Joint Human AI decision making and sense of agency

Bruno Berberian (Bruno.berberian@onera.fr)[1] & Marine Pagliari[1,2], Valerian Chambon[2]
[1] Département d'Études Cognitives, Institut Jean Nicod, ENS, CNRS, PSL University, Paris, France
[2] Information Processing and Systems, Office National d'Etudes et Recherches Aérospatiales, Salon de Provence, France

Abstract

In this study, we address shared Human-AI decision making through the lens of agency: what is the feeling of control experienced by an operator when making a decision that is shared with an AI-based artificial agent? Using an optimization trajectory task, we explored how the level of autonomy and the explainability of the decision support system impact the human operator's feelings of agency, responsibility and confidence. Results indicated that the level of automation modulate the sense of agency, but also the confidence and feeling of responsibility. More interestingly, our results also indicated that explanations could increase the number of mind change, but also the feelings of agency, confidence and responsibility during such mind change. These results are discussed in term of mechanisms underlying the development of a sense of agency and aims to provide tools and models to explore the notion of 'meaningful control' during our interactions with AI-enabled systems.

Keywords: decision making; human-computer interaction; intelligent agents; social cognition.

Introduction

Today, although often assisted by decision-making algorithms, humans remain at the center of critical decisions in many areas (medical diagnoses, judicial decisions, decisions to open fire in armed conflict, etc.). This situation could be disrupted in the near future by advances in technology, particularly those driven by progress in the field of artificial intelligence. Many researchers are now talking about Human AI Hybrid Intelligence, a concept based on collaborative decision-making between human partners and artificial agents. By combining the strengths of human intelligence and AI, these technologies will undoubtedly deliver significant performance gains. However, the ethical, fair, and safe application of AI for supporting decision making raise broader questions (Leslie, 2019).

Central among these is the notion of control, how much control human will keep on this collective decision making. The notion of control has emerged as a key element in the design of socially acceptable AI systems. Particularly, the concept of "meaningful human control" has been proposed as a key element for the design of AI-integrated sociotechnological systems, particularly as a safeguard for appropriate human responsibility (e.g., Christen, Burri, Kandul, & Vörös, 2023; Siebert, Lupetti, et al., 2023; Santoni de Sio & Van den Hoven, 2018). Yet little is currently known about how this control emerges during joint decision making, how it may be disrupted, and how systems can be designed to effectively foster its development.

In order to progress on this topic, we propose to explore the development of the sense of agency during joint Human AI decision making. The sense of agency refers to the experience of controlling one's actions and their consequences in the external world (Haggard & Chambon, 2012; Haggard, 2017). In recent years, extensive research has highlighted the role of sense of agency (SoA) in promoting responsibility and self-regulation, particularly in ethical decision-making. When individuals perceive themselves as in control of their actions and outcomes, they are more likely to take responsibility, thereby reinforcing accountability and moral engagement (Bigenwald & Chambon, 2019). A loss of agency could therefore constitute a form of moral disengagement from our actions that would disturb the mechanism classically used to regulate human behavior (Bandura, 1999). Given its central role in shaping perceptions of control and responsibility, the sense of agency offers a valuable framework for understanding the impact of AI on human experience in joint human-AI decision making.

In this context, we design an experiment which aims to better explore how a decision support algorithm impacts the sense of agency with regard to the final decision made. In particular, the experiment presented aimed to better understand how the level of autonomy of the decision support system and the degree of explanation provided by the system on the proposals made could impact the human operator's sense of agency, as well as their sense of responsibility and confidence. Furthermore, we are interested in the role of explanations in humans' change of mind.

Method

Participants

Twenty participants were recruited for this first experiment (12 women; mean age: 24.56 years, SD = 4.49 years). The participants had no neurological or psychiatric disorders. They signed a written informed consent form before the experiment and received financial compensation for their participation.

Task

The task performed by participants is a trajectory optimization task. Specifically, participants must carry out 'missions' consisting of evaluating and selecting one route from three proposals. These routes are designed to have advantages and disadvantages based on three specific criteria: the length of the journey, the number of waypoints, and the degree of danger, represented by the color of the squares on the map. In particular, the environment is composed of colored boxes representing various types of information (Figure 3) with white box (no particular danger),

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

grey box (slight danger), black box (significant danger), red box (danger proportional to the number shown on the box) and green box (recommended crossing point).

Participant has to select a route for each mission, trying to optimize the three criteria mentioned above as much as possible. It is important to note that, in reality, the three proposed routes are equivalent in terms of overall performance, with the differences lying solely in the weighting of the three criteria. These routes are calculated using an A* algorithm (Metge et al., 2021), ensuring that the route is optimized according to the weighted criteria. Once the participant had made their choice, they could receive or not receive a route suggestion from a decision support system ('autonomy' factor) and this decision could be accompanied or not by an explanation from the system ('explainability' factor). In addition, we manipulate system decision congruency (i.e., the proportion of cases where the AI's decision matches the human decision - 'congruency' factor).

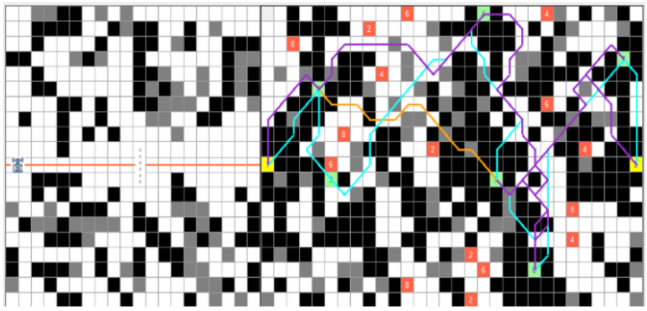

Figure 1: Representation of the environment for each mission.

Procedure

Each mission began with the presentation of three routes pre-established by the A* algorithm. Under all conditions, the first stage of the mission consisted of collecting the participant's choice from among the three proposed routes. Once this choice was recorded, the missions consisted of three distinct phases:

AI decision: for levels 2 to 5 of the AI autonomy variable, the AI's preference was given a few seconds after the participant's choice. Depending on the type of mission (explained vs. unexplained and congruent vs. incongruent), explanations corresponding to the weightings for each of the three criteria (distance travelled, number of waypoints and minimization of danger) were displayed or not, and the AI's preference corresponded to the congruence modality assigned to the mission. It is important to note here that for autonomy level 1, this decision phase is not present since the AI does not give a decision to the human operator.

Implementation of the decision: in this phase, the implementation of the route corresponded to its highlighting, with a view to its imminent validation. The author of the route implementation depended on the level of the AI autonomy variable: for level 1 and 2, implementation was carried out by the participant, and for levels 3 to 5, implementation was automatically carried out by the AI.

Validation: once the route had been implemented, the final phase involved validating the itinerary. This validation also depended on the level of autonomy of the AI, with validation carried out by the participant for levels 1 to 3 and validation carried out by the AI for levels 4 and 5.

Finally, once the route has been validated, feedback is displayed, clearly showing the participant the advantages and disadvantages for each criterion (journey length, number of waypoints and minimization of danger) in the form of gauges ranging from green (criterion fully optimized) to black (criterion not optimized at all).

Design and randomization

The experiment comprised two blocks (Explained vs Unexplained; 50 trials each). Block order was counterbalanced by participant-number parity. Within each block, trials were presented in a pseudo-random order and the congruency manipulation yielded 70% congruent and 30% incongruent trials for Level 2–Level 5. Trial order was independently randomized for each participant subject to these constraints.

Variables

We manipulated three experimental variables:

Level of AI autonomy (5 levels): For each block of 50 trials, we manipulated the level of AI autonomy so that 10 trials of each level were present, each comprising 7 congruent trials and 3 incongruent trials. The levels of AI autonomy follow the classification proposed by Parasuraman et al., (2000):

- Level 1: the participant carries out the decision, implementation and validation without system intervention.
- Level 2: the operator indicates their preference among the three routes proposed, then the system proposes one of the three routes. The participant implements and validates a choice (they may or may not follow the choice proposed by the system).
- Level 3: the participant indicates their preference among the three proposed routes, then the system proposes one of the three routes, the system implements this decision, and the participant must perform the final validation (they may or may not follow the choice proposed by the system).
- Level 4: the participant indicates their preference among the three routes proposed, then the system proposes one of the three routes, the system implements and validates this decision, but the participant can exercise a 'right of veto' over this decision, i.e. they can change this decision and validate another route.
- Level 5: The participant selects their preferred route from the three proposed routes. The system then proposes one of the three routes. The system implements and validates this decision without the participant being able to intervene.

Table 1: Levels of AI autonomy with the respective roles of participant and AI.

LoA	Decision	Implementation	Validation
1	Human	Human	Human
2	AI	Human	Human
3	AI	AI	Human
4	AI	AI	AI + Human veto
5	AI	AI	AI

Explainability (2 levels – Explained vs not explained) : The experiment consists of two blocks of 50 missions, one block consisting of explained missions and one block of unexplained missions. In the explained block, the system provides the weighting for the different route criteria for each mission, corresponding to the Feature-Importance explanation, while in the unexplained block, no explanations are provided. Participants with even participant numbers started with the explained block, and participants with odd participant numbers started with the unexplained block.

Figure 2 : Explanation illustration indicating how each the proposed route optimizes each of the dimensions, with, from top to bottom, the number of validated waypoints, the travel time and the danger associated with the route.

Congruency (2 conditions - congruent vs. incongruent): Throughout the experiment, we adjusted the rate of correspondence between the human operator's decisions and those of the AI for autonomy levels 2 to 5 (it is important to note here that for autonomy level 1, congruence was not manipulated since there is no system decision in this condition). The rate of congruence between the operator's decision and that of the AI was set at 70%. Thus, out of a total of 50 missions per block, 35 missions were 'congruent', i.e. the AI proposed the same decision as that initially chosen by the operator. However, for 15 tasks, the AI suggested a different decision, thus creating an 'incongruent' condition. It is also important to note that the three types of decisions proposed were equally effective, meaning that this manipulation did not create any performance disparities between the congruent and incongruent trials.

Metrics

At the end of each mission with the final route validated, participants received qualitative feedback in the form of a global visual summary: the three gauges indicated how well the selected route addressed each criterion (from well-taken-into-account to poorly taken-into-account). No numerical score, point-based reward, or explicit performance incentive was provided. Then, three questions were asked to the participant:

- *Sense of control*: "How much control did you feel you had over this task?" on a scale of 0% to 100%, where: 0 = no control at all – 100 = complete control.
- *Confidence* in the decision: "How confident are you that this is the right decision?" on a scale of 0% to 100% with 0: not at all confident – 100: totally confident.
- *Responsibility* for the decision: "How responsible do you feel for causing this outcome?" on a scale of 0 to 100 with 0: Not at all responsible – 100: Totally responsible.

We were also interested in behavioral metric, namely the *proportion of mind changes*. It corresponds to the proportion of trials where the choice validated at the end of the task was different from the one initially chosen by the participant. In this paper, we focus on the impact of explainability on this change of mind, as well as the impact of this change of mind on sense of control, confidence and responsibility.

The data were processed using R software and the lme4 package, using GLM for which we included the fixed effects of the independent variables and their interactions for each model, and added the maximum random structure (slope + intercept) per subject.

Results

In this section, we report results on the impact of autonomy, congruence and explainability on our three variables, namely sense of control, confidence and responsibility. We also present the impact of explainability on the number of changes of mind, as well as the link between change of mind, explainability and feeling of responsibility. We will limit the report to the main effect for each variable.

Level of autonomy

The analysis showed that the level of autonomy had significant effect on sense of agency ($F(4, 2961) = 171.25$, $p < .0001$), confidence ($F(4, 2961) = 23.45$, $p < 0.0001$) and feeling of responsibility ($F(4, 2961) = 117.63$, $p < .0001$) (see figure 3). Post hoc analyses showed that the level of autonomy reduced the sense of agency and the feeling of responsibility. For confidence, we observe a bell-shaped relationship, with an initial increase in confidence with the level of autonomy, followed by a decrease in confidence for the highest level of autonomy.

Congruency

The analysis showed that the congruency (Congruent Vs Incongruent) had significant effect on sense of agency ($F(1, 2353) = 271.79$, $p < .0001$), confidence ($F(1, 2353) = 286.37$, $p < .0001$) and feeling of responsibility ($F(1, 2353) = 591.05$, $p < .0001$) (see figure 4). Post hoc analyses showed that the sense of agency, the feeling of responsibility and the confidence decreased in incongruent condition for each level of autonomy.

Explainability

The analysis showed that the explainability (Explained vs Unexplained) had significant effect on sense of agency ($F(1,2353) = 5.8404$; $p = 0.0157$), but no effect significant effect on the confidence ($F(1, 2353) = 0.72$, $p = 0.3950$) and feeling of responsibility ($F(1, 2353) = 1.89$; $p = 0.1693$). Post hoc analyses showed that the sense of agency, increased in presence of explanation, but only for the LoA 4.

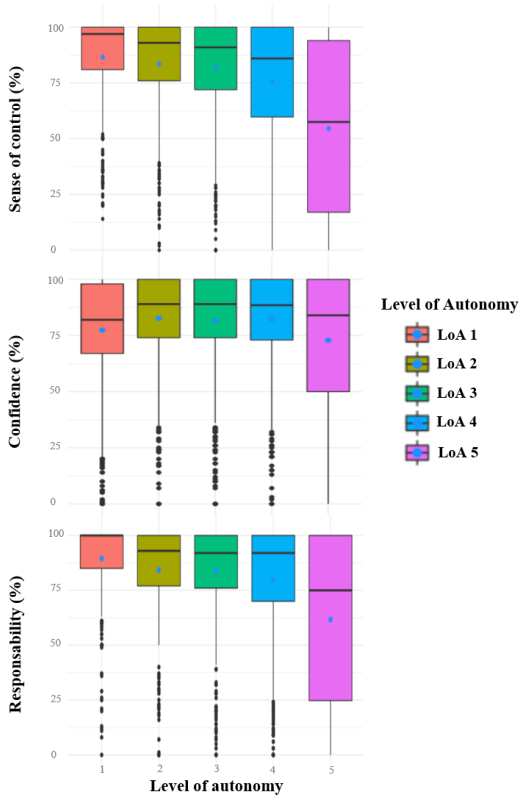

Figure 3: Representation of the effect of AI autonomy levels on sense of control, confidence and responsibility. The horizontal bar represents the median and the blue dot represents the mean for each level shown. The error bars represent the range of data up to 1.5 * IQR (where IQR is the interquartile range).

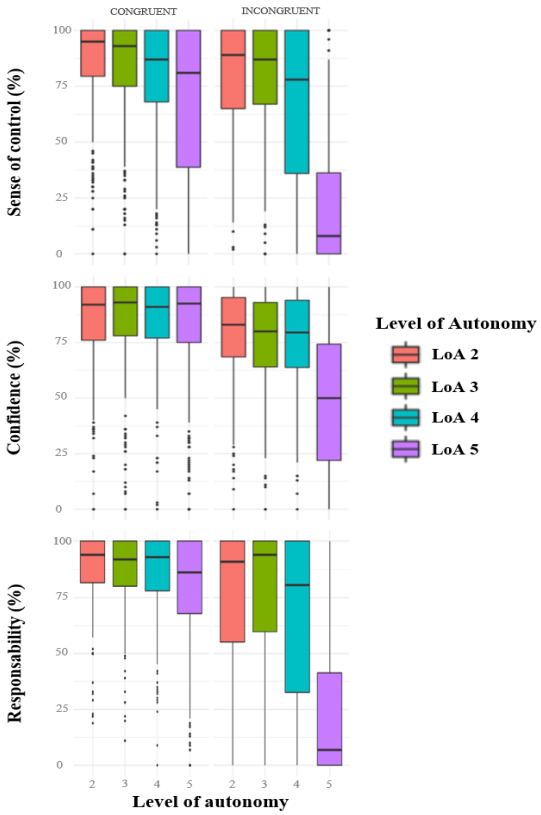

Figure 4: Representation of the effect of congruency for each level of AI autonomy levels on sense of control, confidence and responsibility. The horizontal bar represents the median. The error bars represent the range of data up to 1.5 * IQR (where IQR is the interquartile range).

Explainability and Change of mind

We first performed a Wilcoxon test to explore the impact of explainability on the rate of change of mind (regardless of the level of automation). The analysis showed that the explainability had significant effect on the rate of change of minds, with an increase of the change of minds in presence of explanations ($W = 72414$, p-value < 0.000001) (see Figure 5).

Finally, considering the impact of explainability and minds change on our three variables (i.e., sense of control, confidence and responsibility), the analysis showed that mind change had a significant effect on sense of agency ($F(1, 2962) = 87.53$, $p < .0001$), confidence $F(1, 2962) = 46.62$, $p < .0001$) and feeling of responsibility ($F(1, 2962) = 103.69$, $p < .0001$). Post hoc analyses showed that sense of agency, confidence and feeling of responsibility decreased in missions where participants change their minds compared to missions where they do not change their minds. We also observed a significant interaction between explanation and minds change with respectively $F(1, 2962) = 10.06$, $p = 0.0015$ for sense of agency, $F(1, 2962) = 8.2536$, $p = 0.0041$ for confidence and $F(1, 2962) = 6.5004$, $p = 0.0108$ for responsibility. Post-hoc analyses showed that providing

explanations leads to an increase in feelings of agency, confidence, and responsibility when changing one's mind, whereas this effect is not present when participants do not change their minds.

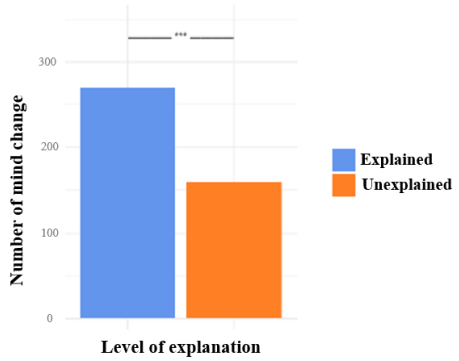

Figure 5: Cumulative observations of changes of mind regarding explainability factor.

Discussion

The present study investigated the impact of explanation and level of autonomy on feeling of agency, confidence and responsibility during shared decision-making. Our results showed that the level of automation modulate the sense of agency, but also the confidence and feeling of responsibility. In addition, incongruent proposal from the AI could dramatically decreases the feelings of agency, confidence and responsibility. More interestingly, our results also indicated that explanations could increase the number of mind change, but also the feelings of agency, confidence and responsibility during such mind change. We will now discuss these different results in turn.

"Optimal" level of automation during joint decision making

A fundamental question concerns the 'optimal' level of autonomy to be implemented in tasks involving AI algorithms. Indeed, in a society where the legal issue of liability for AI decisions has not been resolved, it remains to be determined which tasks could legitimately be automated and, conversely, which tasks would be better delegated exclusively to human operators. The results obtained during our work provide food for thought on this subject.

First, our results suggest that the level of autonomy of an AI significantly influences participants' sense of agency: the greater the level of autonomy of the AI, the less participants feel. In other words, the more autonomous the AI is, the less the user feels they have direct control over the actions or decisions taken and the less responsible they feel for the choices made. These results replicate those obtained by Berberian et al. (2012), which already showed a decrease in sense of agency proportional to the increase in a system's autonomy. At the extreme, the highest level of autonomy, which implies the absence of any possibility of veto for human participants, massively reduces the sense of agency. This result indicates that the absence of a right of veto over the choice of artificial partner drastically alters the development of the sense of agency, thus highlighting the importance of keeping the operator in the decision-making loop. These results echo the work carried out by Caspar and colleagues on the impact of coercion on the sense of agency (Caspar, Christensen, Cleeremans, & Haggard, 2016; Caspar, Beyer, Cleeremans, & Haggard, 2021; Caspar, Cleeremans, Haggard, 2018; Caspar, Vuillaume, Magalhães De Saldanha da Gama, & Cleeremans, 2017; Caspar, Ioumpa, Keysers, & Gazzola, 2020), but also to studies showing that the role of the operator (leader or follower) has a drastic impact on the development of the sense of agency (Chaminade, & Decety, 2002; Pfister, Obhi, Rieger, & Wenke, 2014; Reddish, P, Tong, Jong, & Whitehouse, 2020). In addition, we observe that the decline in the sense of agency with increasing levels of automation is accompanied by a decline in the sense of responsibility for the results obtained, providing further evidence of a relationship between the sense of agency with regard to an action and the sense of responsibility for the consequences of that action (Caspar et al., 2016). Such decrease in feelings of agency and responsibility should be carefully considered as a stronger sense of agency is known to enhance motivation and engagement (Bandura, 1997; Wen et al., 2015) and affect operational efficiency (Pagliari et al., in press).

In contrast, the relation between level of autonomy and confidence seems quite different: when the level of autonomy shifts from a manual task to the system taking over certain functions (selection and implementation of the decision), operators' confidence levels increase. On the other hand, it decreases again when the task is performed under full supervision. This dynamic is consistent with the findings of Ueda et al. (2021), which suggests that the effect of automation on the sense of agency is not linear: while the assistance provided by automation increases operator performance, and therefore the sense of agency, for low to medium levels of automation, the effect becomes detrimental to operators' sense of agency when the level of automation exceeds 90%, even if their performance continues to increase. The reason for this decrease in the sense of agency – despite high performance – is that at these levels of automation, participants are no longer able to attribute the system's (good) performance to themselves (Ueda et al., 2021). While confidence increased at intermediate levels of autonomy, confidence is only beneficial when it tracks performance. Because our task was designed to minimize objective performance differences between options, we interpret confidence here primarily as a subjective metacognitive signal of decision comfort under different autonomy regimes, rather than as evidence that intermediate autonomy improves decision accuracy. Future work should explicitly couple these judgments with tasks where performance varies meaningfully across options.

Taken together, these results reflect the paradoxical effects of automation on human behaviour: automation can lead to an effective loss of control for the human operator, while improving their performance. They thus highlight the importance of finding a compromise between the assistance provided by the decision support system and the participant's involvement in the control loop.

However, our paradigm imposed the level of autonomy rather than allowing participants to choose when and how to rely on the system. In many real-world settings, users can actively calibrate reliance (e.g., requesting advice only under uncertainty). An important next step is therefore to test whether the same agency–responsibility trade-offs arise when autonomy is user-selected, and whether explainability changes the thresholds at which people choose to delegate.

Explanation, change of mind and sense of agency

Beyond this level of operator contribution, the opacity of the systems seems to be an obstacle to the development of their sense of agency. We therefore investigated the impact of explanations regarding the proposals made by the system on participants' sense of control, confidence and responsibility, as well as on their propensity to accept the proposals made by the system.

Our findings reveal no significant main effect of the presence of explanations on our three main variables (i.e., sense of agency, confidence and responsibility). This initial result is not particularly surprising, given that our variables are themselves strongly influenced by the level of automation and by the congruence between the choice proposed (or imposed) by the system and the participant's initial choice. Thus, providing an explanation when what is proposed by the system is already what has been chosen by the participant has no real added value.

On the other hand, we observe a clear interaction between explanation and congruence. First, our results show an increase in the proportion of changes of mind for the 'explained' vs. 'unexplained' parts. This result highlights the importance of explainability on participants' choice behaviour. Indeed, the observation of a different proportion of changes of mind between 'explained' and 'unexplained' tasks suggests that when a decision or action is accompanied by a clear justification, participants may be more inclined to reconsider their initial opinions or choices. A relevant explanation provides context or clarifies a choice situation, enabling the user to better understand the underlying decision-making process. By contrast, the absence of explanation maintains uncertainty, and participants who do not understand the reasons behind the decision are more likely to stick to their initial position (out of caution or lack of trust in the system). This observation could potentially suggest an increase in the acceptability of the system when it provides explanations. More interestingly, providing explanations not only leads to an increase in the number of changes of mind, but also to an increase in the sense of agency, confidence in one's decision, and the responsibility felt when changing one's mind. Indeed, changing one's mind leads to a change in the attribution of responsibility: participants attribute more responsibility to AI for the decisions they have changed, and attribute more responsibility to themselves for the decisions that have not changed. However, we also observe a significant interaction between explainability and change of mind, such that explainability significantly mitigates this decrease in the sense of responsibility when changing one's mind. In other words, while a change of mind leads to a decrease in self-attribution of responsibility, 'explanation' diminishes this effect and induces a reattribution of responsibility to oneself.

Taken together, these results seem to indicate that explanations impact the way in which humans can consider and appropriate decisions proposed by assistance systems. At a more operational level, explainability would therefore make it easier to accept the choices made by a decision support algorithm. Explanation could thus be necessary in order to engage in a form of 'collective' decision-making. Furthermore, given the links between the experience of control and commitment, we can consider this to be a way of increasing operators' commitment to the choices made, even if these choices are guided by assistance.

Limitations

A limitation is that we did not collect trial-level ratings of perceived difficulty or perceived competence. Because agency and confidence are sensitive to the perception of the performance, future work should include manipulation checks (e.g., 'How difficult was this trial?') to better separate the effects of autonomy/explainability from variation in perceived task difficulty. Finally, our conclusions are grounded in explicit, self-reported judgments (control/confidence/responsibility). These reflective judgments can be influenced by contextual inference (e.g., how participants interpret the AI's role), and may not fully capture lower-level sensorimotor aspects of agency. Future work combining explicit ratings with implicit measures would help clarify which components of agency are most affected by autonomy and explainability.

Conclusion

Through this experiment, we focused on the development of a sense of agency when making joint decisions with an artificial partner. We examined how the level of autonomy and opacity of the artificial partner could impact this development, while also questioning the link between a sense of agency, confidence, responsibility and the acceptability of the partner's proposals. From a theoretical perspective, this work provides a better understanding of the mechanisms underlying the development of a sense of agency and its links to feelings of responsibility and confidence. From an applied perspective, it provides the community with tools and models to explore the notion of 'meaningful control' during our interactions with AI-enabled systems, as well as recommendations in terms of autonomy and explainability.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

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

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.

Bigenwald, A., & Chambon, V. (2019). Criminal Responsibility and Neuroscience: No Revolution Yet. *Frontiers in Psychology, 10*, 1406.

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.

Caspar, E. A., Beyer, F., Cleeremans, A., & Haggard, P. (2021). The obedient mind and the volitional brain: A neural basis for preserved sense of agency and sense of responsibility under coercion. *PLOS ONE, 16*(10), e0258884.

Caspar, E. A., Cleeremans, A., & Haggard, P. (2018). Only giving orders? An experimental study of the sense of agency when giving or receiving commands. *PLOS ONE, 13*(9), e0204027.

Caspar, E. A., Vuillaume, L., Magalhães De Saldanha Da Gama, P. A., & Cleeremans, A. (2017). The Influence of (Dis)belief in Free Will on Immoral Behavior. *Frontiers in Psychology, 8*.

Caspar, E. A., Ioumpa, K., Keysers, C., & Gazzola, V. (2020). Obeying orders reduces vicarious brain activation towards victims' pain. *NeuroImage, 222*, 117251.

Chaminade, T., & Decety, J. (2002). Leader or follower? Involvement of the inferior parietal lobule in agency. *Neuroreport, 13*(15), 1975-1978.

Christen, M., Burri, T., Kandul, S., & Vörös, P. (2023). Who is controlling whom? Reframing "meaningful human control" of AI systems in security. *Ethics and information technology, 25*(1), 10.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience, 18*(4), 196-207.

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology, 22*(10), R390–R392.

Leslie, D. (2019). Understanding artificial intelligence ethics and safety. arXiv preprint arXiv:1906.05684.

Metge, A., Maille, N., & Le Blanc, B. (2021, March). Operators and autonomous intelligent agents: human individual characteristics shape the team's efficiency. In *AICA 2021*.

Pagliari, M., Haggard, P., Berberian, B. & Chambon, V. (in press). The Dynamics of Agency in Human–AI Teams: From Follower Roles to Effective Control. *Philosophical Transactions B.*

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.

Pfister, R., Obhi, S. S., Rieger, M., & Wenke, D. (2014). Action and perception in social contexts: Intentional binding for social action effects. *Frontiers in Human Neuroscience, 8*.

Reddish, P., Tong, E. M., Jong, J., & Whitehouse, H. (2020). Interpersonal synchrony affects performers' sense of agency. *Self and Identity, 19*(4), 389-411.

Santoni De Sio, F., & Van Den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI, 5*, 15.

Cavalcante Siebert, L., Lupetti, M. L., Aizenberg, E., Beckers, N., Zgonnikov, A., Veluwenkamp, H., ... & Lagendijk, R. L. (2023). Meaningful human control: actionable properties for AI system development. *AI and Ethics, 3*(1), 241-255.

Ueda, S., Nakashima, R., & Kumada, T. (2021). Influence of levels of automation on the sense of agency during continuous action. *Scientific Reports, 14*.

Wen, W., Yamashita, A., & Asama, H. (2015). The influence of goals on sense of control. *Consciousness and Cognition, 37,* 83–90.