# The sense of agency during Human-Agent Collaboration

Maria Dagioglou
mdagiogl@iit.demokritos.gr
National Centre for Scientific Research 'Demokritos'
Ag. Paraskevi, Greece

Vangelis Karkaletsis
National Centre for Scientific Research 'Demokritos'
Ag. Paraskevi, Greece

## ABSTRACT

In collaborative set-ups, humans and artificial intelligence agents must team-up to achieve common goals. The identity of an agent, meaning its perceived mind and embodiment, will influence human behaviour. The experienced sense of agency, especially in the case of collaboration and action coordination, are expected to affect cooperativeness, team performance and fluency. Recent advances in deep Reinforcement Learning can now facilitate a series of human-agent and human-robot collaboration studies that will allow a better understanding about how the identity attributed to an agent impacts the performance of the partners. In this short paper we discuss some study paradigms we are building to explore the role of embodiment and agent's behaviour to self- and joint- agency experienced by human partners during collaborative scenarios.

## CCS CONCEPTS

• **Human-centered computing** → **Collaborative interaction**; • **Computer systems organization** → **Robotics**; • **Computing methodologies** → **Artificial intelligence**; **Machine learning**.

## KEYWORDS

Human-Agent/Robot Collaboration; Collaborative learning; Sense of Agency; Embodiment

**ACM Reference Format:**
Maria Dagioglou and Vangelis Karkaletsis. 2021. The sense of agency during Human-Agent Collaboration. In *HRI 2021 Workshop: Robo-Identity: Artificial identity and multi-embodiment,March 08, Boulder (Virtual), USA* . ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Successful adoption of human-agent and human-robot interaction systems depends, among other things, on the extent that agent behavior encourages an uninterrupted and natural human behavior within a specific context. For example, agents that assume assistive or companionship roles can be more acceptable if they evoke social attitudes similar to those governing human-human interactions. Both human-like embodiment and behavior play an important role in this [12, 22]. Beyond instruction type of interactions [2], in many applications, such as games [7, 19], industrial work-floors [13, 21] and rehabilitation set-ups [5], humans and Artificial Intelligence (AI) agents, embodied or not, will team-up to achieve common goals. In these cases, robot behavior should facilitate and support joint action. What is an acceptable collaborative agent behavior in scenarios where there is no embodiment at all or the embodiment answers to task requirements that cannot be satisfied with a human-like embodiment?

It is now quite appreciated that in order to build effective collaborative environments, it is necessary to consider human-AI *team performance*, as opposed to isolated AI performance [4], and *human behavior* [3]. A prerequisite for being able to account for expected human behavior in human-AI teams is to understand how humans perceive and respond to AI agents. Humans during their interaction with non-human entities, via the process of anthropomorphization, attribute to them mental and affective states [1]. Such processes also apply to human-agent/robot interactions. Depending on the agent's complexity of behavior and/or embodiment, a human can perceive an agent as an intentional one [16, 22]. Especially in the case of collaboration between a human and an AI agent, attributed identity is expected to affect human actions and as result the performance, the efficiency and the fluency of the team towards achieving the common goal. For example, agent's errors change the behavior of humans during a collaborative game depending on how these errors influence the performance or the reliance [7].

During human-human joint action, people recruit several mechanisms to successfully accomplish a common task. These include joint attention, monitoring of collaborator's actions, shared task representations and spatio-temporal action coordination [18]. Additionally, cooperativeness and team performance are affected by the sense of self- and joint- agency experienced during a joint task. Perceptual distinctiveness of each actor's actions is a factor that influences the sense of control during joint action, as well as the competitiveness or complementarity of partners' roles [8]. Another factor that alters the sense of agency is the fairness of resource distribution, that is how rewards are shared between partners [15].

Similar to other human-human interaction processes [14], some of the mechanisms that people recruit during joint action are expected to be similar to human-agent joint action. For example, interacting with robotic agents seems to affect the sense of agency in a similar way compared to interacting with other humans [6]. However, when joint action involves kinesthetic cues it appears that the movements of a robot partner have a different effect on the sense of agency of a human collaborator compared to movements of a human partner [9]. Predictive mechanisms, or lack thereof, can also have an impact to the collaboration of humans and automated artificial systems [17]. The studies mentioned above emphasize the need for further research regarding how the identity of an agent, its perceived mind and the capabilities of its embodiment, affect human partner action in collaborative systems.


*Robo-Identity Workshop HRI 2021, March 08, Boulder (Virtual), USA*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Recent advances in deep Reinforcement Learning (RL) have demonstrated very promising results in real-world and real-time learning problems [10]. Shafti and colleagues [20] have actually applied a deep RL method to a co-learning paradigm between a human and a Universal Robots UR10 cobot. In their paradigm a tray attached to the end-effector of the cobot is able to rotate around two axes. The human and the robot, each controlling the rotation around one axis, have to learn to collaboratively guide a ball through the tray's obstacles from a starting point to an end goal. The authors experiment with different network parameters as well as with multiple participants and examine, among others, human-robot team performance when dealing with agents of other participants.

Such paradigms, like the one in [20], present the opportunity to study human-agent joint action in real-time and explore how human actions and mechanisms during joint action are affected by an agent's behavior and embodiment. Moreover, co-learning paradigms offer the added value of observing shifts of partners' behavior while they mutually learn and adapt to each other. In the following section we discuss some of the research activities we are currently planning and building, exploiting the co-learning paradigm presented in [20] to investigate how the sense of agency of the human partner is affected when dealing with different versions of agents and environmental conditions.

## 2 RESEARCH METHODS

We first describe three different experimental conditions where humans collaborate with agents that may or may not have an embodiment and where the environment to control is virtual or real. Moreover, we consider the evaluation tools that can be used to measure team performance and collaboration. Finally, we discuss some further design considerations.

In a *virtual-world/mind* condition of the co-learning paradigm [20], a 3D graphic representation of the titling table that simulates the rolling ball physics serves as the environment within which the collaboration occurs. The participants are asked to *collaborate with an AI agent* in order to guide a ball from the starting position to the goal.

In the *real-world/embodied mind* condition of the paradigm, similar to [20], an actual tray is attached to the end-effector of a Universal Robots UR3 cobot. The participants are asked to *collaborate with the cobot* to achieve the goal. They hold in their hands a small tray that is tracked using optical markers. The participants rotate their tray around the axis they control, and these rotations are then mapped to rotations of the cobot tray. However, the feeling of control of the human actor might be affected by the fact that the tray is actually attached to the body of the robot. So in some sense the robot can be considered responsible for executing correctly the actions of its human partner.

To account for such effects, in a third *virtual-world/embodied mind* condition of the study the virtual environment of the titling table is presented to the human partners and they are asked to *collaborate with the cobot* that operates in their proximity and controls the rotation of the other axis of the tray through its movements.

In every condition, team performance will be evaluated using the time spent to achieve the goal. Moreover, participants will evaluate, throughout the learning process, their feeling of control [8] in the task during test trials where the agent is frozen. Finally, human-agent collaboration will be also evaluated based on subjective measures such as fluency, agent contribution and trust [11].

We expect that the experimental conditions described above will shed more light on how, during a collaborative task, the human actor's sense of agency is affected by whether the AI agents are embodied or not. It would also be interesting to explore the effect of a different robot embodiment, like an anthropomorphic one, or to explore if human behavior in the virtual-world/mind condition is affected when we give the agent a face, for example by showing to the participants the image of a robot.

*Design considerations.* A factor that can affect human behavior in the previous paradigms is the type of actions the human partner uses to rotate the tray. For example, human actions can either be conveyed by using specified keys on a keyboard to control the direction of tray rotation or by mapping human movements to tray angles [20]. In the latter case, learning to control one's own movements can increase the difficulty of the task and the overall learning period. In any case, it is also important to consider whether human actions are applied in a continuous or discrete way as this will affect human perception about the causal effect of one's actions, and as a result, the judgement of agency. Finally, hardware limitations and computational processing time can also impose constraints related to delays, human movement accuracy, etc. that will affect design decisions.

Human and agent behavior can be further manipulated in various ways. Introducing noise to chosen actions [8] will impact the expected ball movement and will increase the uncertainty about the dynamics of the system. In what way such external perturbations affect the sense of agency and the human-agent team performance? Finally, another interesting dimension to experiment with, especially in the context of deep RL, is reward attribution [15]. What is the impact of symmetrical (or not) rewards in the course of learning and as a result in self- or joint- judgement of control?

## 3 CONCLUSIONS

Understanding human behavior toward artificial agents is crucial for designing agents that promote fluent human behavior and as a result for the adoption of AI systems. In the present short paper we discussed some of the open questions regarding how attributed identity to an agent can impact human-agent/robot collaboration. How can an agent's perceived mind and embodiment affect self- and joint- sense of agency during joint actions and as a result co-learning and team performance? In pursuit of such questions, we presented a series of studies that we are currently building and discussed some methodological issues that need to be considered, as well as several manipulations that can be exploited to deepen our understanding of how artificial identity is experienced in human-agent joint action.

## ACKNOWLEDGMENTS

This work was supported by the 'Stavros Niarchos Foundation' Industrial Post-doc Fellowship of NCSR 'Demokritos' on *Human-Robot Collaboration: human collaborator representation for robot autonomous decisions* (http://roboskel.iit.demokritos.gr/project/chorus.html).

## SHORT BIOS

**Maria Dagioglou** studied Electrical and Computer Engineering (Democritus Univ. of Thrace, GR, 2007). She holds a MSc in Biomedical Engineering (TU Delft, NL, 2010) and a PhD in Psychology (Univ. of Birmingham, UK, 2014). Since 2014, she has been a research associate at NCSR-D, participating at the Roboskel activity (http://roboskel.iit.demokritos.gr/) and assuming team member roles in national and EU projects. She is currently a Stavros Niarchos Foundation Industrial Post-Doc fellow working on robotics applications for Human-Robot Collaboration at shared workspaces. Her research interests vary from Human-Robot Collaboration to Human Motor Control and Learning.

**Vangelis Karkaletsis** is the Director of the Institute of Informatics Telecommunications (II) at NCSR Demokritos. His research interests are in the areas of content analysis, big data management, knowledge representation, human-machine interaction. He acted as Coordinator, scientific and technical manager of many European and national projects and organised numerous of international conferences, workshops and summer schools (http://karkaletsis.iit.demokritos.gr/).

## REFERENCES


[1] Gabriella Airenti. 2018. The development of anthropomorphism in interaction: intersubjectivity, imagination, and theory of mind. *Frontiers in psychology* 9 (2018), 2136.

[2] Judith Bütepage and Danica Kragic. 2017. Human-robot collaboration: from psychology to social robotics. *arXiv preprint arXiv:1705.10146* (2017).

[3] Micah Carroll, Rohin Shah, Mark K Ho, Thomas L Griffiths, Sanjit A Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-ai coordination. *arXiv preprint arXiv:1910.05789* (2019).

[4] Prithvijit Chattopadhyay, Deshraj Yadav, Viraj Prabhu, Arjun Chandrasekaran, Abhishek Das, Stefan Lee, Dhruv Batra, and Devi Parikh. 2017. Evaluating visual conversational agents via cooperative human-ai games. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 5.

[5] Giorgia Chiriatti, Giacomo Palmieri, and Matteo Claudio Palpacelli. 2020. A framework for the study of human-robot collaboration in rehabilitation practices. In *International Conference on Robotics in Alpe-Adria Danube Region*. Springer, 190–198.

[6] Francesca Ciardo, Frederike Beyer, Davide De Tommaso, and Agnieszka Wykowska. 2020. Attribution of intentional agency towards robots reduces one's own sense of agency. *Cognition* 194 (2020), 104109.

[7] Sylvain Daronnat, Leif Azzopardi, and Martin Halvey. 2020. Impact of agents' errors on performance, reliance and trust in human-agent collaboration. In *Human Factors and Ergonomics Society Annual Meeting*. 1–5.

[8] John A Dewey, Elisabeth Pacherie, and Günther Knoblich. 2014. The phenomenology of controlling a moving object with another person. *Cognition* 132, 3 (2014), 383–397.

[9] Ouriel Grynszpan, Aïsha Sahaï, Nasmeh Hamidi, Elisabeth Pacherie, Bruno Berberian, Lucas Roche, and Ludovic Saint-Bauzel. 2019. The sense of agency in human-human vs human-robot joint action. *Consciousness and Cognition* 75 (2019), 102820.

[10] Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, and Sergey Levine. 2018. Soft Actor-Critic Algorithms and Applications. *CoRR* abs/1812.05905 (2018). arXiv:1812.05905 http://arxiv.org/abs/1812.05905

[11] Guy Hoffman. 2019. Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems* 49, 3 (2019), 209–218.

[12] Dimosthenis Kontogiorgos, Andre Pereira, Olle Andersson, Marco Koivisto, Elena Gonzalez Rabal, Ville Vartiainen, and Joakim Gustafson. 2019. The effects of anthropomorphism and non-verbal social behaviour in virtual assistants. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. 133–140.

[13] Danica Kragic, Joakim Gustafson, Hakan Karaoguz, Patric Jensfelt, and Robert Krug. 2018. Interactive, Collaborative Robots: Challenges and Opportunities.. In *IJCAI*. 18–25.

[14] Nicole C Krämer, Astrid von der Pütten, and Sabrina Eimler. 2012. Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. In *Human-computer interaction: The agency perspective*. Springer, 215–240.

[15] Solène Le Bars, Alexandre Devaux, Tena Nevidal, Valérian Chambon, and Elisabeth Pacherie. 2020. Agents' pivotality and reward fairness modulate sense of agency in cooperative joint action. *Cognition* 195 (2020), 104117.

[16] Minha Lee, Gale Lucas, and Jonathan Gratch. 2021. Comparing mind perception in strategic exchanges: human-agent negotiation, dictator and ultimatum games. *Journal on Multimodal User Interfaces* (2021), 1–14.

[17] Aïsha Sahaï, Elisabeth Pacherie, Ouriel Grynszpan, and Bruno Berberian. 2017. Predictive mechanisms are not involved the same way during human-human vs. human-machine interactions: a review. *Frontiers in neurorobotics* 11 (2017), 52.

[18] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. 2006. Joint action: bodies and minds moving together. *Trends in cognitive sciences* 10, 2 (2006), 70–76.

[19] Konstantinos Sfikas and Antonios Liapis. 2020. Collaborative agent gameplay in the pandemic board game. In *International Conference on the Foundations of Digital Games*. 1–11.

[20] Ali Shafti, Jonas Tjomsland, William Dudley, and A Aldo Faisal. 2020. Real-world human-robot collaborative reinforcement learning. *arXiv preprint arXiv:2003.01156* (2020).

[21] Valeria Villani, Fabio Pini, Francesco Leali, and Cristian Secchi. 2018. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics* 55 (2018), 248–266.

[22] Eva Wiese, Giorgio Metta, and Agnieszka Wykowska. 2017. Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Frontiers in psychology* 8 (2017), 1663.