A sine wave sound (1000 Hz, 200 ms duration), presented via a speaker placed behind the participant, was used during the experiment as the effect of the participant/co-agent's button press for measuring intentional binding.

d) Procedure

Participants were sitting in front of a screen and had to detect, as quickly and as accurately as possible, colored dots that appeared either to the left or to the right side of a central fixation cross. This task could be performed alone, jointly with another human (the experimenter), or jointly with an algorithm implemented in a desktop computer.

In the *standard Simon* condition, participants were sitting in front of the center of the screen. They were instructed to perform as quickly as possible a right key-press with their right index finger when a green dot was displayed on the screen, regardless of whether it appeared on the left or right side of the screen. When a red dot was displayed they were required to execute a left key-press with their left index finger, regardless of whether it appeared on the left or right side of the screen (**Figure 5a**).

In the *individual go/no-go* condition, participants were sitting in front of the right side of the screen and an empty chair was placed in front of the left side of the screen. They were asked to perform a right key-press with their right index finger when presented with a green dot regardless of whether it appeared on the left or right side of the screen, and to not react when presented with a red dot (**Figure 5b**).

In the joint conditions, participants were sitting in front of the right side of the screen. They executed a right key press with their right index finger when presented with a green dot, independently of its location on the screen. When presented with a red dot they were asked to not react. Importantly, the task could be performed jointly with another human agent (*joint HH* condition, **Figure 5c**) or with an automated artificial system, i.e., an algorithm implemented in a desktop computer (*joint HM* condition, **Figure 5d**). When the co-agent was a human, she was sitting on the left side of the screen and performed left key-presses with her left index finger. When the co-agent was the desktop computer, an empty chair was placed in front of the left side of the screen.

In the passive observation conditions, participants were sitting in front of the right side of the screen and they had to observe another agent (another human agent in the *passive observation H* condition and the desktop computer in the *passive observation M* condition) performing the red dot detection (**Figure 5e** & **5d**). Both the human and the algorithm implemented in the desktop computer responded to the presentation of the red dot. When a green dot was displayed no response was required by the co-agent.

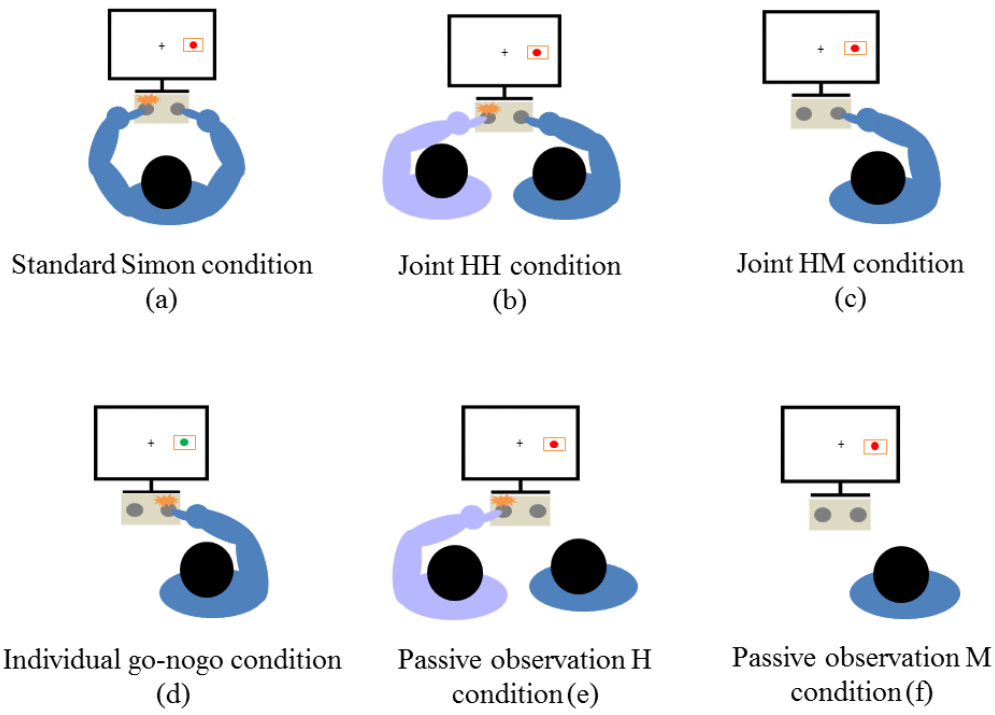

**Figure 5**. Setups of the experimental conditions: *standard Simon* condition (**a**), the *joint HH* condition (**b**), the *joint HM* condition (**c**), the *individual go/no-go* condition (**d**), *passive observation H* condition (**e**), and *passive observation M* condition (**f**).

Each trial started with a fixation cross that appeared at the center of the screen during 500 ms. Thereafter, the target immediately appeared on the screen and participants had at most 1800 ms to press their response key otherwise an error message appeared and the trial was canceled.

Participants were informed of the onset of their own action and the action of their co-agent (human or automated artificial) by the presentation of an empty square displayed around the target for a duration of 200 ms. Participants were required to fixate the computer screen throughout the experiment and to not look at the actions performed by the human agent. When the human co-agent performed the target detection task, the square onset times corresponded to the co-agent's real response times. When the algorithm performed the target detection task, the square onset times were taken from a normal distribution calculated from the mean and standard deviation of the human co-agent's response times computed during a pre-test session. More specifically, square onset times were selected randomly within two standard deviations of this distribution.

Correct target detections were followed by an auditory tone presented after the response at one of three possible Stimulus Onset Asynchronies (SOA) of 400 ms, 900 ms or 1400 ms. The SOA for a given trial was selected randomly. After the presentation of the sound, participants had 4 seconds maximum to verbally report the perceived duration between the onset of the target detection (indicated by the square appearing around the target) and the onset of the auditory tone (see **Figure 6**). This time interval estimates served as an implicit measure of participants' sense of agency (*Haggard, Clark, & Kalogeras, 2002*). Note that participants were not asked to perform the temporal estimations in the *standard Simon* condition as they made estimations for self-generated actions in the *individual go/no-go* condition.

Participants were trained at the beginning of each experimental condition block to estimate and report their perceived duration of the action-tone interval. During this training, they were presented with an empty square that flashed, followed by an auditory tone with a random delay between 200 ms and 2000 ms. They had to verbally report the perceived

duration of this action-tone interval in milliseconds. Then, they were given the correct delay with a visual feedback in order to accurately recalibrate their internal clock. This training session consisted in 20 trials. Thereafter, participants performed 20 trials of the given experimental condition as training. The goal was to familiarize them with the task so that they would associate their key press with the following auditory tones.

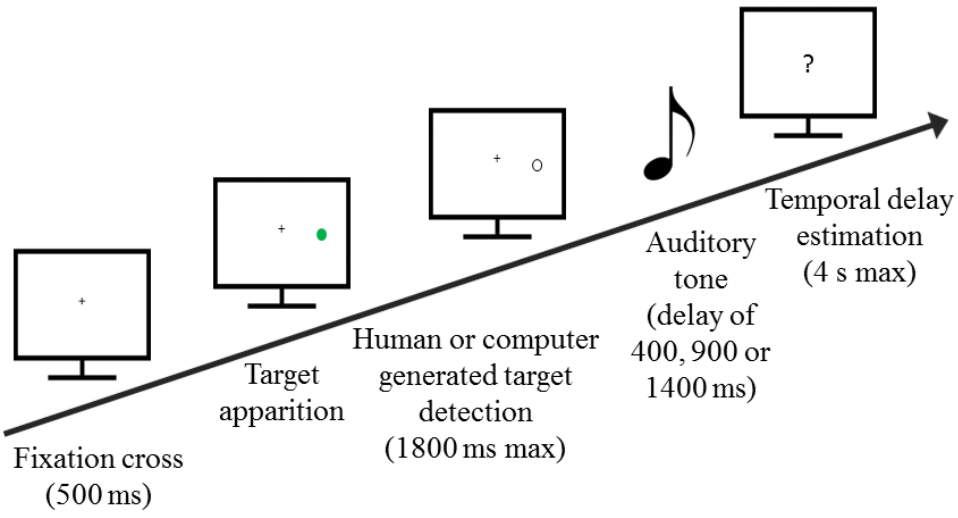


**Figure 6**. Trial timeline. A fixation cross appeared for 500 ms. Then, the target appeared and the agent (the participant or the co-agent) had to detect it before 1800 ms. All target detections were signaled by an empty square around the target. An auditory tone was generated at a delay of 400 ms, 900 ms or 1400 ms after target detection. The participant had to report the temporal delay between the onset of the target detection and the onset of the tone.

The experimental conditions were tested using a within-subject design and the order of conditions that each participant completed was selected randomly. Trials were 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. Participants completed a total of 3600 trials, (6 Tasks (*standard Simon*, *individual go/no-go*, *joint HH*, *joint HM*, *passive observation H*, and *passive observation M*) x 2 Targets (*green dot*, *red dot*) × 2 Congruency levels (*congruent*, *incongruent*) × 3 Delays (*400*, *900*, *1400*) × 50 trials).

### 3.1.3. Data analyses

Our dependent measures were the participants' mean target detection Response Times (RTs) and mean perceived action-tone interval. Statistical analyses were performed with R software (3.3.1 version). Extreme values (the values that were below or above 2 standard deviations from the mean) of the participants' RTs and perceived intervals were excluded from further analyses in order to eliminate outliers and allow for robust statistical analyses. The significance level was set at α = .05. Post-hocs comparisons were made using Tukey's HSD test. To distinguish the participants' trials from the co-agent's trials in the joint conditions, participants' trials were labeled *joint HH self* and *joint HM self*, and the co-agents' trials were labeled *joint HH other* and *joint HM other*.

a) Social Simon Effect (SSE)

This analysis was based exclusively on the data gathered in the conditions in which participants performed an action (*standard Simon*, *individual go/no-go*, *joint HH self*, and the *joint HM* self). The analysis aimed at quantifying the SSE when participants interacted with a human co-agent (*joint go/no-go HH self)* and with an automated artificial system such as an algorithm implemented in a desktop computer (*joint go/no-go HM self)*. Because the order of experimental conditions that each participant completed was not counterbalanced, we introduced the Order (1st, 2nd, 3rd, 4th, 5th, 6th) as a covariate in a linear mixed model ($m_{\widetilde{RT}}$), with Task and Congruency as fixed factors and Participant as random effect. The factor Order indicates when a given task (e.g. *standard Simon*) was presented during the experiment. For instance, if the participant performed firstly the standard Simon task, we would attribute the Order "1st" to this task. Instead, if the standard Simon task was performed second, we would attribute the order "2nd" to this task and so on for the other Order values. This model ($m_{\widetilde{RT}}$) was then compared, using a likelihood ratio test, to a simpler model that included the same fixed and random effects but without order as a covariate ($m_{RT}$):

$m_{RT}$: RT ~ Task * Congruency, random = Participant

$m_{\widetilde{RT}}$: RT ~ Task * Congruency + Order, random = Participant

We found no significant difference between $m_{RT}$ and $m_{\widetilde{RT}}$ (respectively logLik = -697.94 and logLik = -695.97, $\chi^2(5) = 3.94$, $p = .56$). This indicates that including Order as a

covariate did not improve our model of the observed data. Thus, Order was excluded from further analyses. We conducted a within-subjects analysis of variance (ANOVA) on the RTs with Task (*standard Simon*, *individual go/no-go*, *joint HH self*, *joint HM self*) and Congruency (*congruent*, *incongruent*) as factors. The Target (*green dot*, *red dot*) factor was not included in the ANOVA because the SSE does not rely on the target identity but rather on the congruency between the location of the target and the location of the response key. The Delay (*400*, *900*, *1400*) factor was irrelevant for the analysis as the auditory tone was produced after participants' response and therefore could not influence their RTs.

b) Engagement in the task

This analysis aimed at examining if participants' engagement in the joint Tasks (*joint HH self* and *joint HM self*) varied according to the type of Co-agent (*human*, *computer*). We analyzed participants' RTs only on congruent trials because incongruent trials involved interference mechanisms that could induce biases for the engagement analysis. Specifically, we hypothesized that the social feature of the co-agent will increase participants' engagement in the task. We conducted a within-subjects one-way ANOVA on participants' RTs during *congruent* trials with Task (*joint HH self*, *joint HM self*) as a factor.

c) Temporal interval estimation accuracy checking

To verify whether participants correctly perceived the different temporal delays between the onset of the actions and the onset of the subsequent auditory tone, we conducted a prior within-subjects one-way ANOVA on the raw temporal estimations with Delay (*400*, *900*, *1400*) as a factor. This control analysis allowed us to ensure that participants paid attention to the action-sound intervals.

d) Intentional binding (IB)

This analysis aimed at investigates the influence of the co-agent on participants' agency experience. To characterize the IB phenomenon, we considered the mean perceived action-tone interval in the *passive observation M* condition as our baseline, i.e., when no agency was involved at all (see *Poonian, McFadyen, Ogden, & Cunnington, 2015*;

*Wohlschläger, Engbert, & Haggard, 2003*). We measured the IB by subtracting the mean perceived action-tone interval in the *passive observation M* condition from the mean perceived action-tone interval in all other experimental conditions (negative values indicate temporal underestimations in these conditions compared to the baseline). Hence, the *passive observation M* condition was not included as a factor in the IB analyses.

Note that for the intentional binding analyses we did not include the Delay (*400*, *900*, *1400*) as a separate factor. Indeed, in the present experiment we were interested in the way the social context influences IB in general, rather than its influence on IB for different action-tone intervals.

Given that the order of the experimental conditions was not fully counterbalanced between participants, we introduced the Order (*1st*, *2nd*, *3rd*, *4th*, *5th*, *6th*) factor as a covariate in a linear mixed model ($m_{\widetilde{IB}}$) with Action Context (*human individual action*, *human-human joint task*, *human-computer joint task*), Agent (*self, other*) and Congruency (*congruent*, *incongruent*) as fixed factors and Participant as random effect. The Action Context referred to the social context in which the Simon task was performed, i.e., whether the agent was acting alone (*human individual action*), co-acting with another human (*human-human joint task*), or with a computer (*human-computer joint task*). Accordingly, the Human individual action context included the *passive observation H* for the actions that were performed by the human co-agent and the *individual go/no-go* condition for the actions performed by the participant her/himself. Note that there was no temporal estimation measurement in *the standard Simon* condition. That is why this condition was not included in the analyses of the IB phenomenon. The Agent (*self*, *other*) factor referred to the author of the action, i.e., the participants themselves or their co-agent respectively (see **Table 2**). Using a likelihood ratio test, we then compared this model ($m_{\widetilde{IB}}$) with another mixed model ($m_{IB}$) with only Action context, Agent, and Congruency as fixed factors and Participant as random effect:

$m_{IB}$: IB ~ Action Context * Agent * Congruency, random = Participant

$m_{\widetilde{IB}}$: IB ~ Action Context * Agent * Congruency + Order, random = Participant

| | | Factor: Action Context | | |
|---|---|---|---|---|
| | | **human individual action** | **human-human joint task** | **human-computer joint task** |
| Factor : Agent | **self** | individual go/no-go | joint HH self | joint HM self |
| | **other** | passive observation H | joint HH other | joint HM other |

**Table 2**. Factorial design for the intentional binding (IB) analysis.

The analysis showed that the $m_{\widetilde{IB}}$ model including Order as a covariate fitted our data better than the $m_{IB}$ model (respectively logLik = -1135.63 and logLik = -1154.96, $\chi^2(5) = 38.66$, $p < .001$). Hence, we conducted a within-subjects ANOVA on IB with Action Context (*human individual action*, *human-human joint task*, *human-computer joint task*), Agent (*self*, *other*) and Congruency (*congruent*, *incongruent*) as fixed factors and Order (*1st*, *2nd*, *3rd*, *4th*, *5th*, *6th*) as a covariate.

### 3.1.4. Results

a) Social Simon Effect (SSE)

We examined the SSE during a joint task human and with an algorithm implemented in a desktop computer. We assessed the normality of the RTs distributions of the differences between the *congruent* trials and the *incongruent* trials in the *standard Simon*, *individual go/no-go*, *joint HH self* and *joint HM self* conditions using the Shapiro-Wilk test. The analyses showed that none of the RTs distribution deviated from normality (all $W > .90$ and all $p > .10$). We then computed a within-subjects 4 x 2 ANOVA on the RTs with the factors Task (*standard Simon*, *individual go/no-go*, *joint HH self*, *joint HM self*) and Congruency (*congruent*, *incongruent*). We found a significant main effect of Congruency on RTs indicating longer mean RTs on *incongruent* trials compared to *congruent* trials ($F(1,14)$

= 5.47, *p* = .03) but no significant main effect of Task (*F*(3,42) = .80, *p* = .50, ns). Moreover, we found a significant Congruency x Task interaction on RTs (*F*(3,42) = 9.99, *p* < .001).

Post-hoc comparisons revealed that mean RTs on *incongruent* trials was significantly longer than the mean RTs on *congruent* trials in the *standard Simon* condition (respectively 426.19 ms (SD = 58.65) and 409.25 ms (SD = 66.51); *p* = .001) and in the *joint HH self* condition (respectively 403.51 ms (SD = 140.85) and 390.23 ms (SD = 141.80); *p* = .02) but not in the *individual go/no-go* condition (respectively 416.09 ms (SD = 139.21) and 414.32 ms (SD = 140.94); *p* = .99, *ns*) nor in the *joint HM self* condition (respectively 434.78 ms (SD = 121.56) and 444.16 ms (SD = 135.12); *p* = .23, ns). Hence, we observed the classical Simon Effect in the standard Simon task and when participants performed the task with another human agent. No Simon effect was observed when participants interacted with a machine (see **Figure 7**).

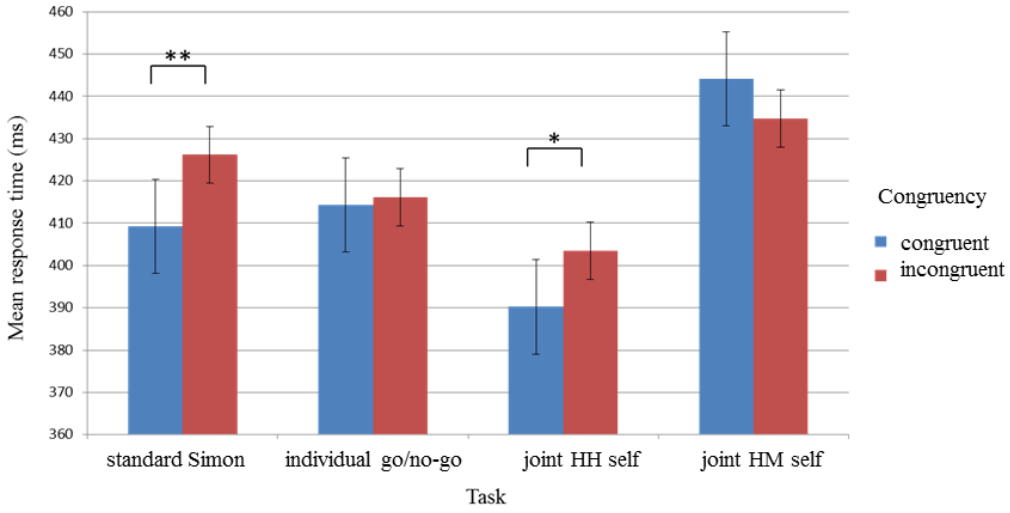


**Figure 7**. Interaction between Congruency and Task on participants' mean response times. Error bars represent standard errors.

b) Engagement in the task

We investigated participants' engagement during the joint tasks according to the nature of the co-agent they partnered with (i.e., human vs. computer). We conducted a within-subjects one-way ANOVA on participants' mean RTs during *congruent* trials with Task (*joint HH self*, *joint HM self*) as a factor. We found a main effect of Task on participants' mean RTs (*F*(1,14) = 14.86, *p* = .002) indicating shorter mean RTs in the *joint*

*HH self* task (390.23 ms, SD = 141.80) compared to the *joint HM self* task (444.16, SD = 135.12). Hence, this result suggests that participants' engagement in the task was better when they cooperated with another human than when they performed the joint task with a machine such as a desktop computer.

c) Temporal interval estimation accuracy checking

We verified whether participants correctly perceived the different action-tone intervals that were used in the task by assessing whether they adjusted their temporal estimations accordingly. We investigated the effect of Delay on participants' mean action-tone interval estimations. We computed a within-subjects one-way ANOVA with the factor Delay (*400*, *900*, *1400*). We found a significant main effect of Delay on participants' mean action-tone interval estimations ($F(2,28) = 316.71$, $p < .0001$).

Post-hoc comparisons revealed that individuals reported significantly higher action-tone intervals when the Delay between the action and the tone was 1400 ms compared to when the delay was 900 ms (respectively 1133.04 ms (SD = 249.62) and 797.83 ms (SD = 198.30); $p < .0001$) and when the delay was 400 ms (433.57 ms (SD = 141.75); $p < .0001$). In addition, individuals reported in average higher action-tone intervals with a 900 ms delay compared to a 400 ms delay ($p < .0001$). Even though this was not a variable of interest (indeed, we only introduced various delays in order to avoid the predictability bias), this effect demonstrated that participants were paying attention to the temporal delays and correctly discriminated the different delays.

d) Intentional binding (IB)

The amount of IB was calculated by subtracting the mean perceived action-tone interval in the *passive observation M* condition from the mean perceived action-tone interval in all the other conditions. We computed a within-subjects 3 x 2 x 2 ANOVA on IB with Action Context (*human individual action*, *human-human joint task*, *human-computer joint task*), Agent (*self*, *other*) and Congruency (*congruent*, *incongruent*) as fixed factors and Order (*1st*, *2nd*, *3rd*, *4th*, *5th*, *6th*) as a covariate.

We found neither a main effect of Agent ($F(1,14) = .11$ $p = .74$, *ns*) nor a main effect of Congruency ($F(1,14) = 0.38$, $p = .54$, *ns*) on IB. However, we found a significant main

effect of Action Context (*F(2,28)* = 10.78, *p* < .001) and a significant Action Context x Agent interaction (*F(2,28)* = 11.24, *p* < .001, see **Figure 8**) on IB. Considering that Congruency did not interact with the other factors, we averaged congruent and incongruent trials for further analyses.

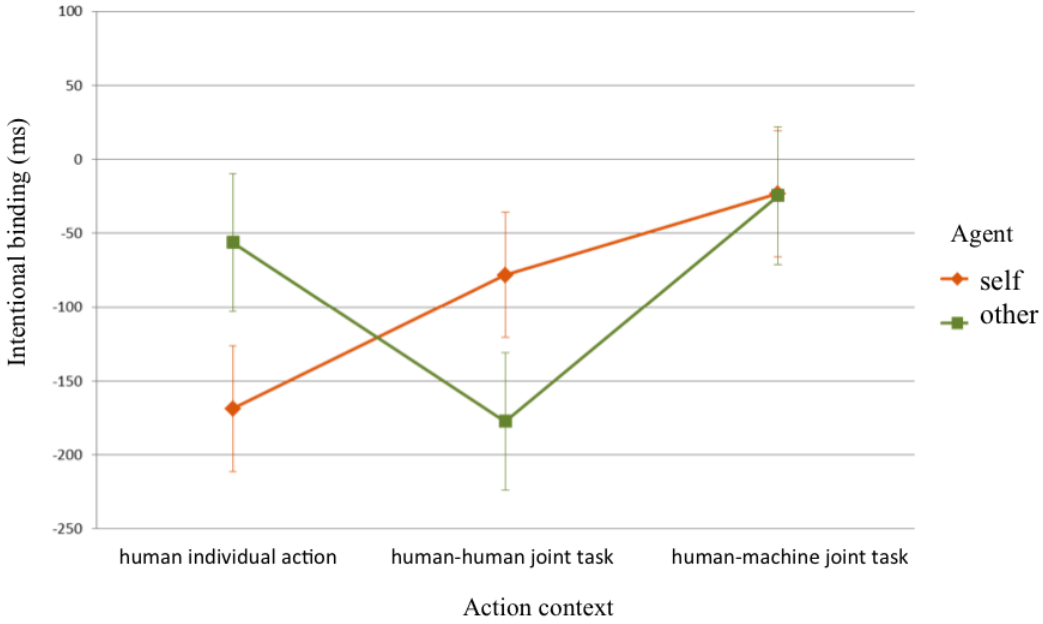


**Figure 8**. Interaction between Action context and Agent on IB. The vertical bars represent the standard error values.

Post-hoc comparisons investigating the main effect of Action Context on IB revealed that participants exhibited greater IB in the *human individual actions* context than in the *human-machine joint tasks* context (respectively -115.84 ms (SD = 155.17) and -31.74 ms (SD = 208.95); *p* < .001). In addition, there was no difference on IB between the *human individual actions* context and the *human-human joint tasks* context (*p* = .24, *ns*). Finally, participants exhibited greater IB in the *human-human joint tasks* context (-130.14 ms (SD = 162.96)) than in the *human-machine joint tasks* context (*p* < .001).

Post-hoc comparisons investigating the Action Context x Agent interaction on IB revealed that IB for *self*-generated actions was greater when participants were acting alone (*individual go/no-go* condition) compared to when they were jointly acting with a machine (respectively -175.05 ms (SD = 165.58) and -30.94 ms (SD = 218.75); *p* < .001). However, IB for *self*-generated actions did not differ when participants were performing the task alone

(*individual go/no-go* condition) compared to when they were interacting jointly with another human agent (-81.04 ms (SD = 165.19); $p$ = .34). Similarly, IB did not differ when participants acted with another human agent compared to when they interacted with a machine, although there was a trend ($p$ = .06, *ns*). Accordingly, this indicates that IB observed for *self*-generated actions did not differ whether participants performed the task alone or interacted with another human agent but tended to strongly decrease when participants interacted with a machine.

Additional post-hoc comparisons on IB for *other*-generated actions revealed greater IB when participants performed the task with another human agent compared to when they were observing the same agent completing the task alone (*passive observation H* condition) (respectively -179.27 ms (SD= 147.46) and -56.62 ms (SD = 119.64); $p$ < .001). In addition, we observed greater IB for *other*-generated actions when participants performed the task jointly with a human agent compared to when they performed it with the computer (-32.54 ms (SD = 202.43); $p$ < .001). However, there was no difference in IB for *other*-generated actions when participants observed a human agent performing the task alone (*passive observation H* condition) compared to when they interacted with the computer ($p$ = .81, *ns*). Accordingly, this indicates that the IB observed for *other*-generated actions strongly increased only when participants performed the task together with a human agent compared to the mere passive observation of the same action. This enhancement was not observed during joint tasks with a machine.

Further post-hoc comparisons showed greater IB for *self*-generated actions compared to *other*-generated actions when participants and the other human agent were performing the task alone (*human individual action* context) ($p$ = .005). During the *human-human joint task*, IB was weaker for *self*-generated actions compared to *other*-generated action ($p$ = .03). There was no difference on IB between *self*- and *other*-generated actions in the *human-machine joint task* ($p$ = 1.00, *ns*). In addition, no difference was observed between the IB for *self*-generated actions performed alone and the IB for *other*-generated actions in the context of *human-human joint task* ($p$ = .95, *ns*). In sum, participants reported stronger IB when they performed the task alone compared to when they passively observed another person performing the same action. However, during the *human-human joint task*, this trend was reversed. Notably, they reported stronger IB for the action of their co-agent than for their own action.

### 3.1.5. Discussion of the experiment

In this study, our aim was (1) to investigate individuals' sense of self-agency and sense of vicarious agency during a joint task with a human versus an automated artificial system such as a desktop computer, and (2) to explore the link between individuals' sense of agency during actions performed by a co-agent and action co-representation mechanisms during a joint task. To assess action co-representation, we used a typical social Simon task where participants had to detect a target that could appear on the same side as their response key or on the opposite side. Accurate target detection triggered an auditory tone after a randomized delay. Participants had to estimate the temporal delay between the target detection and the onset of the tone. This estimate served as an implicit measure of participants' sense of agency (*Haggard, Clark, & Kalogeras, 2002*).

We found the classical SSE with longer response times in *incongruent* trials when participants performed a double target detection task (*standard Simon* task) but not when they performed a single target detection task (*individual go/no-go task*). Moreover, we found the SSE when the participants performed a single target detection task jointly with another human agent performing the complementary action (*joint HH* task). This suggests that, during human-human interactions, participants experienced the action performed by a human as if it was their own action, in regard with the work of Sebanz and colleagues (2003) (*Sebanz, Knoblich, & Prinz, 2003*). However, this effect disappeared when the computer program replaced the human co-agent. These findings are consistent with previous results supporting the claim that the social identity of the co-agent critically influences individuals' action co-representation abilities during a joint task (*Stenzel, Chinellato, Tirado Bou, & del Pobil, 2012; Wen & Hsieh, 2015*).

One possible explanation is that the artificial agent was not considered as an agent either because it was not perceived as acting intentionally or because the causal relationship between the agent and the action effect was opaque or both. Indeed, a previous study investigating the SSE during interactions with a humanoid robot failed to observe this effect when the robot was described as an unintentional device that passively executed a deterministic command whereas the SSE was present when participants were induced to believe that the robot was an intentional active and intelligent human-like agent (*Stenzel, Chinellato, Tirado Bou, & del Pobil, 2012*). In addition, it has been proposed that the perceived co-agent's agency plays an important role in the emergence of SSE (*Stenzel, Dolk, Colzato, Sellaro, Hommel, & Liepelt, 2014*). In their study, Stenzel and colleagues

(2014) asked participants to perform a joint Simon task next to a co-agent who either intentionally performed the response key presses (agency+/intentionality+ condition), or passively placed their finger on the response key that automatically moved up and down (agency-/intentionality- condition), or who was thought to controlled the response key with a brain-computer interface while placing their finger besides the response key (agency-/intentionality+ condition). The authors found that the SSE was manifested only in the agency+/intentionality+ condition, that is to say, when the causality between the co-agent and the action effect was not disrupted. Hence, perceived agency may play an important role in the SSE (*Stenzel, Dolk, Colzato, Sellaro, Hommel, & Liepelt, 2014*). When considering the referential-coding approach to explain the Congruency effect (*Dolk et al., 2011*; *Dolk, Hommel, Prinz, & Liepelt, 2013*; for a review see *Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz, & Liepelt, 2014*), one can argue that the human in our experiment may have been more salient than the computer and could thus have grabbed the attention of the participant to a higher degree, influencing participants' RTs. Unfortunately, this study was not designed to and cannot provide an answer to this debate. However, the decrease in participants' RTs, as well as the development of a sense of vicarious agency (described below) when participants' partnered with a human co-agent compared to a machine co-agent might suggest that the Congruency effect has social matters

Interestingly, we observed that IB phenomenon for *self*-generated actions remained stable regardless of whether participants performed the task alone or with another human agent. However, IB decreased strongly when participants interacted with a machine. This pattern of results indicates that the social context played a pivotal role in implicit self-agency attribution as measured by the intentional binding phenomenon. Even though participants in the present experiment had to perform the same action leading to the same sensory effect in all conditions, they tended to bind their own action and its effect differently according to the social context. The intentional binding for their action was maximal when they were performing the task alone. It decreased drastically as soon as they were collaborating with a machine. Thus, sharing a task with an artificial agent seems to decrease the sense of self-agency. This result corroborate previous findings showing a loss of self-agency during a human-human cooperative task compared to an individual task, even when there was no ambiguity about who had caused the outcome (*Beyer, Sidarus, Bonicalzi, & Haggard, 2017*). This current experiment provides however original evidence that this process of diffusion of responsibility does also occur during human-computer interaction.

More importantly, the sense of vicarious agency for the *other*-generated actions disappeared when the co-agent was the computer, as did the SSE on the RTs. Hence, the IB phenomenon seemed to parallel the SSE effect when comparing the *human-human joint task* with the *human-machine joint task*. This lends support to the view that the sense of agency and the SSE could potentially share common underlying mechanisms, but this needs to be confirmed by neuroimaging studies. As mentioned earlier, there are currently two major accounts explaining the Congruency effect observed in a joint Simon task. The first more classical framework posits that we automatically activate our own sensorimotor representations when partnered with a human co-agent (SSE, *Sebanz, Knoblich, & Prinz, 2003*). The second theory suggests that action coding and thus decision is influenced by salient environmental cues (*Dolk et al., 2011*; *Dolk, Hommel, Prinz, & Liepelt, 2013*; for a review see *Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz, & Liepelt, 2014*). However, a third explanation combining the two views could also be taken into consideration, explaining both the RTs and the IB results of the current experiment. In the case of individual *self*-generated actions, individual's sensorimotor network was activated and attention was focused on the ongoing task, leading to a greater sense of agency. During the *human-human joint task*, the need of participants to coordinate their actions with another human being co-agent induced an attentional shift toward the *other*-generated actions, which in turn engaged the participant's sensorimotor network, leading to a sense of vicarious agency in the specific context of the *human-human joint task*. Finally, during the *human-machine joint task*, the absence of intentional co-agent brought participants to pay less attention to the actions performed by the machine. As a consequence, participants' sensorimotor network was not activated in this case, leading also to a decrease of agency for the actions generated by the machine. Taken together these results suggest that participants experienced a sense of vicarious agency during joint tasks with other human beings but not with desktop computers. As for now, the findings might be considered to provide further evidence in favor of the view that the human representational system is biologically tuned (*Tsai & Brass, 2007*).

Other interesting results were observed in this study. First, IB for the *other*-generated actions was stronger during the *human-human joint task* compared to the *human individual action*. This suggests that being involved in a joint task lead to a stronger representation of the agency of the co-actor compared to the mere passive observation of the same action.

Second, we found that during the *human-human joint task*, IB for the *other*-generated action was stronger compared to IB for the *self*-generated actions. In other words, participants' sense of agency for the co-agent's action was stronger than for the participant's own actions. This outcome contrasts with the experiment reported by Obhi & Hall (2011a), in which the amount of IB was the same for the participant's and the co-agent's action when they were performing a joint task. This difference might depend on the fact that Obhi & Hall (2011a) used a very different design compared to the current experiment. In their experiment, the two participants were required to press the same button jointly, whereas in our experiment the participant and their co-agent pressed different buttons at different times in response to different stimuli. The way agency is allocated among partners may depend on the type of collaboration required by the task. When a participant and a co-actor are required to perform an action together (as in *Obhi & Hall, 2011a*), the IB observed might reflect a summation of both the individual' sense of self-agency and sense of vicarious agency.

Lastly, we did not observe any effect of Congruency on participants' sense of agency. This is in contrast with past studies showing that the fluency of action selection enhances participants' explicit judgment of control (*Chambon & Haggard, 2012; Sidarus & Haggard, 2016*). In our study we did not find any effect of action facilitation on participants' sense of agency. However, unlike previous experiment on the fluency of action, we did not use an explicit rating scale but an implicit measure of agency, namely, the IB phenomenon. This comfort that IB and explicit judgments of agency do not share the same process (*Dewey & Knoblich, 2014*).

**Experiment 1** ventured empirical evidence that action co-representation goes in pace with the ability to develop a vicarious sense of agency for actions generated by a co-agent. Indeed, Participants' RTs showed a classical SSE when they were partnered with another human, but not when they collaborated with an algorithm implemented in a desktop computer. Concomitantly, participants showed a vicarious implicit sense of agency as indicated by IB phenomenon when co-acting with another human agent but not with the artificial automated system.

However, we were not able to investigate the direct link between these two phenomena. Indeed, in the experiment described above, action co-representation mechanisms were

appreciated using the SSE, which was computed from the participant's own RTs. On the contrary, the participants' vicarious sense of agency was assessed from the perceived duration between the onset of the other-generated actions and the onset of the auditory tone. Consequently, direct links between the ability to co-represent other-generated actions and the ability to develop a vicarious sense of agency for these actions could not be established. Hence, we ran an EEG study (**Experiment 2**) wherein both phenomena could be investigated on the same trial.

## 3.2. Experiment 2: Investigation of action co-representation and the sense of agency at the cerebral level

### 3.2.1. Context of the experiment

The present study aimed to explore at the cerebral level the link between individuals' sense of vicarious agency during actions performed by a co-agent and action co-representation mechanisms during a joint task with a human co-agent versus an automated artificial system.

A well-known implicit measure of individuals' sense of self-agency is the sensory attenuation phenomenon (*Poonian, Mcfadyen, Ogden, & Cunnington, 2015; Weiss, Herwig, & Schütz-Bosbach, 2011*). This stems from the CM (see **Figure 3**) that states that because self-generated actions can be better predicted through an internal forward model of the agent's motor system, there is little discrepancy with the actual sensory effects thus attenuating the salience of the following outcomes (*Blakemore, Wolpert, & Frith, 2000*). Hence, it has been shown that during the perception of self-generated auditory tones, participants reported a reduced loudness intensity in comparison with external computer-generated auditory tones (*Weiss, Herwig, & Schütz-Bosbach, 2011*). Likewise, at the cerebral level, it has been showed that self-generated auditory tones engendered a diminution of the amplitude of the auditory N1 event-related potential (ERP) component during their perception compared to computer-generated auditory tones. (*Poonian, Mcfadyen, Ogden, & Cunnington, 2015*). This sensory N1 component is a negative ERP evoked during the perception of auditory stimuli and peaking around 100 ms after the sound onset, with a more pronounced activity at fronto-central sites (*Loehr, 2013*; *Poonian, Mcfadyen, Ogden, & Cunnington, 2015*).

Interestingly, it has been demonstrated that sensory attenuation phenomenon was also sensitive to socially shared process such as individuals' implicit vicarious sense of agency for actions outcomes that have been generated by human peers. Indeed, at the cerebral level, Poonian and colleagues (2015) showed no difference between the amplitude of N1 component when participants listened to auditory tones that were self-generated and auditory tones that were generated by an observed human agent, suggesting similar agentive cerebral processing of self-generated events and those generated by another human being (*Poonian, Mcfadyen, Ogden, & Cunnington, 2015*). Besides, Loehr (2013)

showed that when participants performed self-generated keypresses that triggered auditory tones, the amplitude of N1 component was diminished during the perception of the tones in comparison with the perception of the same tones that were computer-generated. However, the author found that this sensory attenuation phenomenon was even more pronounced during the perception of auditory tones that were jointly generated with a human co-agent, suggesting that sensory attenuation phenomenon is also sensitive to the social context according to which the outcomes was produced (*Loehr, 2013*). In sum, there is evidence that sensory attenuation phenomenon, as indicated by N1 component amplitude diminution, is sensitive to individuals' sense of vicarious agency for outcomes generated by another human being.

Nevertheless, a clear consensus about such a vicarious sensory attenuation phenomenon is missing. For example, at the behavioral level, Weiss and colleagues (2011) found a diminished perception of the loudness of auditory tones specifically when they were self-generated, with no differences in participants' loudness judgments between human other-generated tones and computer-generated tones (*Weiss, Herwig, & Schütz-Bosbach, 2011*). In addition, Desantis and colleagues (2012) made participants believe that they could be the initiator of an auditory tone or that a confederate presented as a genuine participant was the initiator of this tone. In both cases, the tone was always triggered by the participants' actions. The authors found that participants judged the loudness of the tone lower when they were induced the false thought that they were not the initiator of the sensory effect compared to when they believed that they were effectively the initiator of the sensory effect (*Desantis, Weiss, Schütz-Bosbach, & Waszak, 2012*). These results suggest a top-down dissociation in the processing self versus other outcomes.

Regarding the side of action co-representation, it has been argued that during the observation of other-generated actions, the observer's action/observation matching system simulates the motor command allowing the simulation content to be used to predict the consequences of the action, improving action control or implicit action understanding (*Kilner, Friston, & Frith, 2007*; *Pacherie & Dokic, 2006; Picard & Friston, 2014*). Such cognitive representation of other-generated actions into one's own motor system has been investigated using a cerebral marker of action control mechanisms such as the cerebral P3 component (*Sebanz, Knoblich, Prinz, & Wascher, 2006*). The P3 component is a late positive ERP peaking around 300 ms after stimulus onset and which the amplitude has been showed to be enhanced by action control mechanisms such as response inhibition on

no-go trials, that is when participants had to hold back an automatic response activation (*Bokura, Yamaguchi, & Falkenstein, 2001*; *Tekok-Kilic, Shucard, & Shucard, 2001*). In the specific case of a joint Simon task, the principle consisted in that if the co-agent's action was represented into the participant's motor system during a no-go trial, then more inhibitory control was needed in order to suppress this activated representation compared to when there was no action co-representation (because there was nothing to suppress). Thus, the amplitude of P3 component should be more pronounced in conditions where action co-representation is elicited compared to when such a mechanism was not engaged. Consistently, Sebanz and colleagues (2006) demonstrated that the amplitude of P3 component was more pronounced on no-go trials when participants performed a joint go/no-go Simon task with another human compared to the no-go trials when they performed a single go/no-go Simon task, suggesting that relevant actions for a partner during a joint task are co-represented in the co-agent's motor system as they needed to be inhibited (*Sebanz, Knoblich, Prinz, & Wascher, 2006*). Several authors proposed that this kind of motor simulation supports our understanding of intentions of others (*Iacoboni et al., 2005; Rizzolatti, Fogassi, & Gallese, 2001*), enhancing the observer's sense of vicarious agency.

We ran an EEG study where participants had to perform a Social Simon response time task jointly with a human (*joint HH* condition) or desktop computer (*joint HM* condition) co-agent. In a passive observation task, participants were asked to merely observe a believed human co-agent performing the Simon task alone (*passive observation H* condition). Accurate target detections triggered an auditory tone after a certain delay. The amplitude of N1 component during the perception of other-generated tones served as an index of participants' implicit sense of vicarious agency as indicated by sensory attenuation phenomenon (*Poonian, Mcfadyen, Ogden, & Cunnington, 2015*). The amplitude of P3 component during the vision of the co-agent's target served as an indicator of participants' action inhibition mechanisms and thus informed us about action co-representation phenomenon during the tasks (*Sebanz, Knoblich, Prinz, & Wascher, 2006*).

Event related potentials predictions:

**N1 component**

Concerning participants' implicit sense of agency, we hypothesized that self-generated outcomes would induce a stronger sense of agency compared to outcomes that have been generated by an external source. Hence, we expected to observe a reduced N1 activity during the perception of self-generated auditory tones compared to the perception of other-generated tones (regardless the nature of the other, i.e., human or machine).

At the same time, we predicted that during the joint tasks, participants would have an implicit sense of vicarious agency for the human-generated actions but not the computer-generated actions. Hence, we expected that N1 activity would be reduced during the perception of the other-generated tones in the *joint HH* condition compared to the *joint HM* condition.

In addition, we expected that participants' implicit sense of vicarious agency for the human-generated actions would be stronger during the joint task than during the mere passive observation task. Hence, we expected that N1 activity would be reduced during the perception of the other-generated tones in the *joint HH* condition compared to the *passive observation H* condition.

**P3 component**

Concerning action co-representation, we hypothesized that in the joint tasks, participants would have to inhibit the active representations of their human co-agent's but not the computer-generated actions. Moreover, during the passive observation of the other human actions, as no behavioral responses were requested from the participants, no inhibitory mechanisms should be expected. Hence, we expected to observe a greater P3 activity on the co-agent's trials in the *joint HH* condition compared to the *joint HM* condition, and a greater P3 activity on the co-agent's trials in the *joint HM* condition compared to the *passive observation H* condition.

**Link between N1 component and P3 component**

We hypothesized that individuals' implicit sense of vicarious agency as indicated by a diminution of N1 amplitude, was linked to their ability to co-represent their co-agent action as indicated by an enhancement of P3 amplitude. Hence, we expected to observe a negative correlation between N1 amplitude and P3 amplitude.

### 3.2.2. Method

a) Ethic statement

This study was approved by the institutional ethical research committee of the Paris Descartes University (France). The investigation was carried out in accordance with the Declaration of Helsinki and 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.

b) Participants

Twenty naïve participants took part in the experiment (18 women, mean of age = 22.45, SD of age = 4.19). None of them had cognitive or neurological disorders. All the participants had normal or corrected-to-normal vision. Because of an insufficient number of good-quality epochs for ERPs analyses, 3 participants were excluded from analyses leading to a final sample of 17 participants. Participants were given a 30 euros gift card for their participation in the experiment.

c) Material and stimuli

Participants were sitting about 46 cm from a computer screen with a refresh rate resolution of 100 Hz.

Stimulus presentation was controlled using E-prime software (2.0 version). Visual stimuli consisted of four arrows presented in the center of the screen, two blue and two yellow, each pointed either toward the left or the right. Two auditory sin wave sound presented via two speakers placed behind the participant, were used during the experiment

as the effect of the participant/co-agent's detection action for measuring sensory attenuation (1000 Hz, 150 ms duration) or as an error feedback (220 Hz, 150 ms duration).

During the training sessions, the other-generated target detections occurred at a random delay from 200 ms to 400 ms. During the testing sessions, the other-generated target detection were always triggered by an implemented program in the computer and were defined based on the participants' behavior during the training sessions. Indeed, we used the participants' RTs during the training to define the other-generated target detection onsets during the testing phase.

A 64 electrodes EEG system (actiCHamps) was used to record the participants' brain electrical activity during the testing session. The sixty-four active electrodes were fixed on the head surface with the help of a suitable elastic cap (Acticap), according to the international 10/20 electrode positioning. Two external electrodes positioned at the external side of the left and right eyes were used to record the electro-oculogram (EOG) for the purpose of artifact detection. Two active electrodes, TP9 and TP10, that were not relevant for our analyses, were used to record the left and right mastoid signal for which average activity was used as a reference.

EEG raw signal was recorded using Brain Vision Recording software (1.20.0801 version).

d) Procedure

Participants were sitting in front of a screen and had to detect as quickly and as accurately as possible colored arrows that could point either to the left or to the right. This task could be performed jointly with an unseen human co-agent (*joint HH* condition) or an algorithm implemented in a desktop computer (*joint HM* condition) or consisted in observing passively the task performed by an unseen human co-agent (*passive observation H* condition).

In the *joint HH* and *passive observation H* conditions, participants were induced the thought that the *other*-generated actions were performed by a confederate presented as a genuine participant. Instead, in the *joint HM* condition, participants were induced the thought that the *other*-generated actions were generated by the algorithm implemented in

the desktop computer. In facts, the automated program implemented in the computer triggered the *other*-generated actions in all the three experimental conditions.

During the *joint HH* and *joint HM* conditions, participants were sitting in front of a screen located on the right side of the supposed co-agent (human or machine) located in front of another screen and hidden from vision with a plank to allow all experimental conditions being visually identical (see **Figure 9** for a view of the experimental setup). They were told that the two screens were connected so that they could see the same information as their co-agent. Participants were asked to execute a right key press with their right index finger when presented with a yellow arrow, independently of its pointing direction. When presented with a blue arrow, participants were asked to not react, and simply observe their co-agent's detection. Indeed, they were told that their co-agent would have to detect the blues arrows.

In the *passive observation H* condition, participants were sitting in front of a screen located on the right side of a supposed human co-agent located in front of another screen and hidden from vision. They had to observe the confederate detecting the blue arrows. Participants were asked to not react to the yellow arrows.

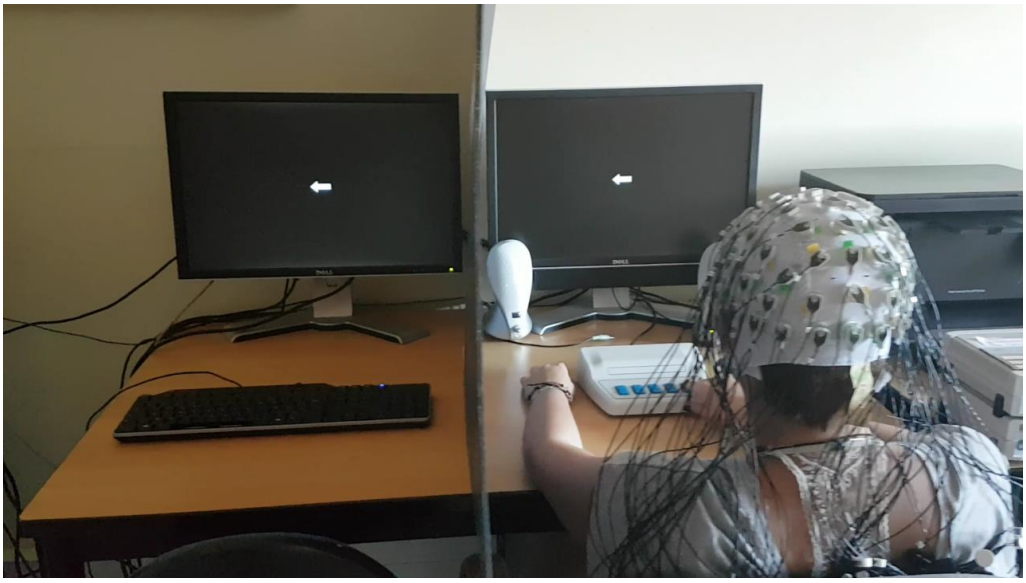

**Figure 9**. Experimental setup during the *joint HM* condition as an example. Participants were located on the right side of their co-agent, hidden from vision with a plank.

In order to reinforce participants' believe, the confederate was sitting on the left side of the participants during the training sessions of the *joint HH* and *passive observation H* conditions and was fully visible. Thereafter, during the testing session, a plank was placed between the two individuals and the blue targets detection was computer triggered. In the *joint HM* condition, there was nobody next to the participant during the training but we placed the plank on the left side of the participants during the testing session so that all the three experimental conditions were visually identical during the testing sessions

Each trial started with a baseline period that lasted for 750 ms, then a fixation cross that appeared at the center of the screen during 500 ms. Thereafter, the target (i.e., the blue or yellow arrow) immediately appeared on the screen. When the target was a yellow arrow, participants had at most 400 ms to press their response key otherwise an error message appeared and the trial was canceled. Correct target detections were followed by an auditory tone (1000 Hz, 150 ms) presented after the response at Stimulus Onset Asynchronies (SOA) of 400 ms. Unlike the joint conditions, in the *passive observation H* condition, participants were asked to not react to the yellow arrows. However, the participants nevertheless were presented the tone after a delay that was unpredictable. A black screen followed the presentation of the tone for 400 ms, which served for N1 component analysis. When participants performed false alarms (i.e., when they made keypresses during the presentation of their co-agent's targets), a different auditory tone (220 Hz, 150 ms) was delivered immediately after the wrong keypresses. Each trial ended with a jitter that lasted a random duration between 300 ms to 500 ms (see **Figure 10**).

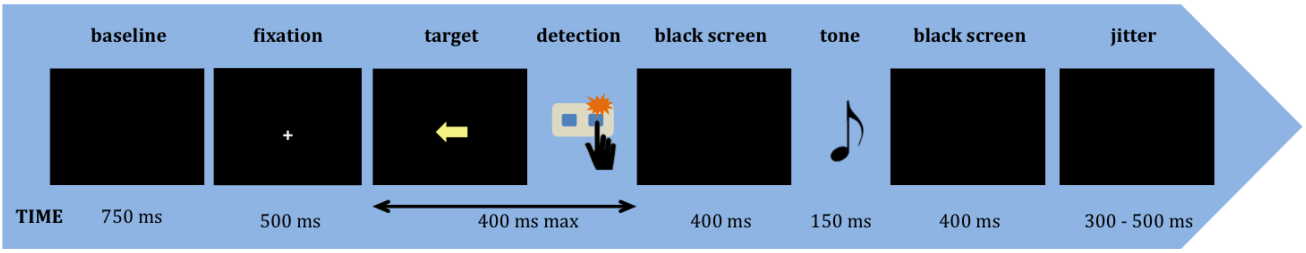


**Figure 10**. Trial timeline. The trial started with a baseline that last for 750 ms followed by a fixation cross for 500 ms. Then, the target appeared (e.g. a yellow arrow pointed toward the left) and participants had 400 ms maximum to detect the target (e.g. to perform a right keypress). Accurate target detection triggered an auditory tone 400 ms after and that was followed by a black screen for 400 ms. The trial ended by a jitter that lasted a random duration between 300 ms to 500 ms.

At the end of the experiment, participants were debriefed in order to ensure that they were not aware about the belief manipulation. There were asked the following questions: (a) “Do you have any comment regarding the experiment?”; (b) “Did you notice any differences in the co-agent performance regarding it was another human or the desktop computer? If yes, please explain”; (c) “Have you got the impression that your co-agent did not perform the task with you actually?”; and (d) “In fact, nobody was performing the task with you and your co-agent detections were always triggered by the computer. Did you suspect that?”.

The experimental conditions were tested using a within-subject design and the order was counterbalanced between participants. Trials were coded as *congruent* when the arrow pointed on the side of the participant’s response key and as *incongruent* when the arrow pointed on the opposite side of the participant’s response key. Our factors were 3 Action context (*joint HH*, *joint HM*, *passive observation H*) x 2 Targets (*blue arrow*, *yellow arrow*) × 2 Congruency levels (*congruent*, *incongruent*) × 100 repetitions leading to 1200 trials.

### 3.2.3. Data analyses

Our dependent measures were the mean target detection Response Times (RTs) and the mean amplitudes of N1 and P3 component around their respective peaks. Statistical analyses were performed with R software (3.3.1 version). Extreme values (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. The significance level was set at $\alpha = .05$. Post-hocs comparisons were corrected using the false discovery rate correction (*Benjamini & Hochberg, 1995*). EEG analyses were performed under EEGLAB Matlab toolbox (*Delorme & Makeig, 2004*). To distinguish participants’ trials from the co-agent’s trials, participants’ trials were labeled *self* trials and the co-agents’ trials were labeled *other* trials.

a) Manipulation of belief

This analysis aimed at verifying that the belief induction (i.e., to perform the task with a human co-agent or with a machine co-agent) was a success for all participants. The

answers of the participants during the debriefing were reported by writing and analyzed by two independent raters. Those raters had to judge whether participants suspected that they had been interacting with a desktop computer. If the two raters disagreed, the judgment of a third rater was requested. The answers participants gave are listed in **Appendix 1**.

b) Social Simon Effect (SSE)

This analysis was based exclusively on the participants' RTs gathered in the conditions in which participants performed an action (i.e., the *self* trials). The analysis aimed at quantifying the SSE when participants interacted with a human co-agent versus a desktop computer. We conducted a within-subjects 2 x 2 ANOVA on participants' RTs with Co-agent (*human*, *computer*) and Congruency (*congruent*, *incongruent*) as factors.

c) Event related potentials analyses

Raw EEG signal was re-referenced to the left and right mastoid signal. Low-pass (48 Hz) and high-pass (0.01 Hz) non-causal filters were applied to the data. Epochs began at 1000 ms before the onset of the target and continued to 1700 ms, and were baseline-corrected using a time window from -1000 ms to -800 ms. After visual inspection of the signal, epochs were rejected if they contained artifacts caused by poor signal or unmandatory movement (artifact rejection threshold = ± 100 µV). Blink artifacts were corrected using Independent Component Analysis (ICA).

After data preprocessing, different types of epochs were selected. For N1 component, data were epoched from -100 ms to 400 ms around the auditory tone. For P3 component, data were epoched from -100 ms to 600 ms around the target. For both components, epochs were baseline-corrected using a time window from -100 ms to 0 ms. Indeed, epochs were baselined again to remove eventual drifts given that the first baseline was applied very far back with respect to both the target and the sound onset.

N1 component

For N1 component, statistical analyses were conducted across 6 centroparietal electrodes (C3, Cz, C4, CP3, CPz and CP4) based existing literature (*Poonian, McFadyen,*

*Ogden, & Cunnington, 2015*). Note that in Poonian and colleagues (2015)' study, frontal electrodes (F3, Fz and F4) were also included in the author's analyses but we decided to not select them for this current analysis after visual inspection of the signal because of a weak N1 activity at the frontal site. For each participant, each experimental condition, and for each electrode separately, the mean voltage during a 24 ms time window centered around the peak (*Poonian, McFadyen, Ogden, & Cunnington, 2015*) of the grand-averaged waveforms for all experimental conditions confounded was used (i.e., from 133 ms to 157 ms after the onset of the auditory tone).

We aimed at investigate differences in the cerebral processing of the tones in regards with the generator (self or external) and the nature of the agent (human or artificial automated) that have triggered sensory consequence (*Poonian, McFadyen, Ogden, & Cunnington,* 2015). The *passive observation H* was not included in the statistical analysis because there were no *self* trials in this experimental condition. To investigate potential effect lateralization, electrodes were grouped into three levels of Localization of electrodes (*left*, *middle*, *right*) according to their positioning on the participants' scalp. The *left* electrodes included C3 and CP3, the *middle* electrodes included Cz and CPz, and the *right* electrodes included C4 and CP4. We conducted a within-subjects 2 x 2 x 2 x 3 ANOVA on N1 mean amplitude with Agent (*self*, *other*), Congruency (*congruent*, *incongruent*), Action context (*joint HH*, *joint HM*) and Localization of electrodes (*left*, *middle*, *right*) as factors.

P3 component

For P3 component, statistical analyses were conducted across 9 centroparietal electrodes (C1, Cz, C2, CP1, CPz, CP2, P1, Pz and P2), based on existing literature (*Polich, 2007*). For each participant, each experimental condition, and each electrode separately, the mean voltage during a 100 ms time window centered around the peak (*Sebanz, Knoblich, Prinz, & Wascher,* 2006) of the grand-averaged waveforms for all experimental conditions confounded was used (i.e., from 304 ms to 404 ms after the onset of the target).

As we were interested in action control mechanisms during the perception of stimuli that referred to the co-agent's action and consistently with previous work, the *self* trials were excluded from this analysis (*Sebanz, Knoblich, Prinz, & Wascher, 2006*). To investigate potential effect lateralization, electrodes were grouped into three levels of Localization of

electrodes (*left*, *middle*, *right*) according to their positioning on the participants' scalp. The *left* electrodes included C1, CP1 and P1, the *middle* electrodes included Cz, CPz and Pz, and the *right* electrodes included C2, CP2 and P2. Hence, we conducted a within-subjects 2 x 3 x 3 ANOVA on P3 mean voltage on the *other* trials with Congruency (*congruent*, *incongruent*), Action context (*joint HH*, *joint HM*, *passive observation H*) and Localization of electrodes (*left*, *middle*, *right*) as factors.

### 3.2.4 Results

a) Manipulation of belief

The answers of the participants during the debriefing were reported by writing and analyzed by two independent raters. There was perfect agreement between the two raters for all participants: none of the participants suspected that the *other*-generated actions were always generated by the algorithm implemented in the desktop computer.

b) Social Simon Effect (SSE)

We examined the SSE during a joint task human and with an algorithm implemented in a desktop computer. This analysis was based exclusively on the participants' RTs gathered in the conditions in which participants performed an action (i.e., the *self* trials). 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 distribution deviated from normality (all $W > 0.90$ and all $p > .10$). We then computed a within-subjects 2 x 2 ANOVA on the RTs with the factors Co-agent (*human*, *computer*) and Congruency (*congruent*, *incongruent*). We found a significant main effect of Congruency ($F(1,16) = 11.19$; $p = .004$) indicating longer mean RTs on *incongruent* trials compared to *congruent* trials (respectively 291.54 ms (SD = 19.65) and 287.74 ms (SD = 18.59), see **Figure 11**). No significant main effect of Co-agent ($F(1,16) = .43$; $p = .52$, *ns*) nor significant Co-agent x Congruency interaction ($F(1,16) = .16$; $p = .70$, *ns*) were found. Hence, we observed the classical Simon Effect when participants performed the joint task with another human agent and when they performed the task with the algorithm implemented in the desktop computer as a co-agent.

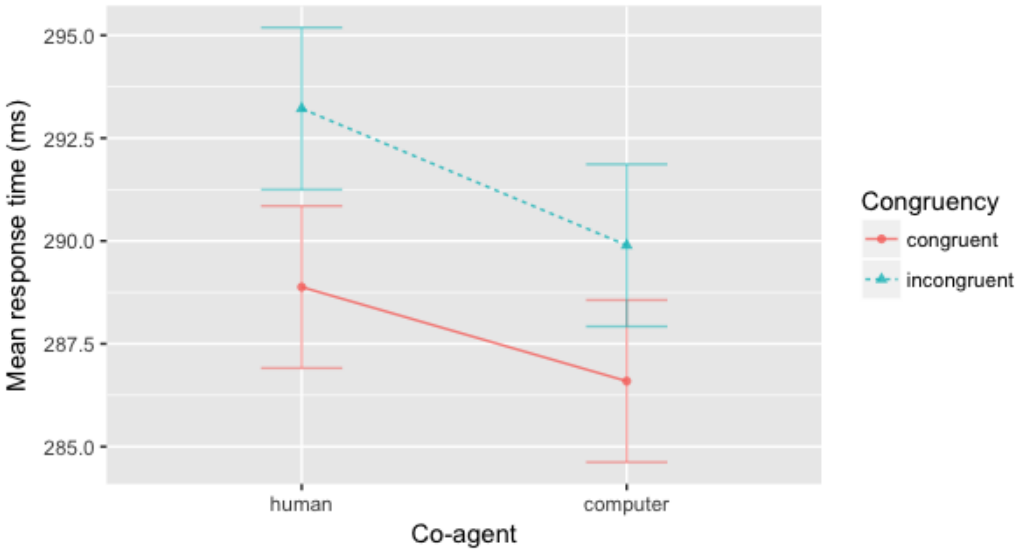


**Figure 11**. Main effect of Congruency on participants' mean response times. Error bars represent standard errors.

To investigate more in details the absence of significant interaction between the type of Co-agent and the Congruency, we ran additional analyses. Specifically, we hypothesized that the belief induction during the first joint task that participant had performed (e.g., the *Joint HH* condition) could potentially be transferred in time and thereby influence the participants' RTs in the other subsequent joint task participants performed (e.g. the *Joint HM* condition). To the purpose of further investigations, participants were split into two groups, G1 and G2, regarding the order of the joint conditions they completed. Participants who performed the *Joint HH* condition before the *Joint HM* condition were placed in the G1 group. Reversely, participants who had performed the *Joint HM* condition first, and the *Joint HM* condition after, were placed in the G2 group. Hence, we computed a 2 x 2 x 2 mixed ANOVA on the RTs with Co-agent (*human*, *computer*) and Congruency (*congruent*, *incongruent*) as within-factors, and Group (*G1*, *G2*) as a between-factor. We found a significant Congruency x Group interaction ($F(1,16) = 4.65$; $p = .046$; see **Figure 12**). However, post-hoc comparisons investigating this interaction revealed no significant differences (all $p > .10$, *ns*). This might be due to the small sample sizes in each modality of the between-factor in comparison with the sample size of the within-factors.

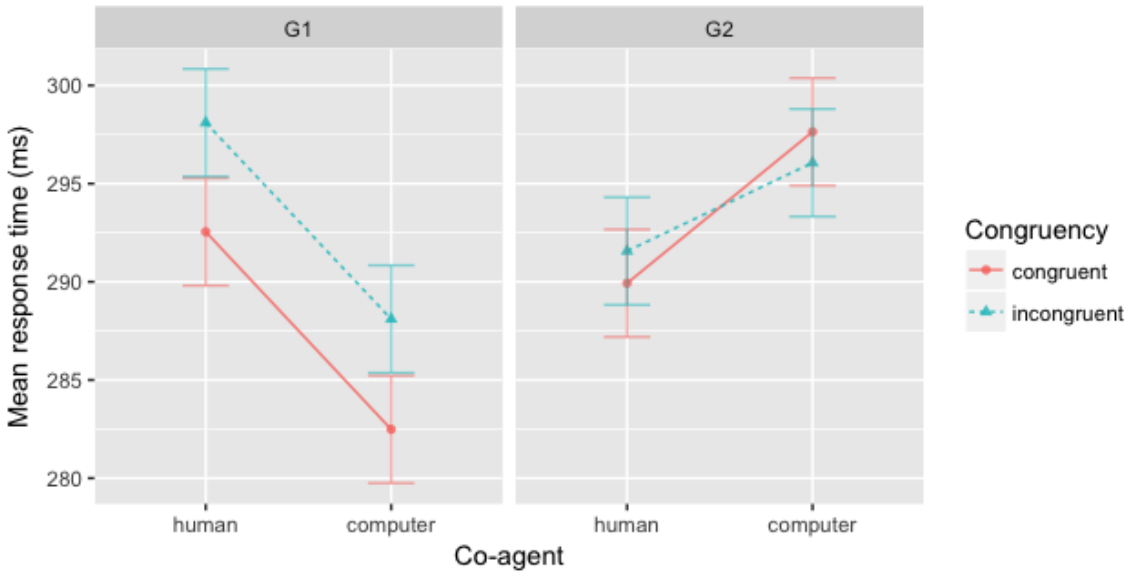


**Figure 12**. Interaction between Group and Congruency on participants' mean response times. Error bars represent standard errors.

Considering these elements, we computed separated within-subject 2 x 2 ANOVAs on participants' RTs for each group, with Co-agent (*human*, *computer*) and Congruency (*congruent*, *incongruent*) as factors.

For participants in G1, we found a significant main effect Congruency on RTs ($F(1,8) = 17.47$; $p = .003$) indicating longer mean RTs on *incongruent* trials compared to *congruent* trials (respectively 293.10 (SD = 22.75) and 287.52 ms (SD = 20.76)). No significant main effect of Co-agent ($F(1,8) = 2.15$; $p = .18$, *ns*) or Co-agent x Congruency interaction ($F(1,8) = .0003$; $p = .98$, *ns*) were found.

For participants in G2, no significant main effects of Co-agent ($F(1,8) = 4.07$; $p = .07$, *ns*) or Congruency ($F(1,8) = .0001$; $p = .98$, *ns*), neither a significant Co-agent x Congruency interaction ($F(1,8) = .62$; $p = .45$, *ns*) were found on participants' RTs.

Hence, these additional analyses suggested that when participants started the experiment believing that they would cooperate in a first task with another human, they exhibited a SSE all along the experiment, even if a desktop computer replaced their co-agent in the subsequent task. On the contrary, when participants started the experiment believing that they would cooperate in a first task with an artificial automated system, they did not manifest a SSE at all, even if their co-agent was a human being.

c) Event related potentials analyses

N1 component

We aimed at investigating differences in the cerebral processing of the tones in regards with the generator (self or external) and the nature of the agent (human or artificial automated) that have triggered sensory consequence (*Poonian, McFadyen, Ogden, & Cunnington,* 2015). We conducted a within-subjects 2 x 2 x 2 x 3 ANOVA on N1 mean amplitude around the peak with Agent (*self*, *other*), Congruency (*congruent*, *incongruent*), Action context (*joint HH*, *joint HM*) and Localization of electrodes (*left*, *middle*, *right*) as factors. We found a significant main effect of Agent on N1 mean amplitude around the peak (*HFe* = .71, *p* = .045) indicating that N1 mean activity was greater when the auditory tone was *self*-generated compared to when it was *other*-generated (respectively -4.72 µV (SD = 4.13) and -3.24 µV (SD = 3.19); *p* = .045; see **Figure 13**). We also found a significant main effect of Localization of electrodes (*HFe* = .69; *p* < *.001*). No other main effects or interactions were significant (all *p* > .10).

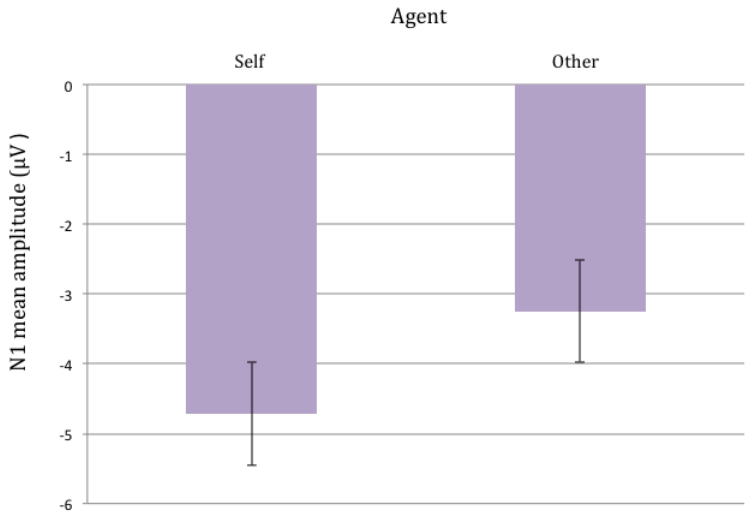


**Figure 13.** Main effect of Agent on the mean amplitude of N1 component.

Post-hoc comparisons investigating the main effect of Localization of electrodes on N1 mean amplitude revealed that N1 mean activity was greater at the *left* electrodes compared to the *right* electrodes (respectively -4.58 µV (SD = 3.67 ) and -2.83 µV (SD =

3.75); $p < .001$). Similarly, N1 mean activity was greater at the *middle* electrodes (-4.54 µV (SD = 3.61)) compared to the *right* electrodes ($p < .001$). However, there was no difference on N1 mean amplitude in the *left* and the *middle* electrodes ($p = .78$, *ns*). In sum, N1 mean amplitude was greater in the *left* and the *middle* electrodes compared to the *right* electrodes.

P3 component

We aimed at investigating action control mechanisms during the perception of visual stimuli that referred to the co-agent's action (*Sebanz, Knoblich, Prinz, & Wascher, 2006*). We conducted a within-subjects 2 x 3 x 3 ANOVA on P3 mean amplitude around the peak on the *other* trials with Congruency (*congruent*, *incongruent*), Action context (*joint HH*, *joint HM*, *passive observation H*) and Localization of electrodes (*left*, *middle*, *right*) as factors. We found a significant main effect of Action context ($HFe = .60$, $p < .001$) on P3 mean amplitude around the peak. No other main effects or interactions were significant (all $p > .10$).

Post-hoc comparisons investigating the main effect of Action context on P3 mean amplitude around the peak revealed that P3 activity was greater for the *other* trials in the *joint HH* context compared to the *joint HM* context (respectively 13.30 µV (SD = 8.32) and 11.97 µV (SD = 7.58); $p < .001$) and to the *passive observation H* context (3.39 µV (SD = 4.56), $p < .001$). Similarly P3 activity was greater for the *other* trials in the *joint HM* context compared to the *passive observation H* context ($p < .001$; see **Figure 14**). Hence, P3 mean amplitude evoked during the perception of visual stimuli that referred to the co-agent's action was at the highest when participants were performed a joint task with another human, then decreased when participants were in a joint task with the desktop computer, to finally drastically decrease when they were passively observing another human agent doing the task.

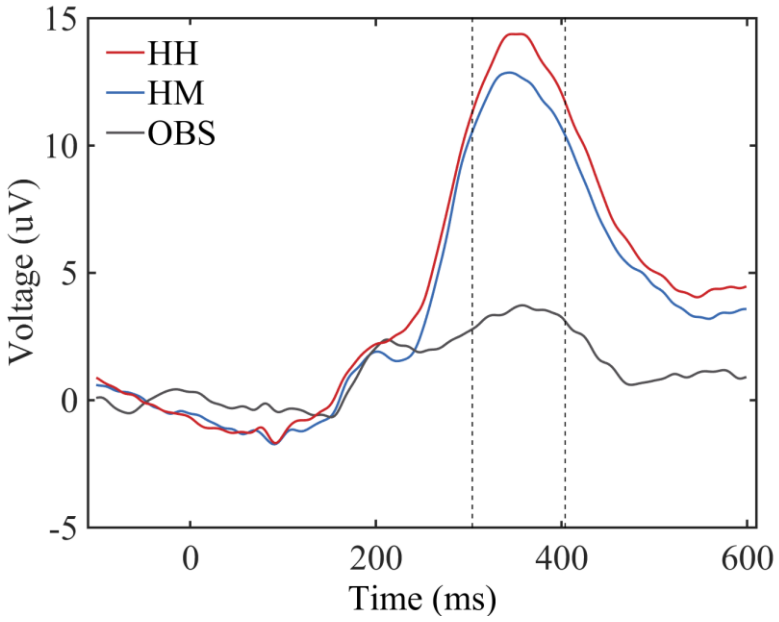


**Figure 14**. Main effect of Action context on the mean amplitude of P3 component

### 3.2.5 Discussion of the experiment

In this study, our aim was to investigate the link between individuals' implicit sense of vicarious agency and action co-representation mechanisms during a joint task with a human co-agent versus an automated artificial system such as a desktop computer. Participants were asked to perform a joint Simon task with a co-agent that was believed to be either another human or an algorithm implemented in a desktop computer. In another condition, they were asked to merely observe a believed human agent performing the Simon task alone. Accurate target detection triggered an auditory tone after a randomized delay. Individuals' sense of agency was estimated using sensory attenuation phenomenon of the cerebral N1 component during the perception of the tones (*Poonian, Mcfadyen, Ogden, & Cunnington, 2015*). To assess action co-representation, we used the amplitude of the P3 component during the vision of the co-agent's target (*Sebanz, Knoblich, Prinz, & Wascher, 2006*).

We found the classical SSE with longer response times in *incongruent* trials when participants performed a double target detection task jointly with a *human* co-agent and, surprisingly, also when they performed the joint task with the desktop *computer*. These results faced the findings of **Experiment 1** that showed that individuals exhibited a SSE only when they partnered with a *human* co-agent. However, even if the task participants ran

was the same in the two experiments, the experimental setups were different. Indeed, in the present experiment, participants did not have direct vision of the co-agents whereas in the first experiment their co-agents were not hidden. Hence, it is possible that at the beginning of the experiment, participants made a substantial effort to conceive a mental representation of their co-agent doing the task. This first construct might be long-lasting, which biased participants' performance. This assumption is consistent with additional results that suggested that when participants were induced the belief that they would cooperate with a *human* co-agent as a first task, they exhibited a SSE in this precise experimental condition but also during the subsequent experimental condition where a desktop *computer* replaced their co-agent. By contrast, participants did not manifest a SSE at all when they started the experiment believing that they would cooperate in a first task with a *computer*, and this absence of SSE lasted even when a believed *human* being replaced their co-agent during the subsequent experimental condition. As a consequence, it could be said that when direct visual information about their co-agent is missing, participants might build a schematic representation of this co-agent doing the task. In the case of a *human* as first co-agent, participants expectedly co-represented the action of their partner (demonstrated by the SSE) but the schematic representation of another human being jointly acting remained and participants did demonstrate a SSE in the case of the joint task with the *computer*. As well, when participants were induced that they would have a *computer* as first co-agent, they expectedly did not co-represent the action of their partner (demonstrated by no SSE, *Tsai, Kuo, Hung, & Tzeng, 2008*) and as the schematic representation of the machine jointly acting remained, participants did not demonstrated a SSE in the case of the subsequent joint task with the *human*.

Nevertheless, when investigating action co-representation at the cerebral level, we found evidence that other-generated actions were differently processed regarding the nature of the co-agent that participants were acting. This suggests that the processing of the type of co-agent participants' were partnered with was distinctly manifested at the behavioral level and at the cerebral level. Indeed, we found that when participants were presented a target that requested a *human* co-agent behavior (*joint HH* condition), the amplitude of P3 component was greater than when participants were presented a target that requested a machine co-agent (*joint HM* condition). Because the amplitude of P3 component has been considered as a marker of response inhibition (*Bokura, Yamaguchi, & Falkenstein, 2001*; *Tekok-Kilic, Shucard, & Shucard, 2001*), it can be admit that during the joint task with the *human*, participants had to hold back the co-representation of their co-

agent' expected action, explaining the greater P3 activity. This results corroborate the findings of Sebanz and colleaues (2006) that investigated P3 component during a joint Simon task between human peers and showed a similar P3 enhancement in this social context (*Sebanz, Knoblich, Prinz, & Wascher, 2006*). Instead, during the joint task with the *computer*, it can be said that participants had no active co-representation of the machine action, and as a consequence, they did not have to bring into play – or at least, in a lesser extent – inhibitory mechanisms. This could explain why the amplitude of P3 component was weaker than in the human joint task. Finally, during the passive observation of the human actions, considering that no inhibitory mechanisms were necessary, we found that the amplitude of P3 component was at it weakest. Taken together, our results suggested that during the joint tasks, participants co-represented into their own motor system the actions of the *human* co-agent whereas such a mechanism was not observed for the *computer* co-agent. However, in this study, a clear dissociation in co-representational abilities for human-generated actions and machine-generated actions was not evident at the behavioral level.

Concerning participants' implicit sense of agency, we found that self-generated auditory tones induced a greater amplitude of N1 component compared to auditory tone that were externally generated, which was the reverse pattern as that expected. Indeed, according to the CM, self-generated sensory effects should be perceptively diminished compared to external generated sensory effects, which should attest a stronger implicit sense of agency in the former case (*Blakemore, Wolpert, & Frith, 2000*). In a backward flow, we found that self-generated action outcomes lead to a stronger cerebral processing than other-generated outcomes. This finding is however consistent with previous work investigating individuals' sense of agency under various contexts (*Caspar, Christensen, Cleeremans, & Haggard, 2016*). In their study, Caspar and colleagues (2016) asked participants to perform self-generated keypresses either deliberately (free-choice condition) or under the expressive order of the experimenter (coercion condition). Participants' keypresses triggered an auditory tone after a variable delay and the cerebral activity of the participants was recorded during the experiment. The authors found the amplitude of N1 component was more pronounced in the free-choice condition than in the coercion condition. As an alternative explanation from the sensory attenuation phenomenon, it has been proposed that outcomes generated under strong agentive experience elicited a more important cerebral processing (*Caspar, Christensen, Cleeremans, & Haggard, 2016*). In a nutshell, we found that self- and external generated auditory tones were differently

perceived at the cerebral level, reflecting separated sensory treatments according to the locus of the generator of the sensory consequence (i.e., internal versus external).

Our results revealed no modulation of the auditory N1 component amplitude between the human joint task and the human-computer joint task, nor for the *self* trials neither for the *other* trials. Hence, we were not able to highlight potential distinctions in participants' sense of vicarious agency according to the nature of the Co-agent they partnered with. These results contrasted with the EEG findings by Loehr (2013) who showed that participants' implicit sense of vicarious agency was the highest in the context of a joint task with another human compared to the passive observation of human-generated or computer-generated actions contexts (*Loehr, 2013*). However, in Loehr (2013)'s study, the action outcome (i.e., the auditory tone) in the human-human joint task was always triggered by the actions of the both individuals. Indeed, after a go signal, the two human partners' keypresses were requested in order to trigger the tone. Hence, they had to act together to produce a shared effect, which in turn might reinforce the *joint* aspect of the task. In our study, even if participants performed jointly the Simon task, the target detections were distributed across the two individuals, leading each one being distinctly responsible for a given sensory outcome. In this case, there was no shared intentionality to produce the auditory tones. This could explain why we did not find modulation of the amplitude of the N1 component in the joint task with the human co-agent compared to the computer-human joint task.

Because of the absence of experimental evidence for sensory attenuation phenomenon as indicated by the N1 amplitude decrease for other-generated actions during a joint task a *human* co-agent compared to a *computer* co-agent, correlational analyses between N1 component and P3 *component* were not conducted. Indeed, our initial goal was to put forward that individuals' implicit sense of vicarious agency as indicated by a diminution of N1 amplitude, was linked to their ability to co-represent their co-agent action as indicated by an enhancement of P3 amplitude for *other*-generated actions.

In summary, in this experiment, we showed a self/other distinction in the sensory processing of auditory outcomes. Additionally, we showed at the cerebral level that during a joint task with a human co-agent, participants demonstrated inhibitory mechanisms when they were presented stimuli that requested their co-agent's action, suggesting that they co-represented into their own motor system the actions of the co-agent as their own. Instead, they did not manifest inhibitory cerebral response when they were presented stimuli that requested a computer co-agent response, suggesting that they did not co-represented such

a behavior in their cerebral motor system. Thus, while humans can represent the actions of human partners, they experience difficulties in mirroring and fitting with traditional automaton-generated actions.

Probably, to envisage the others as similar to us is needed in order to map their actions into our cognitive system and to compute a forward model of these other-generated actions. Indeed, several studies showed that this kind of simulation was made with the help of the observer's own motor expertise (*Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005*; *Kuz, et al. 2015*; *Mann, Williams, Ward, & Janelle, 2007*). Given these elements, human-like automated artificial systems, both at a high level and low levels of integration, could be hypothesized to stimulate co-representational mechanisms of machine-generated actions. Thereby, artificial system humanness could enhance both individuals' sense of self-agency and sense of we-agency during human-machine interactions. Hence, we ran a behavioral study (**Experiment 3**) in order to investigate the effect of machines' physical appearances on individuals' sense of self-agency and sense of vicarious agency during a joint task.

## 3.3. Experiment 3: Reinforcing individuals' sense of we-agency during human-machine interactions using high-level contextual cues

### 3.3.1 Context of the experiment

The present study aimed to investigate the development of the experience of agency during a joint task with different types of artificial partners (i.e., human-like or not).

The first objective of this experiment aimed to clarify the transformation of individuals' agentive experience during human-machine interactions. This referred to the transformation according to which during human-human interactions, individuals' sense of self-agency is diminished in favor of a new agentive status, a sense of "we-agency". This sense of "we-agency" has been showed to take into consideration the performance of the dyad as a whole and lead individuals to experience agency as soon as one of the two had performed a goal-directed action (*Dewey, Pacherie & Knoblich, 2004; van der Wel, Sebanz, & Knoblich, 2015*).

Our second objective was to investigate the impact of artificial system humanness during a joint task on both individuals' sense of self-agency and sense of we-agency. Indeed, while the results previously described (**Experiment 1***; Obhi & Hall, 2011b*) suggest that the sense of agency could be altered when interacting with traditional artificial systems (e.g. an algorithm implemented in a desktop computer), very little is known about to what extent the features of this artificial co-agent promote the development of joint agency during human-machine interactions. To our knowledge, no study has explored the impact of artificial system humanness on the experience of agency during a joint task.

Finally, our third objective was to investigate the factors influencing the sense of agency in the particular context of a joint task. Indeed, it has been showed that egocentric sensory predictions were less involved in the development of individuals' sense of agency when they were engaged in a joint action (*Dewey, Pacherie & Knoblich, 2004*). By contrast, individuals' sense of agency seems more correlated with the degree of control exhibited by the team as a whole (*van der Wel, Sebanz, & Knoblich, 2015*). Hence, when considering individuals' sense of agency as a multiple cues dynamic integration (*Moore & Fletcher, 2012*), the respective weight of sensorimotor, perceptual and environmental cues could potentially change regarding the social context of the action.

This study aimed to explore this modulation during a joint task with different types of artificial co-agents. To address this issue, we ran a behavioral study combining a joint Simon task with an intentional binding task. The main advantages of the joint Simon task was that it consisted in congruent and incongruent trials that allow to investigate sensorimotor processes that have been showed to be integrated during individuals' sense of agency construction (e.g. the fluency of action selection, *Chambon & Haggard, 2012*; *Sidarus & Haggard, 2016; Sidarus, Vuorre, Haggard, 2017*). Additionally, the joint Simon task allowed self-generated and other-generated actions that permit to study individuals' sense of self-agency and sense of vicarious agency. We manipulated the nature of the co-agent participants performed the task with, from totally human, human-like artificial automated, to purely artificial automated. Accurate target detection triggered an auditory tone after a certain delay. Participants had to estimate the temporal delay between the target detection and the onset of the tone. This estimate served as an implicit measure of participants' sense of agency (*Haggard, Clark, & Kalogeras, 2002*). We hypothesized a shift from a sense of self-agency to a sense of "we-agency". We also hypothesized that the more similar to the participants the co-agent would be, the more they would exhibit a strong sense of we-agency.

### 3.3.2. Method

a) Ethic statement

This study was approved by the institutional ethical research committee of the Free University of Brussels (Belgium). The investigation was carried out in accordance with the Declaration of Helsinki and 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.

b) Participants

Twenty-eight healthy adults volunteered to take part in the experiment (22 women, 24 right-handed, mean age 23.61 years, SD of age 3.52 years). All participants had normal or corrected-to-normal vision. None of them had prior knowledge about the purpose of the experiment. Each participant was paid 30 euros for their participation in the experiment.

c) Material and stimuli

Two desktop computers were used to allow pairs of participants to run some sessions of the experiment in parallel. Participants were sitting about 46 cm from the computer screens.

Stimulus presentation was controlled using PsychoPy software (2_PY3 version). Visual stimuli consisted of three dots of 0.5 cm diameter, one white, one blue and the other yellow. A sine wave sound (1000 Hz, 200 ms duration), presented via headphones, was used during the experiment as the effect of the participant/co-agent's button press for measuring intentional binding.

In two experimental conditions, participants had to interact with automated artificial systems. The physical features of the automated artificial system have been manipulated so that it could be a servomotor or a full humanoid robot (see **Figure 15**). When the robots executed goal-directed actions, their response times were taken from a normal distribution calculated from the mean and standard deviation of naïve participants' response times computed during **Experiment 1**.

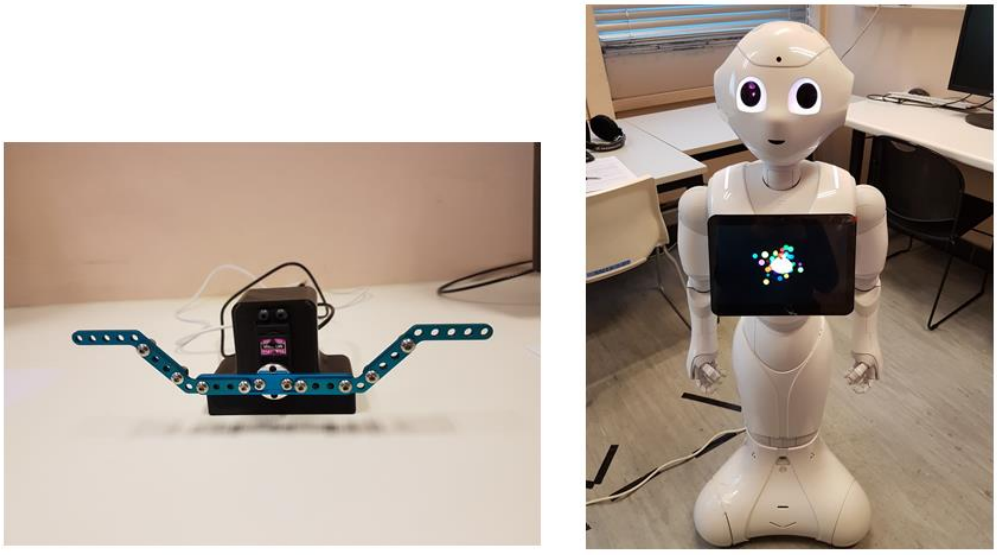

**Figure 15.** Illustration of the two types of robots used during the experiment. The servomotor could execute a keypress with the left or right metallic arm. The humanoid robot (Pepper) could perform a keypress with the left or right fist.

d) Procedure

Participants were sitting in front of one side of a screen and their co-agent in front of the other side. They had to detect, as quickly and as accurately as possible, colored dots that appeared either to the left or to the right side of a central fixation cross. This task could be performed jointly with the servomotor, or with the humanoid robot, or with another naïve participant (see **Figure 16**). When participants partnered with another human, they were matched by gender and handiness.

**Figure 16.** Experimental setups when participants performed the joint task with another human (top), the full humanoid robot (left) and the servomotor (right).

Each trial started with a fixation cross that appeared at the center of the screen during 500 ms followed by the immediate apparition of the target. According to the color of the target (either blue or yellow) participants or their co-agent had at most 1000 ms to press their response key (either left or right) otherwise an error message appeared and the trial

was canceled. Participants were informed of the onset of their own action and the onset of the action of their co-agent by the presentation of a white dot displayed on the target for a duration of 200 ms. Participants were required to fixate the computer screen throughout the experiment and to not look at the actions performed by the co-agent. Correct target detections were followed by an auditory tone presented after the response at one of two possible SOA of 400 ms or 1200 ms. The SOA for a given trial was selected randomly. After the presentation of the sound, 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 17** for a summary). This time interval estimates served as an implicit measure of participants' sense of agency (*Haggard, Clark, & Kalogeras, 2002*).

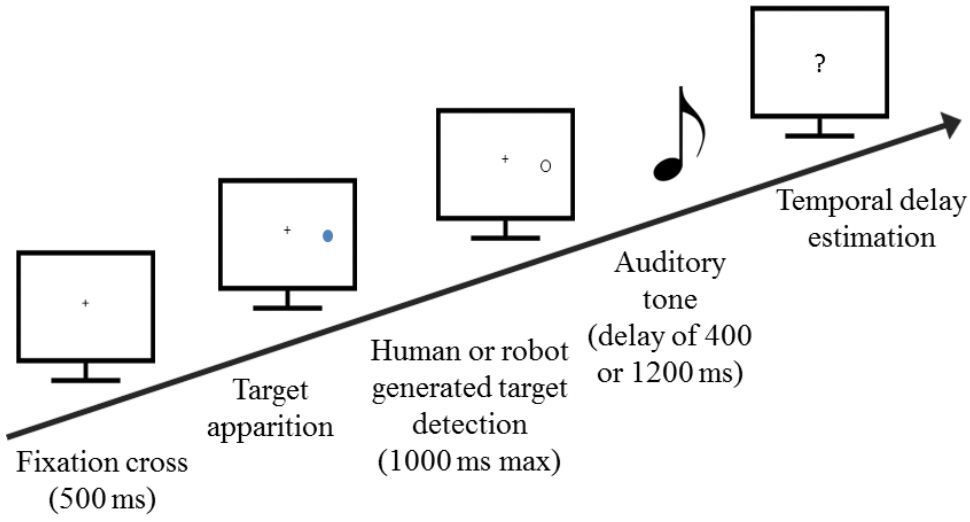


**Figure 17.** Trial timeline. A fixation cross appeared for 500 ms. Then, the target appeared and the agent (the participant or the co-agent) had to detect it before 1000 ms. All target detections were signaled by target that became withe. An auditory tone was generated at a delay of 400 or 1200 ms after target detection. The participant had to report the temporal delay between the onset of the target detection and the onset of the tone.

Participants were trained at the beginning of each experimental condition 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 flashed with a random delay

between 100 ms and 2000 ms. They had to write on a sheet of paper the perceived duration of this interval in milliseconds. Then, they were given the correct delay with a visual feedback in order to accurately recalibrate their internal clock. This training session consisted in 25 trials. Thereafter, participants performed 16 trials of the given experimental condition as training. The goal was to familiarize them with the task so that they would associate their key press with the following auditory tones.

The experimental conditions were tested using a within-subject design and the order of conditions that each participant completed was counterbalanced. The mapping color of the target dot/response key was counterbalanced across the participants but remained the same throughout all the experiment for a given participant. Trials were 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. Participants completed a total of 720 trials: 3 Co-agents (*human*, *humanoid*, *servomotor*) ×2 Targets (*blue dot*, *yellow dot*) × 2 Congruency levels (*congruent*, *incongruent*) × 2 Delays (*400*, *1200*) × 30 trials).

### 3.3.3. Data analyses

Our dependent measures were the mean target detection Response Times (RTs) and the mean perceived action-tone interval. Statistical analyses were performed with R software (3.3.1 version). Extreme values (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. The significance level was set at $\alpha$=.05. Post-hocs comparisons were made using the false discovery rate correction (*Benjamini & Hochberg, 1995*). To distinguish participants' trials from the co-agent's trials, participants' trials were labeled *self* trials and the co-agents' trials were labeled *other* trials.

a) Social Simon Effect (SSE)

This analysis was based exclusively on the data gathered in the conditions in which participants performed an action (i.e., the *self* trials). The analysis aimed at quantifying the SSE according to the nature of the co-agent participants interacted with. We conducted a

within-subjects ANOVA on participants' mean RTs with Co-agent (*human*, *humanoid*, *servomotor*) and Congruency (*congruent*, *incongruent*) as within-subject factors and Hand (*dominant*, *reverse*) as between-subject factor. The Hand factor was introduced in the analysis to investigate whether to perform the task with the dominant hand or not could have an influence on the SSE. The Target (*blue dot*, *yellow dot*) factor was not included in the ANOVA because the SSE does not rely on the target identity but rather on the congruency between the location of the target and the location of the response key. The Delay (*400*, *1200*) factor was irrelevant for the analysis as the auditory tone was produced after participants' response and therefore could not influence their RTs.

b) Engagement in the task

This analysis aimed at examining if the participants' engagement in the task varied according to the type of Co-agent (*human*, *humanoid*, *servomotor*). Hence, we analyzed participants' RTs only on *congruent* trials because *incongruent* trials involved interference mechanisms that could induce biases for the engagement analysis. Specifically, we hypothesized that the social feature of the co-agent will increase participants' engagement in the task. We computed separated linear models for each participant from their mean RTs on *congruent* trials as a function of the type of Co-agent gradually arranged from totally social to non-social. We then extracted the slopes for each participant's model and compared them to the norm mu = 0. Significant positive slopes were supposed to mean that participants' RTs were faster when they partnered with a social agent as opposed to a non-social agent.

c) Temporal interval estimations

This analysis aimed at investigating the influence of the co-agent on participants' agency experience. We conducted a within-subjects ANOVA on participants' mean temporal estimations with Co-agent (*human*, *humanoid*, *servomotor*), Congruency (*congruent*, *incongruent*) and Agent (*self*, *other*) as factors. We did not include the action–tone Delay (*400*, *1200*) as a separate factor. Indeed, in the present experiment we were interested in the way the social context influences action-tone interval estimation in general, rather than its influence on action-tone interval for different temporal intervals.

### 3.3.4. Results

a) Social Simon Effect

We examined the SSE during a joint task according to the nature of the Co-agent (*human*, *humanoid*, *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 distribution deviated from normality (all $W > 0.90$ and all $p > .10$). We then computed a 3 × 2 x 2 ANOVA with the factors Co-agent (*human*, *humanoid*, *servomotor*), Congruency (*congruent, incongruent*) and Hand (*dominant*, *reverse*). We found a significant main effect of Congruency ($F(1,26) = 43.98$, $p < .001$) indicating longer mean RTs on *incongruent* trials compared to *congruent* trials (respectively 351.73 ms (SD = 41.47) and 337.25 (SD = 44.87)) but no significant main effects of Co-agent ($F(2,52) = 1.90$, $p = .16$) or Hand ($F(1,26) =$ , $p = .85$). We found a significant Co-agent x Congruency interaction ($F(2,52) = 6.53$, $p = .003$; see **Figure 18**).

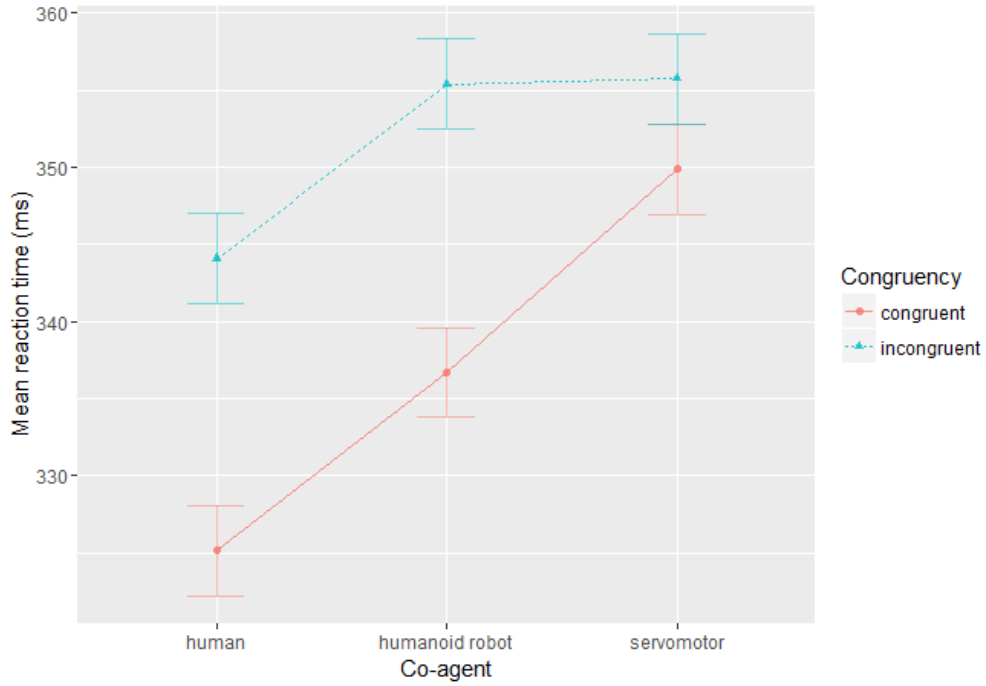


**Figure 18**. Interaction between Congruency and Co-agent on participants' means response times. Error bars represent standard errors.

Post-hoc comparisons investigating this Co-agent x Congruency interaction revealed that mean RTs on *incongruent* trials was significantly longer than the mean RTs on *congruent* trials when the co-agent was a *human* (respectively 344.08 ms (SD = 41.07) and 325.16 ms (SD = 40.36); $p < .001$) and a *humanoid* robot (respectively 355.39 ms (SD = 42.23) and 336.72 ms (SD = 41.60); $p < .001$) but not when it was the *servomotor* (respectively 355.71 ms (SD = 41.84) and 349.86 ms (SD = 50.14), $p = .12$). Hence, we observed a SSE both when participants performed the task with another *human* and with the *humanoid* robot. On the contrary, no SSE was observed when participants interacted with the *servomotor*. These results were independent of whether participants performed the task with their *dominant* hand or the *reverse* hand.

We sought to investigate whether or not the amount of Congruency effect observed on participants' RTs when they partnered with another *human* versus the *humanoid* robot were different from each other. Hence, we ran a dependent t-test on the differences between the participants' RTs on *congruent* and *incongruent* trials, with Co-agent (*human*, *humanoid*) as factor. No significant differences were found on the differences between the participants' RTs on *congruent* and *incongruent* trials according to the nature of the Co-agent ($t(27) = .09$, $p = .93$). Still, we found different effect sizes for the Congruency effect according to the Co-agent (Cohen's $d = .32$ when participants partnered with another *human* and Cohen's $d = .26$ when participants partnered with the *humanoid* robot). These additional results indicates that the amount of SSE did not differ regarding participants were cooperating with a human co-agent or a humanoid robot.

b) Engagement in the task

We sought to examine if the participants' engagement in the task varied according to the type of Co-agent (*human*, *humanoid*, *servomotor*). We computed linear models from each participant's mean RTs on congruent trials as a function of the type of Co-agent from totally social to non-social. Because the distribution of the slopes deviated from normality ($W = .90$, $p = .01$), we computed a one-sample Wilcoxon signed rank test. We found that the mean slope of the regressions were significantly greater than 0 ($W = 329$, $p = .002$, one-tailed). Hence, the more social their co-agent was, the more participants gradually get faster to perform the task.

c) Temporal interval estimation

We investigated the influence of the social context and the target congruency on the participant's perceived action-tone interval between an action (self- or other- generated action) and a subsequent auditory tone. We computed a within-subjects 3 × 2 x 2 ANOVA with the factors Co-agent (*human*, *humanoid*, *servomotor*), Congruency (*congruent, incongruent*) and Agent (*self, other*). We found a significant main effect of Co-agent ($F(2,54) = 5.36$, $p = .008$) and a significant Co-agent x Congruency x Agent interaction (*HFe* = .78, $p = .02$).

Post-hoc comparisons investigating the main effect of Co-agent showed that participants' temporal estimation were shorter when they partnered with another *human* compared to the *humanoid* robot (respectively 631.42 ms (SD = 184.37) and 676.92 ms (SD = 267.07); $p = .03$) and to the *servomotor* (758.89 ms (SD = 255.59); $p < .001$). In addition, participants' temporal estimations were shorter when they partnered with the *humanoid* robot compared to the *servomotor* ($p = .001$; see **Figure 19**). Hence, these results indicated that the humanness of the co-agent modulated participants' perceived durations of the action-tone intervals. Indeed, the more human-like their co-agent was, the more they tended to bind the action to the tone when they reported the intervals durations.

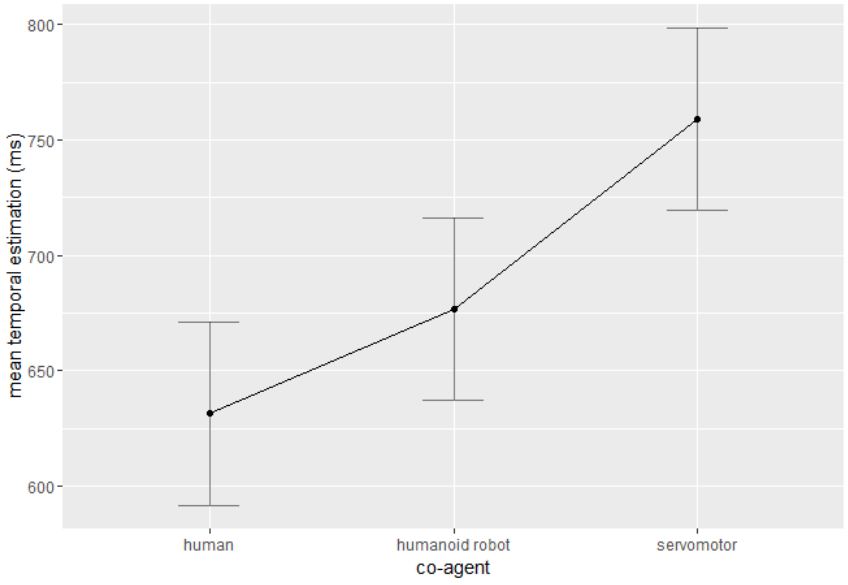


**Figure 19**. Main effect of Co-agent on participants' mean time interval estimations.

Post-hoc comparisons investigated the Co-agent x Congruency x Agent interaction showed that for *self*-generated actions, participants' estimations were shorter on *incongruent* trials when they partnered with another *human* compared to the *servomotor* ($p$ < 001). Similarly, participants' estimations were shorter for on *incongruent* trials when they partnered with the *humanoid* robot compared to the *servomotor* ($p$ < 001). However, there was no difference in participants' estimations on *incongruent* trials when they partnered with another *human* or with the *humanoid* robot ($p$ = .52). For other-generated actions, participants had shorter temporal estimations on *incongruent* trials when they partnered with another *human* compared to the *servomotor* ($p$ = .008). There were no differences on participants' temporal estimations on *incongruent t*rial when they partnered with another *human* compared to the *humanoid* robot ($p$ = .124), or when they partnered with the *humanoid* robot compared to the *servomotor* ($p$ = .362, see **Figure 20**). In sum, participants' perceived intervals duration were modulated only when the target appeared on the opposite side of their response key. Indeed, for *self*-generated actions on *incongruent* trials, participants reported shorter intervals either they performed the joint task with another *human* or with the *humanoid* robot compared to the *servomotor*. For *other*-generated actions on *incongruent* trials, the more human-like their co-agent was, the shorter their temporal estimations were. In particular, they reported shorter intervals durations when it was the *human* co-agent who performed the actions compared to the *servomotor*-generated actions. They made intermediary temporal estimations when the actions were generated by the *humanoid* robot.

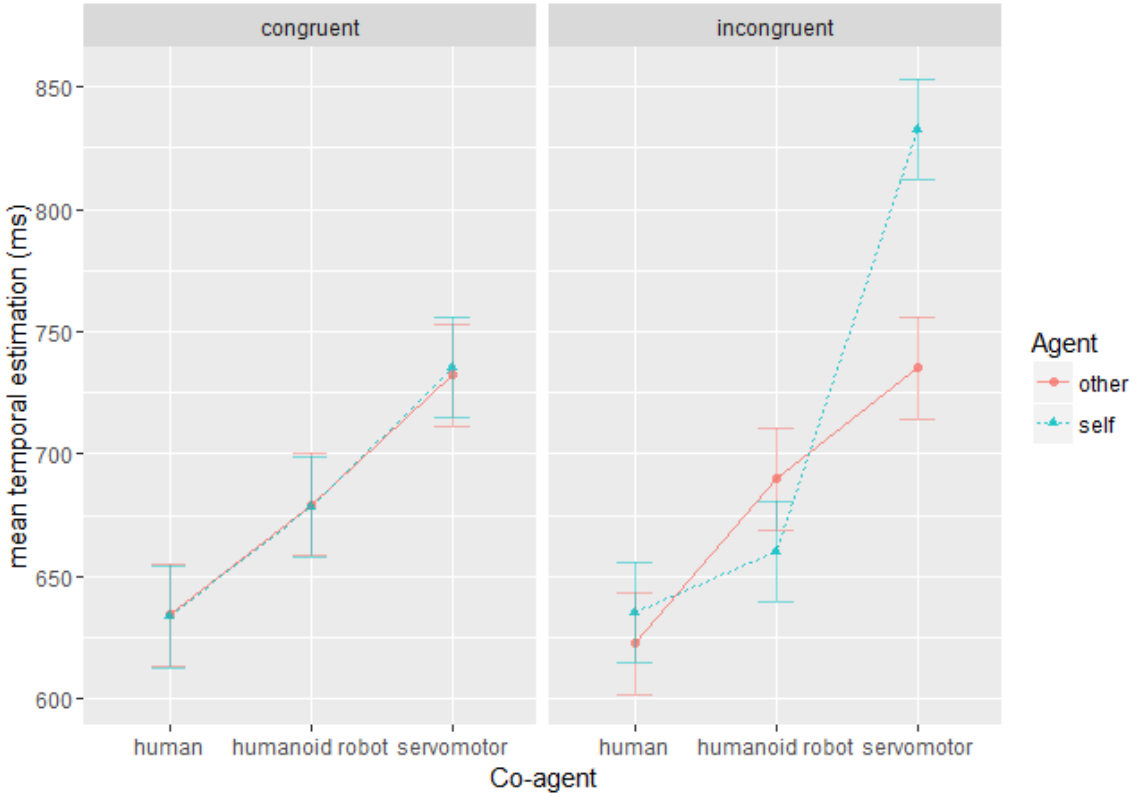


**Figure 20**. Interaction between Co-agent, Congruency and Agent on participants' means temporal interval estimations.

### 3.3.5. Discussion of the experiment

In this study, we aimed to investigate (1) the transformation induced by the development of a new "we" agentive identity regarding the experience of self-agency, (2) the effect of robot humanness on individuals' senses of self-agency and vicarious agency, and (3) the factors influencing the sense of agency during a joint task. We used a joint Simon task where participants had to detect a target that could appear in the same location as the response key or in an incongruent location (*Sebanz, Knoblich, & Prinz, 2003*). We manipulated the type of co-agent so that participants could perform the task jointly with another human, a humanoid robot, or a servomotor. Accurate target detection triggered an auditory tone after a randomized delay. In addition, participants had to estimate the temporal delay between the target detection (self- or other- generated) and the onset of the tone. This measure served as an implicit measure of participants' sense of agency (*Haggard, Clark, & Kalogeras, 2002*).

Regarding the transformation induced by the development of a new "we" agentive identity, we found no main difference in participants' temporal interval estimations between the *self* and *other* trials. These results suggested a shift from a sense of self-agency to a sense of "we-agency" when engaged in a joint task, regardless the nature of the co-agent. Such results corroborated Dewey, Pacherie & Knoblich's results (2004) which shown that experience of agency when engaged in joint action is no more related to egocentric sensory predictions but with the degree of control exhibited by the team as a whole (see also *Van der Wel, Sebanz, & Knoblich, 2015*). These data support the idea that during joint actions, a we-mode is running so that individual actions are turned into common actions (*Searle, 1983; Crivelli and Balconi, 2010*).

Concerning the effect of the robot physical humanness on the individuals' senses of self-agency and vicarious agency, we found that participants reported shorter action-tone 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*, both for their own actions and for their co-agent actions. Thus, our findings sustain that the development of the "we-agency" was modulated by the social feature of the co-agent. Indeed, when interacting with a *servomotor* compared to the other types of co-agents, participants' senses of self-agency and we-agency were at their lowest. It can be supposed that participants might be unable to build the "we-identity" because they shared no common features with the machine. Unlikely, the similarity with the *humanoid* robot could let participants to envisage the machine as a potential social pair, and thus allowing the construction of the "we-identity". Hence, this could explain why the participants' sense of we-agency was at its maximum during the human-human interactions and sharply declined during the human-servomotor interactions. This echoes to Searle (1983)'s talk when he said that recognize the other as similar as oneself and also as a potential agent is a prerequisite to engage oneself in a collaborative activity (*Searle, 1983*). Like so, when interacting in dyad, a we-agency is created but the features of the co-agent will orientate the agentive experience: when two human partners are involved in a task together, they feel a sense of agency for actions that have been triggered by the other as for their own. Furthermore, to be involved in a joint task with a machine that shared human features seems to detain the loss of agency observed during human-computer interactions (**Experiment 1**; *Obhi & Hall, 2011b*).

Interestingly, participants' RTs were also modulated by the physical humanness of the co-agent. Firstly, we found evidence that participants' engagement in the task was linked with the type of co-agent they partnered with. Indeed, the more social their co-agent was, the more participants gradually get faster to detect the congruent targets. This gives weigh to theories claiming that individuals' engagement in a task might rely on their sense of agency (*Caspar, Christensen, Cleeremans, & Haggard, 2016*). Secondly, we found a Congruency effect when participants had to perform the joint Simon task with another *human* and with the *humanoid* robot but not when participants performed the task with a non-human like machine such as the *servomotor*. These findings are consistent with previous results supporting that during human joint actions individuals are faster to detect targets when the target/response key mapping is congruent in comparison with an incongruent mapping *(Sebanz, Knoblich, & Prinz, 2003*). Regarding the joint task with the robots, we furnished additional evidence that manipulating the humanness of the machine could modulate the Congruency effect. Consistently, previous studies using a joint Simon paradigm showed no Congruency effect on participants' RTs when they partnered with non-biological agents such as a wooden hand (*Tsai & Brass, 2007*), a believed unseen desktop computer (both at the behavioral and the cerebral levels, *Tsai, Kuo, Hung & Tzeng, 2008*, but see *Wen & Hsieh, 2015*), or a visible desktop computer (**Experiment 1**). Yet, the congruency effect reappeared when an intentional stance was given to the machine, as suggested by previous work (*Stenzel, Chinellato,Tiradou Bou, del Pobil, Lappe, & Liepelt, 2012*). In the study of Stenzel and colleagues (2012), the participants were sitting next to a humanoid robot described either as an intelligent and active agent or a passive machine acting in a deterministic way. Interestingly, the authors found a congruency effect when the robot was introduced as a human-like intentional robot that can actively act but not when the robot was introduced as a deterministic machine. This finding emphases that the mere belief of intentionality is sufficient to modulate the congruency effect during a joint Simon task with a machine (*Stenzel, Chinellato,Tiradou Bou, del Pobil, Lappe, & Liepelt, 2012*). In our experiment, however, we were not able to separate the intentional stance belief of the *humanoid* robot from its physical humanness and their potential separated effects on participants' RTs. We acknowledge this constituted one major limitation of the study. Moreover, participants' RTs revealed that the Congruency effect did not differ in amplitude when participants performed the task with another *human* co-agent or with the *humanoid* robot suggesting that the biological characteristic of the co-agent *perse* was not responsible for the Congruency effect but rather the ability to envisage the co-agent as a social partner.

Hence, social robots can elicit similar cognitive processes that are involved during human-human interactions. However, note should be taken that it is unclear whether such a congruency effects could be attributed to the participants' ability to integrate the behavior of their co-agent into their own motor scheme (*Sebanz, Knoblich, & Prinz, 2003*) or rather to an attentional effect thereby providing a spatial referential for action coding (*Dolk, Hommel, Prinz, & Liepelt, 2013; Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz, & Liepelt, 2014*).

Finally, our findings also suggest differences in the processing of individuals' sense of self-agency and sense of we-agency. Indeed, the classical *doxa* is that individuals' sense of agency is informed by the dynamic integration of both internal motoric cues and external situational cues, with a higher motor cueing (*Moore & Fletcher, 2012*). In line with the involvement of motor cues in individuals' sense of agency construction, some authors have proposed that when action selection was easy (e.g. on congruent trials), participants' sense of self-agency was stronger compared to conflictual action selection (e.g. on incongruent trials) (*Chambon, Sidarus, & Haggard, 2014; Sidarus & Haggard, 2016*). However, in our study, no such involvement of motor cue integration was found on participants' sense of we-agency, suggesting that the fluency of action did not enter into consideration in the we-agency construction. Consequently, it can be said that the weight of the internal cues is weakened when individuals are not considered as single entities but as whole in a "we-identity".

To sum up, in **Experiment 3**, found a shift from a sense of self-agency to a sense of "we-agency" when engaged in a joint task. We showed that the physical appearance of an automated artificial system could modulate this sense of we-agency during a joint task. Indeed, participants exhibited a stronger sense of we-agency when they cooperated with another human compared to a (non-human-like) servomotor. Interestingly, an intermediate level of sense of we-agency was found during participants' interactions with a full humanoid robot. These results suggest that the sense of "we-agency" can be modulated through top-down processes. The next experiment consisted in investigating whether individuals' sense of we-agency could be modulated with a bottom-up approach. The aim of **Experiment 4** was to investigate whether individuals' sense of we-agency was sensitive to sensorimotor information such as haptic feedbacks about other-generated (human or machine) actions.

## 3.4. Experiment 4: modulating the sense of we-agency using low levels sensorimotor cues

### 3.4.1. Context of the experiment

The aim of this study was to investigate the effect of haptic feedbacks on individuals' sense of we-agency during human-human joint actions versus human-machine joint actions.

To our knowledge, the involvement of tactile reafferences in individuals' sense of we-agency during human-human joint task has been scarcely investigated (*Obhi & Hall, 2011b*). In Obhi and Hall (2011a, experiment 1 and experiment 2)' study, naive volunteers were paired and had to act jointly to press a spacebar of a keyboard at the time of their own choosing. In the first experiment, the participants were instructed that as soon as one of them had pressed the spacebar, the other had to join him or her by pressing the spacebar as quickly as possible. The first press of the spacebar triggered an auditory tone after 200 ms. In this experimental setup, both participants intended to trigger the sensory consequence. In the second experiment, one of the participants was instructed to press the space bar at the moment of his or her own choosing (the initiator) while the other had to join him or her by pressing the spacebar immediately after (the follower). In this context, the sensory consequence was always triggered by the personal intention of the initiator. In both experiments, participants had to report their perceived sense of causal responsibility using a scale (this served as an explicit measure of the participants' sense of we-agency). Participants also had to estimate the onset of the initiator's action and onset the auditory tone (this served as an implicit measure of the participants' sense of we-agency). In both experiments, both the initiator and the follower demonstrated intentional binding (*Obhi & Hall, 2011a*). This implicit sense of agency developed for the actions of another individual engaged in joint action has been explained through the concept of "we-identity" (*Balconi, 2010*). However, the authors' study (*Obhi & Hall, 2011a*) bore some methodological limitations. The paired participants were sitting side by side without any visual occlusion. Thus, it was not possible to dissociate the influence of the visual information exchanged between the partners from the influence of the kinesthetic information generated by the spacebar that the participants shared on the sense of we-agency. Therefore, in the current experiment, we wanted to create a protocol wherein the visual information regarding the

performance of the partner would be absent but the kinesthetic sensations related to the other's action would be preserved.

Furthermore, with respect to human-machine interactions, we sought to investigate the effects of artificial-generated tactile reafferences on individuals' sense of agency during joint actions with artificial automated systems. Indeed, it has been shown that the human action/observation matching system could be sensitive to artificial motions (*Craighero, Jacono, & Mele, 2016*; *Gazzola, Rizzolatti, Wicker & Keyser, 2007*). Consequently, it could be think that kinesthetic feedbacks about machine-generated actions could provide a support for action and intention understanding, hence allowing a sense of we-agency.

In this context, we ran a behavioral study wherein participants had to manipulate a pair of haptic interfaces jointly with an unseen human co-agent. The display allowed participants to receive sensorimotor information about their co-agent. Critically, during half of the experiment, unknown to participants, a robotic system replaced the human co-agent. This artificial system generated human-like motions. Hence, the type of motor information participants had as a feedback from their co-agent was manipulated {human, artificial human-like}. An auditory tone was presented after a certain delay following the end of the joint action. Participants were asked to report estimate of the time interval between the end of the joint action and the onset of the tone. This served as an implicit measure of participants' sense of agency (*Haggard, Clark, & Kalogeras, 2002*). Participants were also asked to report their perceived contribution on the action outcome. This served as an explicit measure of participants' judgment of agency (JoA) (*van der Wel, Sebanz, & Knoblich, 2012*). We predicted that human-generated reafferences would enhance participants' sense of we-agency during the human joint actions. During the joint task with the robotic system, we predicted that artificial reafferences would not be totally detrimental to individuals' sense of we-agency as it would allow machine's intentions understanding and as participants were induced the thought to interact with a human peer. Yet, we believed the sense of we-agency would be weaker with the artificial reafferences compared to human-generated reafferences.

### 3.4.2. Method

a) Ethic statement

This study was approved by the institutional ethical research committee of the Paris Descartes University (France). The investigation was carried out in accordance with the Declaration of Helsinki and 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. Participants were not paid for their participation in the experiment.

b) Participants

Twenty-six healthy adults volunteered to take part in the experiment (17 women, mean age 21.73 years, SD of age 3.80 years). One participant was excluded from the sample because the debriefing interview suggested that he had some doubts about his co-agent being a robot and not a human. All participants were right-handed and had normal or corrected-to-normal vision. None of them had prior knowledge about the purpose of the experiment. Participants were not paid for their participation in the experiment.

c) Material and stimuli

Participants were sitting 46 cm from a computer screen with a refresh rate of 120 Hz.

Stimulus presentation was controlled using a homemade software and hardware from the ISIR laboratory. Two auditory tones (1000 Hz and 2000 Hz, each 120 ms duration), presented via headphones, were used during the experiment as the effects of the participant/co-agent's actions (e.g., handle pushes bumping stop positions) for measuring IB phenomenon.

A pair of two haptic interfaces developed at the ISIR laboratory was used (*Roche & Saint-Bauzel, 2016*). This consisted in interconnected handles (see **Figure 21**) that could be moved rightward or leftward by the participants' index finger or by a robotic program. The robotic program was implemented to randomly turn the handle right or left after a random delay between 0 and 3 seconds. If participants started moving their handle before the machine co-agent, then the robotic program followed the participants' lead. If the machine co-agent was first to move, but participants exerted forces on their handle in the opposite direction above a 2 Newton threshold, the robotic program changed its direction to

follow the participant for safety reasons. The machine co-agent's handle simulated human motion using minimum jerk optimization. Tactile receptors located on the handles served to measure exerted forces.

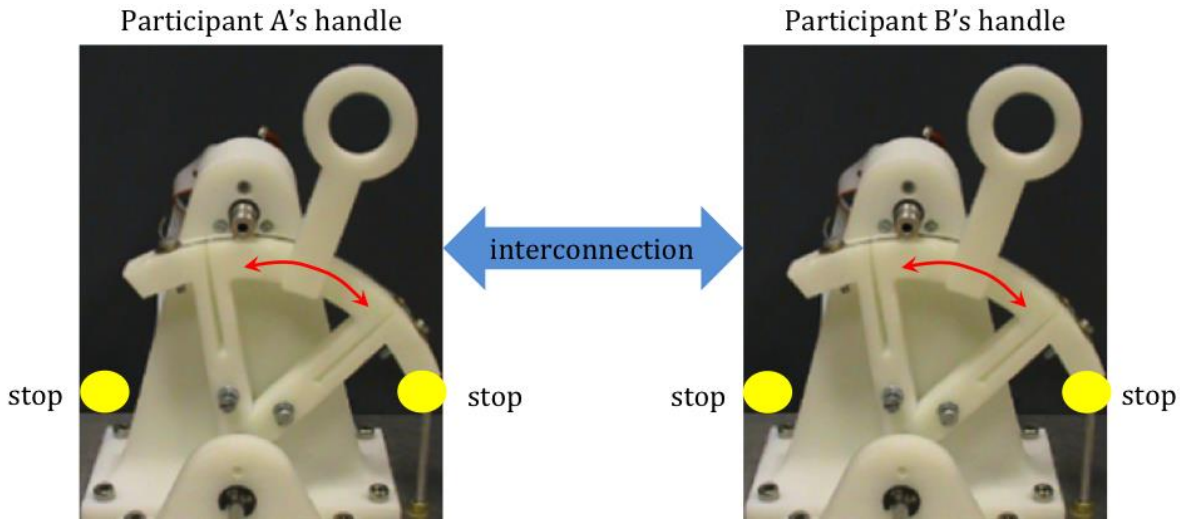


**Figure 21**. Illustration of the haptic devices used in the experiment designed at the ISIR lab by Roche and Saint-Bauzel (2016).

d) Procedure

Participants were paired and placed side by side, separated by a curtain, each sitting in front of a screen. A pink noise was played in participants' headphones to prevent any sound cue from each other. In order to make participants believe that they would always be interacting with a human co-agent, they were explained that during the whole experiment, they would have to jointly manipulate a pair of connected handles that allow sensorimotor information about the other's performance. In fact, in the *joint HM* condition, their co-agent was replaced by a robotic program but participants were not informed about this manipulation.

In the *baseline* condition, participants were presented two different auditory tones separated by a delay of 700 ms or 1300 ms. Participants were told that this delay could range between 0 to 2000 ms, and had 5 seconds to provide their estimation by moving a horizontal slider that extended from 0 to 2000 ms.

In the *joint HH* condition, the paired participants were asked to turn their handle rightward or leftward, at the time of their own choosing but within a 3 seconds time window

after the onset of a fixation cross. Each handle reproduced the forces applied to the other handle. Participants were told that if their co-agent initiated the action first, they would have to follow her/his lead. The two participants were asked to cooperate together and equate the number of times they each initiated the action. When the two connected handles reached stop positions, an auditory tone was presented after a SOA of 700 ms or 1300 ms. Participants were told that the delay could range between 0 to 2000 ms. Rightward handle pushes were always associated with a specific type of auditory tone (e.g., 1000 Hz), whereas leftward handle pushes were always associated with another type of auditory tone (e.g., 2000 Hz). The mapping action/tone has been counterbalanced across participants.

In the *joint HM* condition, the paired participants were told the same instructions as in the *joint HH* condition. Hence, they believed that they could receive sensorimotor information about the other human agent's performance but in reality, they were cooperating with a robotic program that simulated human motion.

In both operant conditions (i.e., the *joint HH* condition and the *joint HM* condition), participants had to estimate de delay between the onset of the moment their handle reached the stop position and the onset of the auditory tone, by moving a horizontal slider that extended from 0 to 2000 ms. This time interval estimates served as an implicit measure of participants' sense of agency (*Haggard, Clark, & Kalogeras, 2002*). Participants were also asked to rate from 0 ("absolutely not") to 100 ("absolutely") how much they thought they had contributed to causing the auditory tone by moving a slider on a ruler representing the percentage of their contribution. This served as an explicit measure of participants' judgment of agency (JoA) (*van der Wel, Sebanz, & Knoblich, 2012*, see **Figure 22** for a summary).

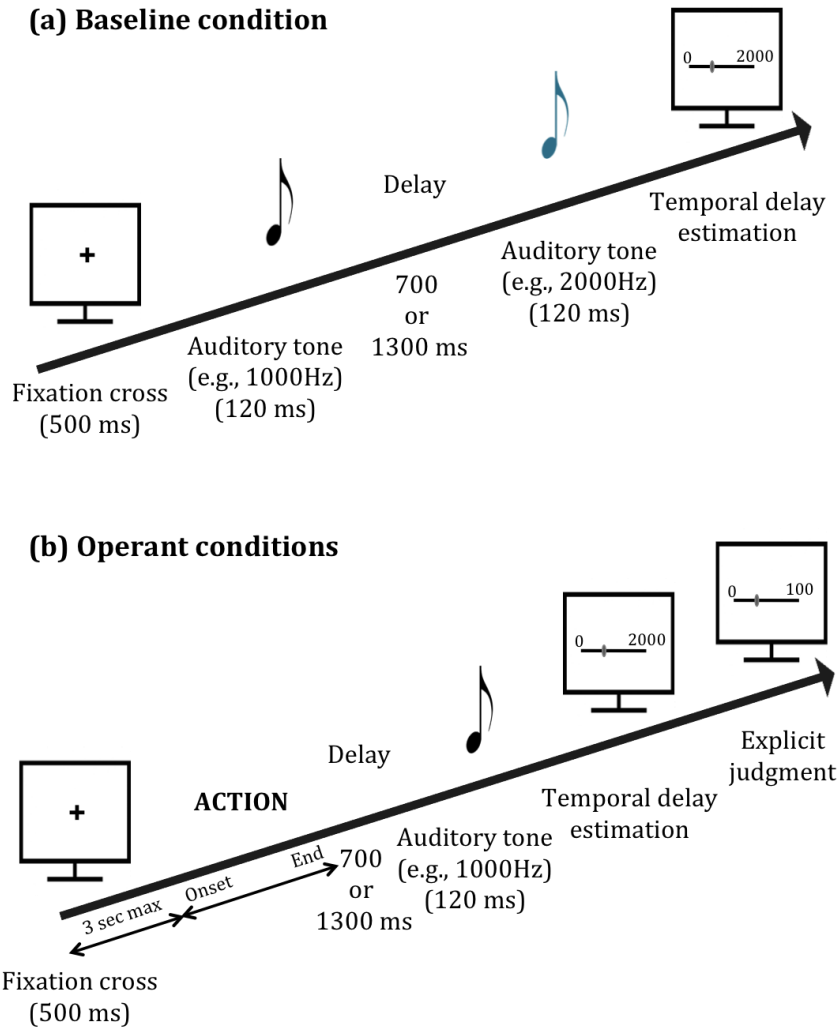


**Figure 22**. Trial timelines in the (a) baseline condition and (b) operant conditions.

At the beginning of the experiment, participants were trained to estimate temporal intervals in milliseconds. During this training, they were presented two different auditory tones separated with a random delay between 500 ms and 1500 ms. Participants were asked to report the perceived duration of random delay by moving a horizontal slider that extended from 0 to 2000 ms. Then, a visual feedback about their estimation error was given in order to accurately recalibrate their internal clock. This training session consisted in 30 trials. In addition, participants had an association learning session which consisted in performing rightward handle pushes that were followed by a specific auditory tone (e.g., 1000 Hz) and leftward handle pushes that were followed by a different auditory tone (e.g., 2000 Hz). The aim of the association learning session was to make participants associate their actions with the onset of the subsequent tone. The association learning session consisted in 20 trials.

At the end of the experiment, all participants were individually interviewed to verify they had been believed interacting with another human during the entire experiment. They were asked the three following questions: (a) “Do you have any comment regarding the experiment?”; (b) “Did you notice a difference between the two sessions where you interacted with your co-agent?”; and (c) “In fact, you were interacting with a human partner in one session and with a robot in the other. Did you suspect that?”.

The experimental conditions were tested using a within-subject design. Participants always started with the *baseline* condition and the order of the two operant conditions was counterbalanced between participants. The baseline condition contained 40 trials and each operant condition contained 120 trials. For every individual trial of the operant conditions, participants were attributed a posteriori an “initiator” or a “follower” role based on the forces applied on the handles. More precisely, the initiator role was assigned to the first agent who applied a 0.2 N force on her/his/its handle (i.e., the threshold discriminating involuntary micro-movements, *Roche & Saint-Bauzel, 2016*)**.** Our factors were 3 Tasks (*baseline*, *joint HH*, *joint HM*) x 2 Delays (*700*, *1300*) x 2 Roles (*initiator*, *follower*).

### 3.4.3. Data analyses

Our dependent measures were the participants’ mean temporal interval estimations, the participants’ mean rating of their contribution to the action outcome, the number of times the movement of the participants’ handle changed direction before reaching the stop position, and the sum of forces applied by the participants to their handle. Statistical analyses were performed with R software (3.3.1 version). The significance level was set at $\alpha = .05$. Post-hoc comparisons were made using the false discovery rate correction (*Benjamini & Hochberg, 1995*).

a) Manipulation of belief

This analysis aimed at verifying that the belief induction (i.e., that paired participants believed that they had been interacting with one another during the entire experiment) was a success. The answers of the participants were recorded and analyzed by two independent raters. Those raters had to judge whether participants suspected that they had been interacting with a desktop computer. If the two raters disagreed, the judgment of a third rater was requested. Those raters had to judge whether participants suspected that

they had been interacting with a robot. If the two raters disagreed, the judgment of a third rater was requested. The answers participants gave to the first question (a) are listed in **Appendix 2**.

b) Temporal interval estimation accuracy checking

To verify whether participants correctly perceived the different time interval between the two events they had to judge, we conducted a prior within-subjects one-way ANOVA on the raw temporal estimations with Delay (*700*, *1300*) as a factor. This control analysis allowed us to ensure that participants paid attention to the time intervals. The Delay was not included in further analyses as a factor.

c) Intentional Binding (IB)

Firstly, we sought to investigate participants' implicit sense of agency. Specifically, we aimed at verifying whether participants' mean temporal interval estimations differed when the auditory tones were generated by voluntary joint actions (i.e., in the operant conditions) compared to when no actions were involved (i.e., in the *baseline* condition). One outlier was removed from the dataset of the analysis because her corresponding temporal estimations were always beyond two standard deviations from the mean of the group. The Delay (*700*, *1300*) factor was excluded from the analysis. Indeed, in the present experiment we were interested in the way sensorimotor cues influence IB in general, rather than its influence on IB for different action-tone interval delays. Because the occurrence of the auditory tones did not involve any participants' actions in the *baseline* condition, the Role (*initiator*, *follower*) factor was excluded from this analysis. Hence, we conducted a within-subjects one-way ANOVA on participants' mean temporal interval estimations with Task (*baseline*, *joint HH*, *joint HM*) as a factor.

Secondly, we sought to investigate whether participants' mean temporal interval estimations were modulated by the asymmetry of implicit social role establishment (i.e., being the initiator or the follower of an action) during haptic joint actions, in regard with the nature of the co-agent participants partnered with (i.e., human or machine). Therefore, we conducted a within-subjects 2 x 2 ANOVA on participants' mean temporal interval estimations with Co-agent (*human*, *robot*) and Role (*initiator*, *follower*) as factors.

d) Explicit Judgment of Agency (JoA)

Consistently with IB analyses, we sought to investigate whether participants' explicit judgments of agency were modulated by the asymmetry of implicit social role establishment (i.e., being the initiator or the follower of an action) during haptic joint actions, in regard with the nature of the co-agent participants partnered with (i.e., human or machine). We computed a within-subjects 2 x 2 ANOVA on participants' mean rating of their contribution to the action outcome with Co-agent (*human*, *robot*) and Role (*initiator*, f*ollower*) as factors.

e) Movement parameters

The first movement parameters analysis concerned the number of times the movement of participants' handle changed direction before reaching the stop position. This analysis aimed at quantifying how sensorimotor communication was achieved in a joint task with another human versus a robot as a Co-agent. We assessed the normality of the distribution (i.e., the difference between the *joint HH* and the *joint HM* task) of participants' number of times they changed direction, which showed a significant deviation from normality ($W$ = .73; $p$ <.001). Hence, we computed a non-parametric Wilcoxon signed rank test on the average number of times the movement of participants' handle changed direction with Co-agent as a factor (*human*, *robot*).

The second movement parameter analysis concerned the sum of forces applied by the participants to their handle. We aimed at evaluating whether participants exerted a different amount of force regarding the type of Co-agent they partnered with. We assessed the normality of the distribution (i.e., the difference between the *joint HH* and the *joint HM* task) of participants' mean sum of forces applied on the handle, which showed a significant deviation from normality ($W$ = .63; $p$ < .001). Hence, we computed a non-parametric Wilcoxon signed rank test on the mean sum of forces applied by participants to their handle with Co-agent as a factor (*human*, *robot*).

### 3.4.4. Results

a) Manipulation of belief

The answers of the participants during the debriefing were analyzed by two independent raters. When the two raters disagreed, the judgment of a third rater was requested. There was perfect agreement between the two initial raters for 24 participants and, of the two remaining participants, one was excluded. The answers that included participants gave to the first question are listed in **Appendix 2**.

b) Temporal interval estimation checking

We verified whether participants correctly perceived the different temporal intervals that were used in the task by assessing whether they adjusted their temporal estimations accordingly. We investigated the effect of Delay on participants' mean temporal estimations. We computed a within-subjects one-way ANOVA with the factor Delay (*700*, *1300*). We found a significant main effect of Delay on mean temporal interval estimations ($F(1,24) = 192.75$; $p < .001$) indicating that participants reported greater intervals when the delay between the two events to judge was 1300 ms compared to when it was equal to 700 ms (respectively 1079.50 ms (SD = 184.28) and 762.75 ms (SD = 129.78)). Even though this was not a variable of interest (indeed, we only introduced various delays in order to avoid the predictability bias), this indicates that participants were paying attention to the delays and correctly discriminated the different temporal intervals that were presented.

c) Intentional Binding (IB)

Firstly, we examined whether participants' mean temporal interval estimations differed in the *joint HH* condition and *joint HM* condition compared to the *baseline* condition. We assessed the normality of the distributions of participants' mean interval estimations. The analyses showed that none of the distributions deviated from normality (all $W > 0.90$ and all $p > .10$). We then computed a within-subjects one-way ANOVA with the factor Task (*baseline*, *joint HH*, *joint HM*). We found a significant main effect of Task ($F(2,48) = 5.64$; $p = .01$) revealing that participants reported greater interval estimations in the *joint HM* condition compared to the *joint HH* condition (respectively 940.38 ms (SD = 148.95) and 875.45 ms (SD = 176.62); $p = .002$). There was a trend according to which participants

reported greater interval estimations in the *baseline* condition (947.54 ms (SD = 311.30)) compared to the *joint HH* condition ($p$ = .08). However, there was no significant difference on participants' interval estimations between the *joint HM* condition and the *baseline* condition ($p$ = .83, *ns*, see **Figure 23**). These results indicated that participants did not exhibit IB when the auditory outcome was generated by a human-robot joint action with respect to auditory outcomes that were externally presented. In contrast, they reported IB when the auditory outcome was generated by a joint human-human action.

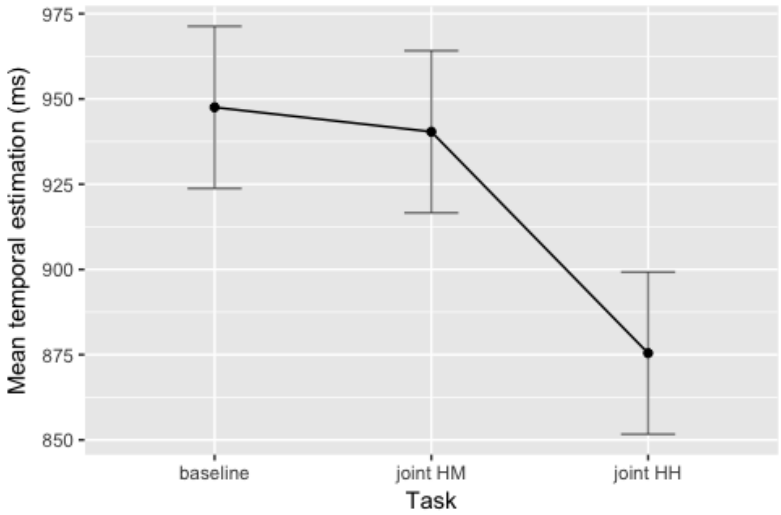


**Figure 23**. Main effect of Task on participants' mean temporal interval estimations.

Secondly, we investigated the impact of implicit social role such as being the initiator or the follower of a joint action in regard with the nature of the co-agent participants partnered with (i.e., human or robot). We conducted a within-subjects 2 x 2 ANOVA on participants' mean temporal interval estimations with the factors Co-agent (*human*, *robot*) and Role (*initiator*, *follower*). We found a significant main effect of Co-agent ($F$(1,24) = 9.33; $p$ = .01) indicating that participants reported longer temporal intervals when their Co-agent was a *robot* compared to when their Co-agent was another *human*. No significant main effect of Role ($F$(1,24) = .51; $p$ = .48, *ns*) or significant Task x Role interaction ($F$(1,24) = .01; $p$ = .92, *ns*) were found. Hence, being the i*nitiator* or the *follower* of a joint action did not modulate participants' temporal estimations.

d) Explicit Judgment of Agency (JoA)

We investigated the impact of implicit social role such as being the initiator or the follower of a joint action in regard with the nature of the co-agent participants partnered with (i.e., human or robot) on participants' explicit sense of agency. We computed a within-subjects 2 x 2 ANOVA on participants' mean rating of their contribution to the action outcome with the factors co-agent (*human*, *robot*) and Role (*initiator*, f*ollower*). We found a significant main effect of Co-agent ($F(1,25) = 11.75$; $p = .002$) and a significant main effect of Role ($F(1,25) = 90.38$; $p < .001$) on the participants' mean ratings. No significant Co-agent x Role interaction was found (($F(1,25) = .02$; $p = .89$).

We investigated the impact of implicit social role such as being the initiator or the follower of a joint action in regard with the nature of the co-agent participants partnered with (i.e., human or robot) on participants' explicit sense of agency. We computed a within-subjects 2 x 2 ANOVA on participants' mean rating of their contribution to the action outcome with the factors co-agent (*human*, *robot*) and Role (*initiator*, f*ollower*). We found a significant main effect of Co-agent ($F(1,25) = 11.75$; $p = .002$) and a significant main effect of Role ($F(1,25) = 90.38$; $p < .001$) on the participants' mean ratings. No significant Co-agent x Role interaction was found (($F(1,25) = .02$; $p = .89$).

Post-hoc comparisons investigating the main effect of Co-agent revealed that participants' ratings were higher when their Co-agent was a *robot* compared to when their Co-agent was another *human* (respectively 60.81% (SD = 15.63) and 50.79 % (SD = 18.16); see **Figure 24**). These results indicate that participants demonstrated a stronger JoA when they were in a joint task with a *robot* compared to when they performed the same task with another *human*.

Post-hoc comparisons investigating the main effect of Role revealed that the participants' ratings were higher when they had the Role of an *initiator* compared to when they had the Role of a *follower* (respectively 65.10% (SD = 14.90) and 46.50% (SD = 15.10); see **Figure 25**). These results indicated that participants demonstrated a stronger explicit judgment of agency when they initiated the joint actions compared to when they followed the lead of an action initiated by a co-agent.

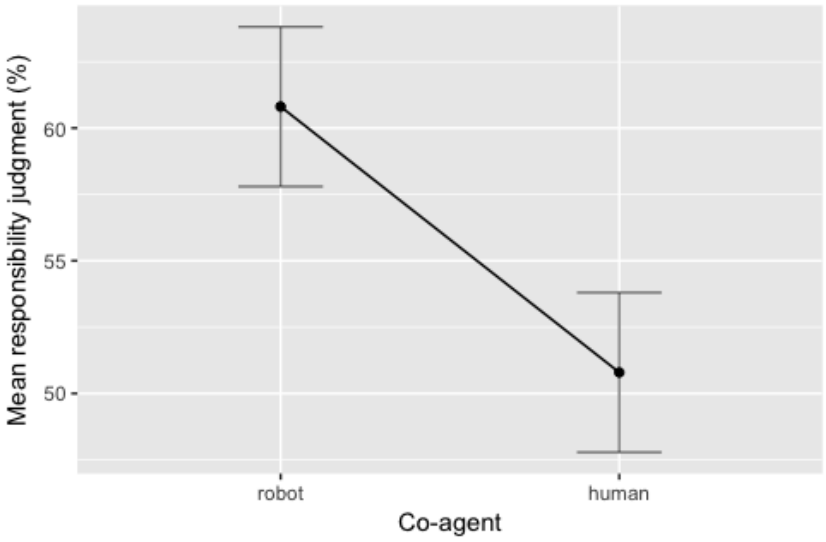


**Figure 24**. Main effect of Co-agent on participants' mean explicit judgment of responsibility for the action outcome.

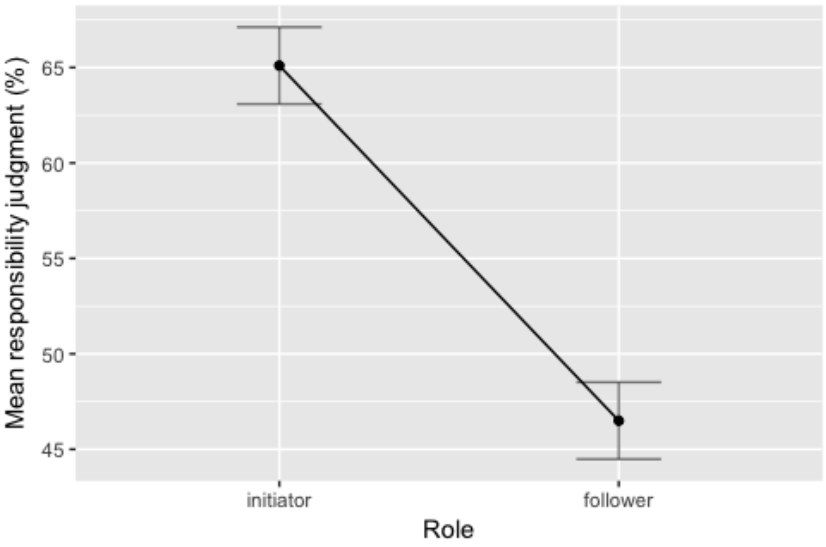


**Figure 25**. Main effect of Role on participants' mean explicit judgment of responsibility for the action outcome.

e) Movement parameters

We investigated how sensorimotor communication was achieved in a joint task with another human or with a robot as a co-agent. We computed a Wilcoxon signed rank test on the average number of times the movement of participants' handle changed direction

before reaching the stop position with Co-agent as factor (*human*, *robot*). We found a significant main effect of Co-agent on the average number of times the movement changed direction ($W = 321$; $p < .001$) indicating a larger average number when the Co-agent was a *human* compared to when it was a *robot* (respectively mean = 1.49 and mean = 0.16). These results indicated that participants had more sensorimotor exchanges with the human co-agent in comparison with the robot co-agent.

We also investigated whether participants exerted a different amount of force depending on the type of Co-agent they partnered with. We computed a non-parametric Wilcoxon signed rank test on the mean sum of forces applied by participants to their handle with Co-agent as factor (*human*, *robot*). We found no significant main effect of Co-agent but a near trend according to with participants (with the human co-agent mean = .15 N and with the robot co-agent mean = .05 N; $W = 103$; $p = .11$, *ns*).

### 3.4.5. Discussion of the experiment

In this study, we sought to investigate how sensorimotor communication was achieved in a joint task with another human or with a robot as a co-agent. Participants were asked to manipulate a pair of haptic interfaces jointly with another human and could receive haptic feedbacks from each other (*joint HH* condition). During half of the experiment, unknown to participants, the sensorimotor feedback they received was generated by a robotic system (*joint HM* condition). A sensory consequence was presented after a certain delay following the end of the joint action. Participants were asked to report their estimate of the time interval between the end of the joint action and the onset of the consequence. This served as an implicit measure of participants' sense of agency (*Haggard, Clark, & Kalogeras, 2002*). Participants were also asked to report their perceived contribution on the action outcome. This served as an explicit measure of participants' judgment of agency (JoA) (*van der Wel, Sebanz, & Knoblich, 2012*).

We found that participants' demonstrated an implicit sense of we-agency as indicated by IB phenomenon when the auditory outcome was generated by a joint human-human action. On the contrary, they did not manifest such an implicit sense of we-agency indicated by no IB at all when the auditory outcome was generated by a human-robot joint action. In addition, we found that being the i*nitiator* or the *follower* of a joint action did not modulate IB, suggesting a lack of agency when participants interacted with the artificial

system even when they *intentionally* attempted to generate a sensory consequence in the environment (i.e. when they were the *initiator* of the joint action). Given that the tactile reafferences could potentially allow machine's intentions understanding and that participants were induced the thought to interact with another naïve participant, it could be expected that participants' sense of we-agency would not totally be declined in the *joint HM* condition. However, the changes in kinesthetic/tactile reafferences yielded by the artificial system sufficed to modulate participants' sense of agency. This gives supplementary weight to the computational account of the sense of agency that proposed a higher weighting for internal sensorimotor cues in comparison with contextual cues in one's sense of agency construction (*Moore & Fletcher, 2012*). Indeed, during the human joint actions, the haptic feedback informed participants that they were cooperating with another human, which explained why participants' sense of we-agency was maximal. However, during the joint task with the robotic system, even if participants though they were cooperating with another human, the haptic feedback implicitly informed participants that they were not in a joint task with an intentional agent, which in turn decreased their sense of agency. Hence, participants' belief had little weight in the construction of their sense of agency given that motor cues were available.

In addition, our results showed a clear dissociation between the explicit and implicit measures of participants' sense of agency. First, at the explicit level, we found that participants demonstrated a stronger explicit JoA when they were in a joint task with a *robot* compared to when they performed the same task with another *human,* whereas the reverse pattern was found when investigating IB. Second, we found that participants reported a stronger explicit judgment of agency when they were the *initiator* of the joint actions compared to when they were the *follower* of an action initiated by their co-agent whereas these social roles did not impact IB. This corroborates the meta-analysis by Dewey and Knoblich (2014) that stated that implicit and explicit measures were not correlated and might refer to distinct underlying neurocognitive mechanisms (*Dewey & Knoblich, 2014*).

Interestingly, the analyses of participants' movements during the joint task revealed that participants had more sensorimotor exchanges with the human co-agent in comparison with the robotic system. Exacerbations in kinematics could be interpreted as an implicit form of communication between the two agents involved in a cooperative task (*Vesper & Richardson, 2014*). Therefore, it could be said that participants were more able to extract the intention of their co-agent during the human-human joint task compared the human-

machine joint task. Indeed, movement kinematics might be a rich source of relevant cues that could be used to make predictions about a co-agent's action. Accordingly, several studies pointed out that intentions were embedded in agents' kinematics and that individuals were able to read informative intentions from actions with the help of their action/observation matching system (*Iacoboni et al., 2005*; *Manera, Becchio, Cavallo, Sartori, & Castiello, 2011*; *Sebanz & Shiffrar, 2009*). For example, Sebanz and Shiffrar (experiment 2, 2009) asked participants to watch videos of point light displays representing a basketball player making a pass or a deception. The authors asked participants – expert and novice basketball players - to report whether the agent on the video intended to fake or make a pass. The results showed that both expert and novice were able to discriminate from the kinematics the accurate intention of the observed agent, with a better performance in the expert group (*Sebanz & Shiffrar, 2009*). These results emphasize the involvement of the action/observation matching system in intention reading. Our results suggest that the participants' action/observation matching system was not sensitive to artificial-generated reafferences. This could explain why participants in our experiment were more sensitive to the human reafferences, enhancing their sense of we-agency.

In sum, we found that the implementation of kinesthetic machine-generated feedbacks in a haptic device did not suffice to make participant exhibit a sense of we-agency during their interaction with a robotic system with respect to human-human joint actions. Importantly, participants' prior belief did not modulate heir sense of agency as they had a direct access to sensorimotor information about their co-agent's intentions, which is in accordance with the weighing given to each cue in the computational model of agency by Moore and Fletcher (*Moore & Fletcher, 2012*).

# Chapter 4: General discussion of the thesis

## 4.1. A better understanding of the loss of agency

During the first stage of the thesis (**Experiment 1** and **Experiment 2**), we proposed to characterize the cognitive processes underlying the individuals' decreased sense of agency during joint tasks with automated artificial systems that is classically reported (*Berberian, Sarrazin, Le Blaye, & Haggard, 2012; Coyle, Moore, Kristensson, Fletcher, & Blackwell, 2012*; *Obhi & Hall, 2011b*). Specifically, we raised the question of the relationship between individuals' ability to co-represent their co-agent's actions into their own cognitive system and the development of the sense of we-agency. In both experiments, participants were asked to perform a joint Simon task with a co-agent that could be either another human or an algorithm implemented in a desktop computer. An action outcome (e.g., an auditory tone) was presented after a certain delay following the participants' or their co-agent' actions.

**Experiment 1** revealed at the behavioral level the convergence of action co-representation, measured with the SSE (*Sebanz, Knoblich, & Prinz, 2003*), and the vicarious sense of agency, measured with IB phenomenon (*Haggard, Clark, & Kalogeras, 2002*). Indeed, participants were able to co-represent human co-agent's actions and to develop a sense of agency both for the self- and other-generated action outcomes. Such abilities were impaired with the computer-generated actions. The ability to co-represent other-generated actions into one's motor system has already been proposed to function only for biological motions (*Mann, Williams, Ward, & Janelle,* 2007; *Tai, Scherfler, Brooks, Sawamoto, & Castiello, 2004*). One explanation could be that because the action schemes of the computer were not part of participants' motor repertoire (indeed, the computer made abstract actions generated by an implemented algorithm), they were not able to use their action/observation matching system in order to simulate the machine-generated actions. This echoes the issue of artificial system opacity that individuals face during their interactions with machines (*Taleb, 2012*). The inability to co-represent machine-generated actions – that are generated in an inappropriate form – might prevent individuals from understand the machine's intentions, creating a loss of agency.

**Experiment 2** aimed at confirming these assumptions at the cerebral level. The protocol of **Experiment 1** was replicated with slight adaptations (see the Method part of Experiment 2, section III.2.3, for more details) to adjust to EEG recording. We did not succeed in showing modulations in participants' sense of agency, measured by the N1 component amplitude (*Poonian, McFadyen, Ogden, & Cunnington, 2015*), according to the

nature of their co-agent (i.e., human versus computer). Because the nature of the participants' co-agent induced modulations on IB in **Experiment 1**, our findings comforts the claim according to which implicit measures of individuals' sense of agency such as IB and sensory attenuation phenomena might tap into different processes (*Dewey & Knoblich, 2014*). Nevertheless, we found a reliable modulation of action co-representation, measured by the P3 component amplitude (*Sebanz, Knoblich, Prinz, & Wascher, 2006*), demonstrating that computer-generated actions were not represented into the participants' brain during the joint task, whereas the actions generated by the human co-agent were. Taken together, our findings suggested that during human joint tasks, the ability to simulate a co-agent's action into one own motor system could provide a better understanding of the co-agent's intention based on one's own previous experiences. This kind of motor simulation might sustain the ability to build a sense of vicarious agency (*Obhi & Hall, 2011b*; *Sahaï, Pacherie, Grynszpan, & Berberian, 2017*). During human-machine joint tasks, as individuals did not consider the artificial system as intentional, they did not use their sensorimotor system to simulate and try to understand the machine-generated action and consequently failed to develop a sense of vicarious agency. In other words, predictive mechanisms classically involved during action simulation seemed to operate distinctively according to the type of action (human-generated vs. computer-generated), hence modulating individuals' sense of vicarious agency. This led to the question of how artificial automated systems should be designed to promote action understanding and thus guarantee meaningful human-machine interactions. An increasing number of studies has provided evidence that humanized automated artificial systems could enhance individuals' sense of agency during their interactions with machines (*Caspar, Desantis, Dienes, Cleeremans, & Haggard, 2016*; *Wohlschläger, Haggard, Gesierich, & Prinz, 2003*).

## 4.2. Reallocation of agency during a joint task with a full humanoid

To address this issue, the second stage of the thesis (**Experiment 3** and **Experiment 4**) consisted in studying the influence of the humanness of the co-agent on individuals' sense of agency. In **Experiment 3,** using a top-down approach, we investigated the influence of high-level properties such as the physical appearance of the machines on participants' sense of agency during a joint Simon task. Hence, participant could interact with another human, a full humanoid robot or a servomotor as a co-agent. Participants' sense of agency was assessed using IB phenomenon (*Haggard, Clark, & Kalogeras, 2002*).

We found that participants co-represented the actions generated by the human co-agent and by the humanoid robot but not the action generated by the servomotor. Hence, the results of **Experiment 3** corroborate the social component of action co-representation observed in **Experiment 1**. Indeed, in **Experiment 1** we revealed an absence of action co-representation effect when participant's performed the Simon task jointly with a non-human like machine such as a desktop computer. Nevertheless, one limitation of **Experiment 1** consisted in the fact that during the human-human joint task, there was the physical presence of another agent sitting next to the participants. In contrast, in the human-computer joint task, an algorithm implemented in a desktop computer triggered the other-generated target detections and there was no physical presence next to the participants. Hence, one could argue that the absence of Congruency effect observed in the machine condition of **Experiment 1** could be due to the mere absence of a physical presence next to the participants during the task, leading to an absence of spatial referential for action coding (referential coding hypothesis, see *Dolk, Hommel, Prinz, & Liepelt, 2013; Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz, & Liepelt, 2014*). However, in **Experiment 3**, the non-human-like machine (i.e., the servomotor) was physically present next to the participants during the task and participant still exhibited no Congruency effect. This support that the mere presence of a co-agent is not sufficient to generated a Congruency effect. Rather, the way the action outcome has been generated – by a co-agent that *could* produce similar or dissimilar actions as those hold in one's motor repertoire - might play a crucial role. It is to be noted that the humanoid robot actions did not follow biological motion laws but were rough punches on the response key. Hence, this experiment strictly manipulated prior thoughts such as the identity of the co-agent (human, humanoid or non-human like). Another aspect to be considered to explain our findings deals with empathy. Because human-like robots elicit empathic behaviors in humans as opposed with non-human-like robots (*Riek, Rabinowitch, Chakrabarti, & Robinson, 2009*), it is conceivable that individuals were more likely to create similarity boundaries and affiliation with humanoid robots (*De Vignemont & Singer, 2006*). Indeed, a linear relation has been shown to exist between the degree of anthropomorphism of robots and the activation of brain areas involved in the processing of other minds (*Krach, Hegel, Wrede, Sagerer, Binkofski, & Kircher, 2008).* Taken together, our findings (**Experiment 1**, **Experiment 2** and **Experiment 3**) suggest a clear self-integration of stimuli that refers to a social (human or humanoid) co-agent's action as proposed by the SSE framework (*Sebanz, Knoblich, & Prinz, 2003*).

Interestingly, in **Experiment 3**, we highlighted a shift in participants' agentive experience, from a sense of self-agency to a sense of "we-agency". This sense of we-agency was modulated by the automated artificial system' physical appearance during the joint task Indeed, participants' sense of agency was stronger for self- and other- generated actions when they cooperated with another human compared to a servomotor as a co-agent. Critically, this loss of agency observed during the joint task with the non-human-like machine was attenuated both for self- and other-generated actions during participants' interactions with a full humanoid robot. Individuals' sense of "we-ness" during collaborative tasks with other humans and its decline during human-machine joint tasks were demonstrated by numerous investigations (*Dewey, Pacherie & Knoblich, 2014; Obhi & Hall, 2011a; Obhi & Hall, 2011b*; *van der Wel, Sebanz, & Knoblich, 2012*). On the one hand, it has been shown that individuals could develop a sense of we-agency during joint actions with other humans (*Dewey, Pacherie & Knoblich, 2014; Obhi & Hall, 2011a; Obhi & Hall, 2011b*; *van der Wel, Sebanz, & Knoblich, 2012*). On the other hand, a loss of self-agency both for self-generated actions and other-generated actions performed in the context of a joint task with a computer has been demonstrated (**Experiment 1**, *Obhi & Hall, 2011b*). Few investigations focused on individuals' sense of agency for actions generated by a human-like machine (*Caspar, Desantis, Dienes, Cleeremans, & Haggard, 2016*). In Caspar and colleagues' study, the robotic hand was monitored by the participants' own actions. Indeed, they were wearing gloves with sensors that allowed the machine to move. In the robot homologous condition, the robotic hand performed the same action as participants. In the robot non-homologous condition, the robotic hand performed a different action than the participants. The authors reported that participants' exhibited a sense of agency as indicated by IB when the robotic hand triggered a sensory effect that was consistent with a previous learning session, only when the robotic hand performed the same action as participants (*Caspar, Desantis, Dienes, Cleeremans, & Haggard, 2016*). In this case, the sensory outcome was always generated by the joint actions of the machine and the participants. However, in **Experiment 3**, the robot-generated actions and outcomes did not involve participants at all as participants were responsible for one specific type of target and the humanoid robot for another specific type of target. Hence, we originally demonstrated here that participants could exhibit a sense of we-agency that was purely top-down driven during a joint task with a humanoid robot.

## 4.3. Individuals' agency is modulated by kinesthetic information during haptic feedback

Finally, using a bottom-up approach, we focused on the influence of the sensorimotor information generated by the machine on participants' sense of agency. In **Experiment 4**, participants were asked to manipulate a pair of haptic interfaces jointly with another human co-agent. Participants were told that they could receive haptic feedbacks from each other. Unknown to them, during half of the experiment, the sensorimotor feedback they received were generated by a robotic system. Participants' sense of we-agency was assessed using IB phenomenon and explicit judgments of agency. Our results showed that participants reported higher judgments of agency when they cooperated with the robot compared to the human co-agent. In contrast, at the implicit level, participants demonstrated a sense of we-agency only during the human-human joint actions in comparison with the human-machine joint actions. This sense of we-agency during the human-human joint actions was not modulated by social roles such as being the initiator of the follower of the joint action, which emphasizes the concept of *we*-agency. At the same time, we found that participants had more implicit motor communication with the human co-agent in comparison with the robotic system. We postulated that participants were better at extracting the intention of their co-agent during the human-human joint task compared the human-machine joint task, enhancing their sense of we-agency. Indeed, several studies pointed out that intentions were embedded in agents' kinematics and that individuals were able to read intentions from actions with the help of their action/observation matching system (*Iacoboni et al., 2005*; *Manera, Becchio, Cavallo, Sartori, & Castiello, 2011*; *Sebanz & Shiffrar, 2009*). Taken together, our results suggest that individuals are highly sensitive to sensorimotor information about their co-agent even when they operate at an unconscious level. Indeed, human-generated reafferences, but not machine-generated reafferences, led participants to exhibit a sense of we-agency, even though in both cases the participants thought they were cooperating with another human. These findings raise the important question of the motion laws that would need to be implemented in devices such as industrial robotic arms or haptic feedback handles in aircrafts for example (*Cheffi, Rakotomamonjy, Binet, Bidaud, & Sarrazin, 2015*). Indeed, the purpose of motorized haptic feedback in such apparatus is to make possible pilot assistance in degraded visual environments in order to avoid obstacles (*Binet & Rakotomamonjy, 2018*). However, ours findings suggest that individuals could not develop a sense of shared control (or we-agency) when they are given artificial-generated reafferences. Therefore, it one might think

that this type of artificial motorized feedback, although commonplace, is inappropriate in case the pilot need to take back control of the system as it keeps humans out of the loop of control. In addition, burdening the pilot with extra information that is inadequate might sustain the loss of situation awareness (*Régis, Dehais, Rachelson, Thooris, Pizziol, Causse, & Tessier, 2014*). The implementation of haptic feedbacks that follow biological motion law should arouse interest in future designs of haptic handles or robotic arms for assembling cells in industrial fields.

## 4.4. The multiple cue integrative model to explain the sense of we-agency

The multifactorial cues integrative model of the sense of agency proposed by Moore and Fletcher (2012) proved extremely useful in helping us explain the findings of our four experiments Indeed, as said earlier, Moore and Fletcher's model postulates that individuals' sense of agency emerged from the dynamic integration of both internal sensorimotor and external contextual cues. The influence of each cue in the construction of the sense of agency has been proposed to be dependent of its reliability, with a higher weighting for internal motoric cues by default. However, the authors conceded that contextual cues could be given greater weight when sensorimotor information is weak (*Moore & Fletcher, 2012*). We aimed at clarifying the contribution of contextual cues in the construction of individuals' sense of *we*-agency. In **Experiment 1**, **Experiment 2** and **Experiment 3**, participants were requested to perform a joint Simon task with a co-agent (visible or not). Participants had to detect one type of target with one specific action (e.g., to detect green dots with a right keypress) while their co-agent had to detect another type of target with another specific action (e.g., to detect red dots with a left keypress). In this case, due to the experimental setup, no direct low level sensorimotor information about the participants' co-agent (e.g., kinesthetic feedbacks) was available. Hence, the high-level contextual cues such as the belief about their co-agent during the joint task were the only available information that could modulate participants' sense of agency. The more social their co-agent was, the stronger the sense of agency participants exhibited (not manifested with sensory attenuation phenomenon) both for self- and other- generated actions. Specifically, we propose that when low-level sensorimotor information about the other's state is missing to inform individuals' sense of we-agency during a joint task, contextual cues such as the belief about a co-agent will stimulate the activation of the individual' action/observation

matching system if this co-agent is considered as an intentional and social agent. This attraction for social cues might probably originate from early neural wiring (*Castiello, Becchio, Zoia, Nelini, Sartori, et al., 2010*).

In **Experiment 4**, participants were requested to perform a joint action with a co-agent from whom they received sensorimotor feedbacks. Participants were instructed to push rightward or leftward a haptic interface and to join the lead if their co-agent initiated the move first. In this specific case, the direct availability of low-level motor cues about the performance of the co-agent led to a bottom-up functioning of the action/observation matching system (see *predictive coding account*, *Kilner, Friston, & Frith, 2007*; *Picard & Friston, 2014*) that allowed intention understanding. Their understanding of their co-agent's intentions led participants to experience a strong sense of we-agency during the human-human cooperative context. However, during the human-machine joint actions, even if participants thought they were cooperating with another human, tactile reafferences yielded by the robotic system failed to activate the participants' action/observation matching system, thus preventing intention understanding and weakening their sense of we-agency (see **Figure 26** for a summary).

In this view, the key concept to develop a sense of we-agency is the mutuality of action understanding and in particular the intention embedded in the other-generated action. Such comprehensive mechanisms have been proposed to rely on individuals' own motor experience through their action/observation matching system (*Sahaï, Pacherie, Grynszpan, & Berberian, 2017*). In accordance with the computational integrative model of agency by Moore and Fletcher (2012), we propose that sensorimotor cueing plays a major role in individuals' sense of we-agency.

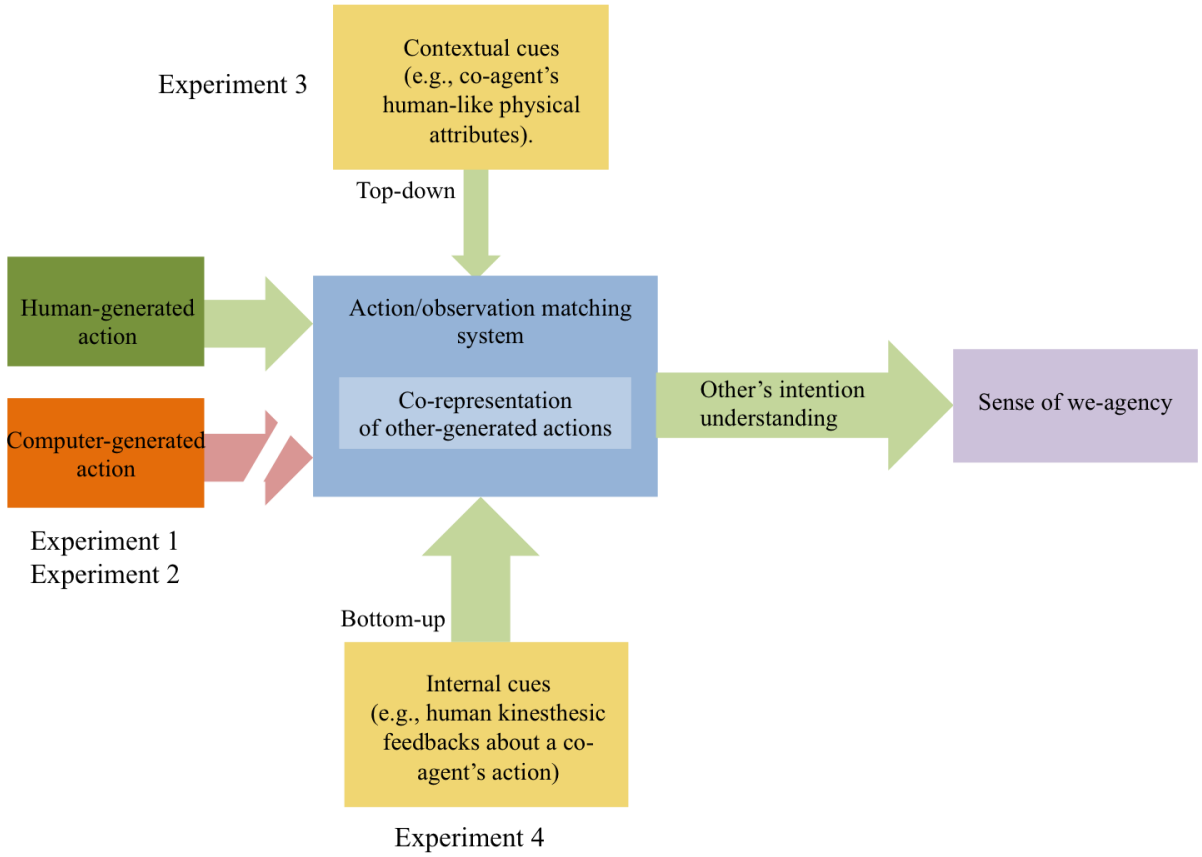


**Figure 26**. Balanced sheet of integrative cueing for the sense of we-agency.

## 4.5. The sense of agency and the out-of-the-loop (OOTL) problematic

This thesis work originated from a well-established operational issue, the out-of-the-loop (OOTL) problem encountered when humans are interacting with highly artificial automated systems (*Kaber, Onal, & Endsley, 2000; Parasuraman, Molloy, & Singh, 1993*; *Parasuraman & Riley, 1997*). Indeed, automation has steadily created a gap between the human operators and the loop of control, disconnecting them from the machine' actions and outcomes, hence causing a large set of OOTL performance problems (*Merat & Jamson, 2008*; *Thackray & Touchstone, 1989*; *Warm, Parasuraman, & Matthews, 2008*).

Cognitive engineering literature has discussed at length the OOTL phenomenon and the implication of different mechanisms in its emergence: decrements in vigilance such as reduced sensitivity to important signals, complacent or excessive trust in system ability, and loss of operator situation awareness have been pointed as potential contributor to this phenomenon (*Kaber, Onal, & Endsley, 2000*). In consequences, different solutions have been proposed by the human factors society.

Certainly, these different concepts (i.e., vigilance decrement, complacency and loss of situation awareness) have been useful in understanding human-machine performance in complex work environments. If we do not deny the scientific viability of these constructs, nor their diagnostic value in assisting human factor/engineering practitioners to formulate specific solutions, we also argue that these constructs lack a level of detail and thus fail to fully account for a psychological mechanism needed to reconnect the human in the loop of control. With regard to this, Billings (1996) nicely pointed this deficit with the situation awareness issue:

> *''The most serious shortcoming of the situation awareness construct as we have thought about it to date, however, is that it's too neat, too holistic and too seductive. We heard here that deficient situation awareness was a causal factor in many airline accidents associated with human error. We must avoid this trap: deficient situation awareness doesn't cause anything. Faulty spatial perception, diverted attention, inability to acquire data in the time available, deficient decision making, perhaps, but not a deficient abstraction!'' (Billings 1996)*

Two decades later, these different constructs have been partly linked to information processing or other psychological processes (for situation awareness, see *Durso & Sethumadhavan, 2008*; for trust, see *Lee & See, 2004*), but it remains that the underlying cognitive mechanisms of the OOTL phenomenon remains largely unknown. This PhD thesis aimed to introduce new tools and models to think about human-automation interaction.

First of all, we claimed that the understanding of the diminution of individuals' sense of agency during human-machine interactions provides a theoretical framework to better understand the OOTL operational phenomenon. Indeed, the Gordian knot of the OOTL performance problems might arise from individuals' impairments in building a strong sense of we-agency with (traditional) artificial automated systems. Yet, individuals' sense of agency during self-generated actions and outcomes has been proposed to be closely linked to their engagement in the task (*Bandura, 1999*). Therefore, the loss of agency experienced during human-machine interactions might sustain individuals' moral disengagement, leading to the OOTL performance problems classically reported.

Undeniably, designers could also benefit from the development of scientific knowledge in the science of agency. Firstly, the outcomes of this thesis could contribute to the assessment of whether individuals are in or out of the loop of control when testing devices. Indeed, as pointed out by Shneiderman and Plaisant (2004), users need the sense that they are in charge of machines and that the machines are responsive to their actions (*Shneiderman & Plaisant, 2004*). This sense of control could be assessed with IB phenomenon, as this implicit measure of individuals' sense of agency is sensitive to the cognitive distance individuals experience during the perception of machine-generated action outcomes (*Obhi & Hall, 2011b*; *Wohlschläger, Haggard, Gesierich, & Prinz, 2003*).

Secondly, this thesis furnished some guidelines for machine conception to keep humans in the loop of control. Indeed, as Norman pointed out, the key for designers is to "socialize our interactions with technology" (*Norman, 2010*). Considering that we use our own cognitive toolkit in order to understand and predict other-generated actions and outcomes, machine humanization can be proposed as a potential mean to overcome the loss of agency during human-machine interactions. This could be accomplished at a low-level of information processing with the implementation of biological motion laws in machines when it is possible, or at a higher level of abstraction with the conception of machine that hold a human-like physical appearance.

In sum, the contribution of this thesis in the study of the OOTL problem consists in three main aspects: the refined comprehension of the underlying processes involved in the OOTL performance problems; the proposal of an objective measure of OOTL episodes during human-machine interactions; and finally, the recommendations for machine designs.

## 4.6. Directions for future research

As said before, the main objective of the thesis was to get a more precise understanding of the underlying mechanisms involved in the loss of agency during human-machine interactions and to propose means to overcome this loss. In order to control parasite variables in our investigations, simple laboratory tasks (e.g., the Simon task, **Experiment 1**, **Experiment 2** and **Experiment 3***;* or left/right handles pushes, **Experiment 4**) were used. Although this does not detract from their relevance to the various issues raised in this thesis, they were not representative of the type of human-machine interactions that human operators experience in their daily life. Hence, it could be relevant to use more

ecological tasks with aircraft simulators for instance and in the same time, a targeted population (e.g., pilots). Indeed, the large range of stimuli displayed in rich environment such as aircraft simulators might generate a higher mental workload in human operators that is closer to reality.

Furthermore, our findings argue in favor of the implementation of new forms of automation to promote successful human-machine interactions. In particular, we think that machine humanizing could enhance individuals' sense of agency during their interactions with artificial systems. Obviously, not all forms of machine humanness are feasible in industrial artificial systems. On the one hand, it must be admitted that simulating the laws of biomechanics is far from simple. On the other hand, it is hard to envisage the presence of a full humanoid robot in aircrafts or air traffic control environments. However, empathy toward human-like machines has been proposed to be at the basis of social cooperation and prosocial behavior (*Riek, Rabinowitch, Chakrabarti, & Robinson, 2009*). Hence, the implementation of interfaces with conversational virtual agents might offer a good alternative in the socialization of our interactions with automated artificial system (*Pecune, Cafaro, Ochs, & Pelachaud, 2012).* It would be interesting to investigate whether such type of co-agents would reinforce individuals' sense of we-agency as humans have been shown to be sensitive to the observation of artificial humanoid avatars' actions (*Ferstl, Bülthoff, & de la Rosa, 2017*).

Finally, it would be relevant to examine the occurrence of mind-wandering episodes during interaction with human-like machines as this phenomenon has been closely linked with the OOTL problem (*Gouraud, Delorme, & Berberian, 2017*). Mind-wandering could be defined as the human mind's propensity to generate thoughts unrelated to the task at hand (*Christoff, 2012*). Therefore, the loss of sense of agency imposed by automation might be at the roots of mind-wandering episodes. It can be hypothesized that human-like machines would keep humans focused on the task because they would elicit their action/observation matching system during their cooperative interactions. Another aspect to be examined could be the impact of artificial systems humanization on individuals' performance during human-machine interactions. Indeed, previous research indicated that participants' error monitoring cerebral activity was reduced during the supervision of computer-generated errors compared to human-generated errors (*Somon, Campagne, Delorme, & Berberian, 2019*). It could be hypothesized that human-like machines would help regain similar error detection performances as during human-human interactions. Detecting when artificial

systems are failing and stepping in when such a failure arises is essential to improving safety in critical situations. Insuring that humans retain a sense of agency during human-machine joint tasks is therefore a major concern for optimizing interactions.

## 4.7. Conclusion

In a nutshell, one major consequence of automation is the emergence of new types of issues experienced by individuals during their interactions with artificial automated systems, namely, the OOTL performance problems (*Kaber, Onal, & Endsley, 2000; Parasuraman, Molloy, & Singh, 1993*; *Parasuraman & Riley, 1997*). In this thesis, we bring a cognitive explanation of this operational problem. Specifically, we proposed that individuals' loss of agency observed during their interactions with (traditional) artificial automated systems could be a contributor to the OOTL phenomenon. The challenge is then to try to allocate a fair sense of agency to each co-agent, with no preeminence of the human or the machine, to ensure that the human be able to step in when the automation fails. Because human-like automata are more likely to stimulate individuals' action/observation matching system (as opposed to traditional machines), human-like automata may facilitate action co-representation, which in turn would mediates individuals' sense of we-agency during joint tasks. However, further investigations are needed to examine the reallocation of agency with this kind of new machines in applied contexts. Last but not least, this thesis work furnishes possible solutions for how human-machine interactions should be approached for the sake of keeping the human user in the loop of control.

# Appendixes

**Appendix 1**: Answers of participants to the question of the debriefing interview These answers were transcribed and translated from French into English. (Experiment 2).

| **(a) Do you have any comment regarding the experiment?** | **(b) Did you notice any differences in the co-agent performance regarding it was another human or the desktop computer? If yes, please explain** | **(c) Have you got the impression that your co-agent did not perform the task with you actually?** | **(d) In fact, nobody was performing the task with you and your co-agent detections were always triggered by the computer. Did you suspect that?** |
|---|---|---|---|
| "It was good to have some breaks between the experimental blocks." | "No, I did not notice anything." | "No." | "No." |
| "The tasks were repetitive." | "No." | "No." | "No." |
| "Not especially." | "No." | "No." | "No." |
| "No." | "No, they made the same amount of errors." | "No." | "No." |
| "The tasks were repetitive". | "Yes, the human partner made more errors than the machine." | "No." | "No." |
| "It was hard to stay focused." | "No." | "No." | "No." |
| "No." | "No." | "No." | "No." |

| | | | |
|---|---|---|---|
| “No, I am curious to know more about it.” | “No.” | “No.” | “No.” |
| “I was very tired during the last experimental block.” | “No.” | “No.” | “No.” |
| “The experiment is tiresome.” | “No but the human co-agent made more noise than the machine.” | “No.” | “No.” |
| “The tasks were too repetitive and too long.” | “No.” | “No.” | “No.” |
| “It was tiresome to stay in the darkness”. | “It is strange but the machine made more errors than the human co-agent. It was supposed to the opposite.” | “No” | “I had some doubts at the beginning of the experiment about a possible trick because I ran a lot of experiments in social psychology but then I performed the task and I did not notice anything.” |
| “No.” | “No.” | “No.” | “No.” |
| “No.” | “The human made more errors.” | “No.” | “No.” |
| “It was too long.” | “No.” | “No.” | “No.” |
| “The tasks were very repetitive.” | “No.” | “No.” | “No.” |
| “The tasks were easy but too long.” | “Not especially.” | “No.” | “No.” |

**Appendix 2**: Answers of included participants to the first question of the debriefing interview ("Do you have any comment regarding the experiment?). These answers were transcribed and translated from French into English (Experiment 4).

| |
|---|
| Sometimes, we often disagreed. One of us started, the other one did not follow and finally we sometimes would run out of time. Well, only two or three times, otherwise it was ok. The white noise was tiring after a while. |
| I do not know. I thought it was always the same time between the small click and the beep. Well, except in some cases when it changed. Otherwise, I believe it was fun. |
| I thought the noise was disturbing. |
| No, not really. But then, I still do not understand the purpose of the experiment. But I find it interesting. We want to know, we want to go deeper. Because, after a while, we know what we are doing but we do not know the goal of the experiment. We believe things and then we are waiting for an answer. And then, no, I have nothing to say. |
| The noise, it's not super nice. It gets tiring to do all those repetitions. That's all. |
| Yes, actually, in this part, after we act, there are two choices, that is, estimating the duration between the stopper and the sound and then estimating responsibility. I think that the longer I evaluated the duration to be, the lesser I felt responsibility. |
| Nothing particular |
| We lose track of time a bit, especially when time has to be evaluated. I had the impression that when the sound came faster, I would rate each of us as 50/50. I had the impression that when we were more together, the time was shorter. I noticed that I would often move the two gauges similarly. I mean… well… or the opposite, the longer it was, the more I felt he was leading. |
| Not right now. It's true that it was a bit repetitive. In the beginning, I wasn't sure I understood the instructions and what was expected from me. Then it was ok. The first phase of the experiment was useful to actually understand what was expected in terms of instructions. |

| |
|---|
| Nothing |
| Not really. Well, actually, I did not really have the impression of feeling what he was doing on the handle. Sometimes, I felt it clearly and sometimes I had an impression as if his finger was not placed on it. |
| Well, I find that estimating time is difficult. Mainly, this is what I found was not easy to find. |
| I think that the noise… Well, then, it might be just me, because I rarely listen to music with headphones. |
| For the two experiments that lasted more than 20 minutes, I would say that from the middle onwards my concentration was often less important than at the beginning. I could not concentrate on the two elements, that is, on the duration and whether it was really me who initiated the movement or not. However, when only duration had to be estimated, then I felt more comfortable, for instance. |
| No that's it. Difficult to concentrate on the long run. Well, after some time I felt tired. This is why I had the impression of not being as concentrated. That's all. |
| The only thing is related to the sound system. Regarding the exercise and the experiment itself, the training is not adequate. I found it harder when we did the combined experiment together than during training. Apart from that, it's ok. |
| I think I noticed something: When we were together and we had to concentrate on following the other or not, I found that the time between arriving at the end and the beep was shorter. It never went very far. Well I think, just that. |
| No. Tiredness is annoying. I had the impression that the effect was the opposite of what was expected. If the goal is that we would become more and more synchronous, then the exact opposite happened. |
| I had the impression that I was always the one to choose whether the sound was high or low. Even when my partner decided to go toward the high pitch sound, if I decided to go for the low, I went back to the low. |
| I was less able to focus on the sound. I preferred focusing on the distribution of reactions rather than the sound. Therefore, I was more concentrated on this. Then, I wonder how this will be useful for you. |

| |
|---|
| No |
| Not really. Well, the part with the sound was more complicated. This is the thing where it was the hardest for me to really give the right one. |
| It becomes more and more difficult to perceive the time scales. I was thinking about a clock. I imagined the clock hand. Truly, when it happens like that. I think that I didn't estimate as well at the end. Then, regarding the finger, I think that following or generating the movement changes our perception of the noise. |
| During the first part, I was leading. In the second part, I noticed that, even if I didn't put my hand, it moved. Especially in the second part, when I was responding, as soon as I took a little bit of time to adjust, he would take over. |
| It was mostly when we had our finger on the handle, when we had to reach the stopper, at the beginning, it was not very easy to differentiate whether the other participant or myself was going in one direction or the other. As the experiment continued, it became clearer and clearer that he was pushing in one direction or that I was pushing in one direction. I do not know if this is because he started pushing more rapidly or with more conviction. |

# References

Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and cognition*, *14*(3), 439-458. doi : 10.1016/j.concog.2004.11.001

Balconi, M. (2010). *Neuropsychology of the Sense of Agency*. New York, NY: Nova Science Publishers.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and social psychology review*, *3*(3), 193-209.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of personality and social psychology*, *71*(2), 364. doi : 10.1037/0022-3514.71.2.364

Baron, R. A. (1988). Negative effects of destructive criticism: Impact on conflict, self-efficacy, and task performance. *Journal of Applied Psychology*, *73*(2), 199.

Billings, C. E. (1996). Human-centered aviation automation: Principles and guidelines.

Billings, C. E. (1997). Aviation automation: The search for a human centered approach.

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.

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. doi : 10.1093/scan/nsw160

Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself?. *Neuroreport*, *11*(11). R11

Bokura, H., Yamaguchi, S., & Kobayashi, S. (2001). Electrophysiological correlates for response inhibition in a Go/NoGo task. *Clinical Neurophysiology*, *112*(12), 2224-2232.

Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion changes the sense of agency in the human brain. *Current biology*, *26*(5), 585-592. doi : 10.1016/j.cub.2015.12.067

Caspar, E. A., Desantis, A., Dienes, Z., Cleeremans, A., & Haggard, P. (2016). The sense of agency as tracking control. *PloS one*, *11*(10), e0163892. doi : 10.1371/journal.pone.0163892

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2004). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cerebral cortex*, *15*(8), 1243-1249. doi : 10.1093/cercor/bhi007

Castiello, U., Becchio, C., Zoia, S., Nelini, C., Sartori, L., Blason, L., ... & Gallese, V. (2010). Wired to be social: the ontogeny of human interaction. *PloS one*, *5*(10), e13199.

Chambon, V., & Haggard, P. (2012). Sense of control depends on fluency of action selection, not motor performance. *Cognition*, *125*(3), 441-451. doi : 10.1016/j.cognition.2012.07.011

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. doi : 10.3389/fnhum.2014.00320

Courgeon, M., Rautureau, G., Martin, J. C., & Grynszpan, O. (2014). Joint attention simulation using eye-tracking and virtual humans. *IEEE Transactions on Affective Computing*, *5*(3), 238-250.

Coyle, D., Moore, J., Kristensson, P. O., Fletcher, P., & Blackwell, A. (2012, May). I did that! Measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2025-2034). ACM.

Craighero, L., Jacono, M., & Mele, S. (2016). Resonating with the ghost of a hand: a TMS experiment. *Neuropsychologia*, *84*, 181-192.

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.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Desantis, A., Roussel, C., & Waszak, F. (2011). On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition*, *20*(4), 1211-1220. doi : 10.1016/j.concog.2011.02.012

Desantis, A., Weiss, C., Schütz-Bosbach, S., & Waszak, F. (2012). Believing and perceiving: authorship belief modulates sensory attenuation. *PLoS One*, *7*(5), e37959. doi : 10.1371/journal.pone.0037959

Dewey, J. A., & Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? *PloS one*, *9*(10), e110118. doi : 10.1371/journal.pone.0110118

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. (2014). The joint Simon effect: a review and theoretical integration. *Frontiers in Psychology*, *5*, 974. doi : 10.3389/fpsyg.2014.00974

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

Durso, F. T., & Sethumadhavan, A. (2008). Situation awareness: Understanding dynamic environments. *Human Factors*, *50*(3), 442-448.

Endsley, M. R. (2016). *Designing for situation awareness: An approach to user-centered design*. CRC press.

Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors*, *37*(2), 381-394. doi : 10.1518/001872095779064555

Ferstl, Y., Bülthoff, H., & de la Rosa, S. (2017). Action recognition is sensitive to the identity of the actor. *Cognition*, *166*, 201-206. doi : 10.1016/j.cognition.2017.05.036

Frith, C. D. (2005). The self in action: lessons from delusions of control. *Consciousness and cognition*, *14*(4), 752-770.

Frith, C. D. (2014). *The cognitive neuropsychology of schizophrenia*. Psychology press.

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

Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage* 35, 1674-1684. doi: 10.1016/j.neuroimage.2007.02.003

Grynszpan, O., Weiss, P. L., Perez-Diaz, F., & Gal, E. (2014). Innovative technology-based interventions for autism spectrum disorders: a meta-analysis. *Autism*, *18*(4), 346-361.

Haggard, P. (2005). Conscious intention and motor cognition. *Trends in cognitive sciences*, *9*(6), 290-295. doi : 10.1016/j.tics.2005.04.012

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 196.

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology*, *22*(10), R390-R392. doi : 10.1016/j.cub.2012.02.040

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

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS biology*, *3*(3), e79.

Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human factors and ergonomics in manufacturing & service industries*, *10*(4), 409-430.

Keysers, C., & Gazzola, V. (2009). Expanding the mirror: vicarious activity for actions, emotions, and sensations. *Current opinion in neurobiology*, *19*(6), 666-671.

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

Lee, J. D. (2006). Affect, attention, and automation. *Attention: From theory to practice*, 73-89.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50-80.

Limerick, H., Coyle, D., & Moore, J. W. (2014). The experience of agency in human-computer interactions: a review. *Frontiers in human neuroscience*, *8*, 643. doi : 10.3389/fnhum.2014.00643

Loehr, J. D. (2013). Sensory attenuation for jointly produced action effects. *Frontiers in psychology*, *4*, 172.

Mann, D. T., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology*, *29*(4), 457-478.

Merat, N., & Jamson, A. H. (2008). The effect of stimulus modality on signal detection: Implications for assessing the safety of in-vehicle technology. *Human Factors*, *50*(1), 145-158. doi : 10.1518/001872008X250656

Moore, J. W., Teufel, C., Subramaniam, N., Davis, G., & Fletcher, P. C. (2013). Attribution of intentional causation influences the perception of observed movements: behavioral evidence and neural correlates. *Frontiers in psychology*, *4*, 23. doi : 10.3389/fpsyg.2013.00023

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

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and cognition*, *18*(4), 1056-1064.

Moretto, G., Walsh, E., & Haggard, P. (2011). Experience of agency and sense of responsibility. *Consciousness and cognition*, *20*(4), 1847-1854.

Norman, D. A. (1986). Cognitive engineering. *User centered system design*, *31*, 61.

Norman, D. A., (2010). Living with complexity. Cambridge, Massachusetts: The MIT Press

Niedermeier, D., & Lambregts, A. A. (2012). Fly-by-wire augmented manual control-basic design considerations. In *International Congress of the Aeronautical Sciences* (Vol. 100, p. 7).

Obhi, S., S., and Hall, P. (2011a). Sense of agency and intentional binding in joint action. *Experimental Brain Research* 211(3-4), 655–662. doi: 10.1007/s00221-011-2675-2

Obhi, S., S., and 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. doi : ijn_00352565

Pacherie, E. (2012). Action. In The Cambridge handbook of cognitive science, Cambridge University Press (Ed.), 92-111.

Pacherie, E., & Dokic, J. (2006). From mirror neurons to joint actions. *Cognitive Systems Research*, *7*(2-3), 101-112. doi : 10.1016/j.cogsys.2005.11.012

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced 'complacency'. *The International Journal of Aviation Psychology*, *3*(1), 1-23. doi : 10.1207/s15327108ijap0301_1

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230-253. doi : 10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297. doi : 10.1109/3468.844354

Pecune, F., Cafaro, A., Ochs, M., & Pelachaud, C. (2016, September). Evaluating Social Attitudes of a Virtual Tutor. In *International Conference on Intelligent Virtual Agents* (pp. 245-255). Springer, Cham.

Perani, D., Fazio, F., Borghese, N., A., Tettamanti, M., Ferrari, S., Decety, J., et al. . (2001). Different brain correlates for watching real and virtual hand actions. *Neuroimage* 14, 749-758. doi : 10.1006/nimg.2001.0872

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

Pineda, J. A., Allison, B. Z., & Vankov, A. (2000). The effects of self-movement, observation, and imagination on/spl mu/rhythms and readiness potentials (RP's): toward a brain-computer interface (BCI). *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 219-222. doi : 10.1109/86.847822

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.

Poonian, S. K., & Cunnington, R. (2013). Intentional binding in self-made and observed actions. *Experimental brain research*, *229*(3), 419-427. doi : 10.1007/s00221-013-3505-5

Poonian, S. K., McFadyen, J., Ogden, J., & Cunnington, R. (2015). Implicit agency in observed actions: evidence for N1 suppression of tones caused by self-made and observed actions. *Journal of cognitive neuroscience*, *27*(4), 752-764. doi : 10.1162/jocn_a_00745

Ravenet, B., Ochs, M., & Pelachaud, C. (2012, November). A computational model of social attitude effects on the nonverbal behavior for a relational agent. In *Workshop Affect Compagnon Artificiel Interaction*.

Richerson, P. J., & Boyd, R. (1997). Homage to Malthus, Ricardo, and Boserup: Toward a general theory of population, economic growth, environmental deterioration, wealth, and poverty. *Human Ecology Review*, 85-90.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, *27*, 169-192.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews neuroscience*, *2*(9), 661. doi : 10.1038/35090060

Roche, L., & Saint-Bauzel, L. (2016, October). Implementation of haptic communication in comanipulative tasks: a statistical state machine model. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 2670-2675). IEEE.

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.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, *2*, 1926-1943.

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 : 10.1162/jocn_a_00745

Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of cognitive neuroscience*, *18*(5), 803-817. doi : 10.1162/jocn.2006.18.5.803

Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic bulletin & review*, *16*(1), 170-175.

Searle, J. R., & Willis, S. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge university press.

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

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

Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: An ERP study of action planning and control in coacting individuals. *Journal of cognitive neuroscience*, *18*(5), 859-870. doi : 10.1162/jocn.2006.18.5.859

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

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

Sidarus, N., Vuorre, M., & Haggard, P. (2017). How action selection influences the sense of agency: an ERP study. *NeuroImage*, *150*, 1-13. doi : 10.1016/j.neuroimage.2017.02.015

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.

Stenzel, A., Chinellato, Tirado Bou, & del Pobil, (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

Stevens, A., Horberry, T., & Regan, M. A. (2017). Driver Acceptance of New Technology: Synthesis and Perspectives. *Driver Acceptance of New Technology: Theory, Measurement and Optimisation*, 335.

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

Tai, Y. F., Scherfler, C., Brooks, D. J., Sawamoto, N., & Castiello, U. (2004). The human premotor cortex is' mirror'only for biological actions. *Current biology*, *14*(2), 117-120.

Tekok-Kilic, A., Shucard, J. L., & Shucard, D. W. (2001). Stimulus modality and Go/NoGo effects on P3 during parallel visual and auditory continuous performance tasks. *Psychophysiology*, *38*(3), 578-589. doi : 10.1017/S0048577201991279

Thackray, R. I., & Touchstone, R. M. (1989). Detection efficiency on an air traffic control monitoring task with and without computer aiding. *Aviation, Space, and Environmental Medicine*.

Tsai, C. C., & Brass, M. (2007). Does the human motor system simulate Pinocchio's actions? Coacting with a human hand versus a wooden hand in a dyadic interaction. *Psychological Science*, *18*(12), 1058-1062. doi : 10.1111/j.1467-9280.2007.02025.x

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

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.

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

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human factors*, *50*(3), 433-441. doi : 10.1518/001872008X312152

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American psychologist*, *54*(7), 480. doi : 10.1037/0003-066X.54.7.480

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011). The self in action effects: selective attenuation of self-generated sounds. *Cognition*, *121*(2), 207-218. doi : 10.1016/j.cognition.2011.06.011

Wen, T., & Hsieh, S. (2015). Neuroimaging of the joint Simon effect with believed biological and non-biological co-actors. Frontiers in human neuroscience, 9, 483.

Wiley, D. A. (2002). The coming collision between the automated instruction and learning communities camps of online learning research. *Working Draft Retrieved June*, *22*, 2003.

Wohlschläger, A., Engbert, K., & Haggard, P. (2003). Intentionality as a constituting condition for the own self—and other selves. *Consciousness and cognition*, *12*(4), 708-716. doi : 10.1016/S1053-8100(03)00083-7

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. doi : 10.1046/j.0956-7976.2003.psci_1469.x

Woods, D. D., & Tinapple, D. (1999, September). W3: Watching human factors watch people at work. In *Presidential address, presented at the 43rd Annual Meeting of the Human Factors and Ergonomics Society, Houston, TX.*

## Agentivité conjointe lors des Interactions Homme-Machine : comment concevoir des agents plus coopératifs ?

A travers l'automatisation des systèmes, les mutations technologiques ont progressivement éloigné l'opérateur humain des actions et des effets générés par les machines. Ce phénomène de « sortie de boucle » (Kaber, Onal, & Endsley, 2000) a généré un ensemble de difficultés pour l'opérateur humain qui restent encore aujourd'hui difficile à appréhender et compenser. Au cours de cette thèse, nous avons proposé un cadre théorique, celui de l'Agentivité, permettant d'expliquer et de compenser ces difficultés. Le sens d'agentivité peut se définir comme le sentiment de contrôle de ses propres actions, et à travers elles, des évènements de l'environnement extérieur (Haggard & Chambon, 2012). De manière intéressante, il a été montré qu'au cours des interactions purement humaines, les individus pouvaient ressentir un sens d'agentivité pour des actions et effets générés par autrui (ou « sens d'agentivité conjoint»). En revanche, la capacité à développer un sens d'agentivité conjoint et à se représenter les actions générées par un système artificiel automatisé semble altérée lors des interactions homme-machine (Obhi & Hall, 2011b).
La première partie de la thèse a consisté à examiner finement les mécanismes sous-jacents de cette perte d'agentivité lors de tâches conjointes avec les systèmes automatisés, à la fois au niveau comportemental (Expérience 1) et au niveau cérébral (Expérience 2). En utilisant une tâche de Simon conjoint couplée à une tâche de liage intentionnel, nous avons montré que la capacité à développer un sens d'agentivité conjoint allait de pair avec la capacité de se représenter dans son propre système cognitif les actions générées par un partenaire. Nous avons également montré que ces deux processus étaient altérés lors de nos interactions avec une machine.
La deuxième partie de la thèse visait à déterminer sur quelles propriétés de la machine il était possible d'agir pour créer un sens d'agentivité conjoint avec des systèmes artificiels automatisés. Premièrement, en utilisant une approche top-down, nous avons montré que l'apparence physique humanoïde d'une machine atténuait la perte d'agentivité classiquement reportée lors des interactions homme-machine (Expérience 3). Deuxièmement, en utilisant une approche bottom-up, nous avons montré que le sens d'agentivité conjoint était sensible aux informations sensorimotrices générées par des retours haptiques (Expérience 4). Les implications de nos résultats sont discutées au regard de la littérature sur l'Agentivité et de la problématique opérationnelle du phénomène de « sortie de boucle ».

**Mots-clés :** INTERACTION HOMME-ROBOT ; ACTION CONJOINTE ; SORTIE DE BOUCLE ; AGENTIVITE ; LIAGE INTENTIONNEL ; CO-REPRÉSENTATION

## Joint agency in Human-Machine Interactions: How to design more cooperative agents ?

System automation has steadily created a gap between the human operators and the loop of control (i.e., "out-of-the-loop" (OOTL) problem), disconnecting them from the machines' actions and outcomes (Kaber, Onal, & Endsley, 2000). In this thesis, we aimed at investigating how to keep the human operators in the loop of control. We based our investigations on the theoretical framework of the science of agency. The sense of agency can be defined as the experience of controlling one's own actions, and, through them, the events in the outside world (Haggard & Chambon, 2012). Interestingly, it has been shown that during human-human interactions, individuals could exhibit a sense of agency for other-generated actions and outcomes (or sense of "we-agency") while such ability was impaired for machine-generated actions and outcomes (Obhi & Hall, 2011b).
The first stage of the thesis sought to finely examine the cognitive processes underlying individuals' loss of agency during joint tasks with automated artificial systems, both at the behavioral (Experiment 1) and at the cerebral (Experiment 2) levels. Using a joint Simon task coupled with an intentional binding task, we found that the ability to develop a sense of agency for other-generated action outcomes was in pace with the ability to represent other-generated actions into one own cognitive system, which were both impaired during human-machine interactions.
The second stage of the thesis sought to investigate on which characteristics of the machine it was possible to act in order to regain the human operator's sense of agency. Firstly, using a top-down approach, we found that humanoid machines could mitigate the loss of agency during joint tasks (Experiment 3). Secondly, using a bottom-up approach, we found that individuals' sense of agency was sensitive to kinesthetic feedbacks (Experiment 4). Implications of our findings are discussed in regard with the literature on the sense of agency and the operational OOTL issue.

**Keywords :** HUMAN-ROBOT INTERACTION ; JOINT ACTION ; OUT-OF-THE-LOOP ; AGENCY ; INTENTIONAL BINDING ; SHARED REPRESENTATION

ONERA
THE FRENCH AEROSPACE LAB

BP 80100 - Chemin de la Hunière FR-91123 PALAISEAU CEDEX - Tél. : +33 1 80 38 60 60 - Fax : +33 1 80 38 65 10

www.onera.fr