



HAL
open science

How humans comply with a (potentially) faulty robot: Effects of multidimensional transparency

Loïck Simon, Clément Guérin, Philippe Rauffet, Christine Chauvin, Éric
Martin

► **To cite this version:**

Loïck Simon, Clément Guérin, Philippe Rauffet, Christine Chauvin, Éric Martin. How humans comply with a (potentially) faulty robot: Effects of multidimensional transparency. *IEEE Transactions on Human-Machine Systems*, 2023, pp.1-10. 10.1109/THMS.2023.3273773 . hal-04089023

HAL Id: hal-04089023

<https://hal.science/hal-04089023v1>

Submitted on 5 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

How humans comply with a (potentially) faulty robot: Effects of multidimensional transparency

Loïck Simon, Clément Guérin, Philippe Rauffet, Christine Chauvin and Éric Martin

Abstract—This paper deals with how a human operator follows a request made by a robot in an industrial context. This robot is potentially myopic, i.e., its request could be based on partial and limited information. Therefore, it could possibly be faulty. This exploratory study aims to analyze whether multidimensional agent transparency may have an effect on two drivers of compliance identified in the literature: trust in signal and risk perception. In this experiment, we manipulated different agent transparency levels combined with two dimensions of agent transparency: robot-TO-human transparency (rTOh) and robot-OF-human transparency (rOFh). Results mainly show that adding rOFh to rTOh transparency changes human compliance with the agent and moderates both trust and risk perception. Moreover, task performance and completion time are shown to vary according to the different transparency levels. Our results show that transparency has no effect on mental workload. From a methodological perspective, this study shows the importance of distinguishing the different types of information about which a robot can be transparent, especially the combination of rTOh and rOFh transparencies. From a practical point of view, the study shows that the agent transparency framework needs to be considered carefully when designing human-robot teaming in the context of Industry 4.0 in the case of robots that may be unreliable due to their myopia.

Index Terms— Agent transparency, Compliance, Trust, Risk perception, Industry 4.0.

I. INTRODUCTION

Industry 4.0 is characterized by new technologies (e.g., digital twin, cobotics, virtual and augmented realities) to optimize, customize, and make production reconfigurable [1][2]. These technological innovations make it possible to control increasingly complex industrial processes that are increasingly using autonomous agents and intelligent cyber-physical systems. If working with these new agents brings new capabilities in Industry 4.0, they can also generate new interferences. Research in the field of Intelligent Manufacturing Systems shows that the development of autonomous agents can disturb cooperative activities with humans, especially when considering the phenomenon of myopia [3]. Myopia represents the lack of information of an autonomous agent regarding the scope (between a local and a global level) and the time horizon

(between near and distant future) of the supervised process and the other entities of the system [4]. This myopia can be used as a compromise between reactivity and optimality of the decisions taken by these autonomous agents. Thus, it may bring about human operators' perception of a faulty autonomous agent's behavior.

Several studies have focused on the effects of faulty robots in collaborative situations with humans, in particular the impact on the trust granted to the autonomous agent. These effects seem to differ depending on the nature of the associated risks, i.e., depending on consequences that may be more or less important or damaging. Rossi *et al.* [5] and Wang *et al.* [6] investigate the effects of different degrees of error consequences on trust. They show that trust decreases if the robot makes mistakes associated with serious consequences. Salem *et al.* [7] explain that trust can also be influenced when the consequences of the task are revocable. Correia *et al.* [8] also observe a decrease in the trust given to a faulty robot, but justifying the technical failure could mitigate the negative impacts on trust when the consequences of the failure are not significant. When these consequences are significant or serious, i.e., when the failure is likely to compromise the task, the justification would have no effect on trust. Daronnat *et al.* [9] investigate the effect of different types of agents' errors (mistakes, lapses, slips, and violation) on performance and trust. They conclude that while autonomous agents are likely to be imperfect, a complete lack of information input is better than inaccurate information.

Designing autonomous agent transparency is an issue for successful human-robot interaction [6][10]. Transparency is a key factor that influences humans' willingness to interact with an autonomous agent such as a robot. Transparency helps develop the capacity of robots to inspire trust and to be considered as legitimate teammates. In this article, we use the term "agent" to refer to an autonomous agent (in our study, a collaborative robot, i.e., a cobot), and we focused on how this agent may provide information to human operators (hereinafter referred to as "operators") depending on different transparency conditions. Our general objective was to investigate the effects

This paper was submitted on 14/09/2022. The research presented in this paper is carried out within the framework of the HUMANISM N° ANR-17-CE10-0009 research program, funded by the French National Research Agency (ANR).
Loïck Simon - Lab-STICC - CNRS UMR 6285, University of South Brittany, France. loick.simon@univ-ubs.fr (corresponding author)
Clément Guérin - Lab-STICC - CNRS UMR 6285, University of South Brittany, France. clement.guerin@univ-ubs.fr

Philippe Rauffet - Lab-STICC - CNRS UMR 6285, University of South Brittany, France. philippe.rauffet@univ-ubs.fr
Christine Chauvin - Lab-STICC - CNRS UMR 6285, University of South Brittany, France. christine.chauvin@univ-ubs.fr
Eric Martin - Lab-STICC - CNRS UMR 6285, University of South Brittany, France. eric.martin@univ-ubs.fr

of these conditions, defined along several dimensions of transparency, on compliance with an autonomous agent that has partial observability (or myopia) in an industrial context, i.e. an order preparation line.

II. RELATED WORKS

A. Compliance with an autonomous agent: Trust and perception of risk

In several studies, compliance has been operationalized as the human decision following a request or robot's recommendation (e.g., [7][6]). In Chen *et al.* [11], compliance is measured in terms of the agreement rate with the decision support system. This compliance is known to be affected by trust as it is a behavioral expression of attitude (i.e., trust) [12]. For Freedy *et al.* [13], trust has an effect on accepting suggestions from agents (i.e., being compliant). Compliance is also seen, in an objective way, as an observable behavioral indicator of trust [11]. Trust is especially compromised when the behavior of the agent is faulty, as trust is positively correlated to the perception of reliability [14].

Moreover, for Lee and See [15] and Chancey *et al.* [16], the perception of risk also needs to be considered to understand how human operators comply with an autonomous agent, given that it is not unusual to use prior work on human-automation interaction in order to investigate human-autonomy teaming [17]. Chancey *et al.* [16] maintain that the uncertainty linked to risk plays a role in mediating trust. For Mayer, Davis and Schoorman cited by Stuck *et al.* [18], “*An individual actually taking a risk is defined as exhibiting behavioral trust*” (p. 4). In other words, risk perception is seen as a cognitive process that moderates trust and bears upon human compliance with robots, i.e., the decision to follow the advice provided by the robot.

This relationship between compliance, trust, and risk perception is synthesized in Fig. 1, based on the work of Lee and See [15] and Chancey *et al.* [16]. The two main factors explaining a compliant behavior are (1) **Trust**, i.e., how humans perceive whether a system is faulty or not when it detects a new situation and emits a signal and whether they can rely on the agent and (2) **Risk perception**, i.e., how humans perceive the consequences of the signal emitted by the agent on the task and on other aspects.

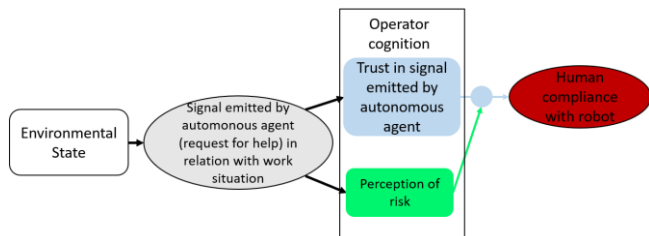


Fig. 1. Model of human–autonomous agent trust on compliance behavior (adapted from Chancey *et al.* [16])

Finally, these combined effects of trust and risk perception upon compliance are also mentioned by Hancock *et al.* [19]. The authors explain that the appropriate calibration of

compliance (i.e., to avoid over- or under-compliance) is dependent on how humans cognitively process two types of information: robot or automation characteristics (e.g., in terms of reliability and rationale) and environmental characteristics (e.g., in terms of assessment of the work situation). For these authors, the robot's characteristics play a huge role in human compliance, and they especially cite among these characteristics the notion of agent transparency.

B. Agent transparency: Definitions and framework

Roundtree *et al.* [20] identify several definitions of transparency in the literature. Transparency has been described as a process, a method, a mechanism, a property, or an emergent characteristic. By means of transparency, autonomous agents provide information about their abilities or capabilities to a human operator. In this paper, we adopt Chen *et al.*'s [21] definition of agent transparency as “*the descriptive quality of an interface pertaining to its abilities to afford an operator's comprehension about an intelligent agent's intent, performance, future plans, and reasoning process*” (p.2). The terms “agent transparency” and “transparency” are interchangeable.

As suggested in the two literature reviews on transparency proposed by Bhaskara *et al.* [22] and Rajabiyazdi and Jamieson [23], two main approaches have been adopted in the literature to describe and operationalize the concept of transparency: Chen *et al.*'s [21] situation awareness–based agent transparency, known as the SAT model, and Lyons' framework [24] of transparency for human-robot interaction. In their review, Rajabiyazdi and Jamieson [23] specify that Lyons' framework has been cited 119 times and Chen *et al.*'s model 91 times. In these two approaches, transparency is unidirectional: an autonomous agent is the sender, and the human operator is the recipient.

The SAT model proposed by Chen *et al.* [21] is based on the Situation Awareness theory of Endsley [25], on the Beliefs, Desires, Intentions (BDI) agent framework of Rao and Georgeff [26], and on the Process, Purpose, and Performance (PPP) framework related to trust of Lee and See [15]. The SAT model incorporates three levels of agent transparency. At level 1 (L1), the agent provides basic information about its current state, goals, intentions, plans, progress, and current and proposed actions. At level 2 (L2), the agent provides a rationale that justifies its action or decision. At this second level, the human operator is given information about the agent's reasoning, behavioral capabilities, and the constraints it considers. At level 3 (L3), the agent provides a projection of future outcomes (e.g., success rate). At this third level, the human operator is given information regarding the agent's anticipation of the future state, predicted consequences, and uncertainties.

Lyons' [24] model involves two dimensions of agent transparency:

- The robot-TO-human transparency (rTOh) concerns information about the robot that is communicated to the human operator. The agent may be transparent about its intent or purpose (the intentional model), the current task

or the previous tasks conducted (the task model), the processes performed that led to a decision or an action (the analytical model), or aspects of the environment (the environmental model).

- The robot-OF-human transparency (rOFh) concerns the robot’s awareness of “the others”, which is communicated to the human operator. The agent may be transparent about its perception of the operator’s state (the operator model) or the task allocation (the teamwork model) that is responsible for a task or set of tasks.

Chen *et al.* [21] have attempted to integrate the Lyons’s teamwork model into the original SAT model, in order to better support the interaction between the agent and the human operator. This extension of the SAT model that includes teammates is called the Dynamic SAT model [27].

C. The impact of multidimensional transparency on operators regarding compliance

Previous studies have examined the impact of agent transparency levels upon trust, risk perception, mental workload, and response time. These studies were applied mainly to the military domain [28][29][30][31] and transportation systems (monitoring and supervision [29][30][31], an aircraft landing aid [32], or driving an autonomous car [33][34]). These studies, however, do not deal with multiple dimensions of transparency and their interaction.

Impact on trust. Trust has been examined as an effect of agent transparency level. In their study of landing assistance systems for pilots, Lyons *et al.* [32] observe that trust was higher when transparency was at a higher level, i.e., when operators had access to the logic of success calculations. Similarly, Selkowitz *et al.* [29] indicate that the more transparent the autonomous agent was, the higher the level of trust in the agent was. Chen *et al.* [21] state that transparency concerning unreliable elements tends to lower the operator’s trust level in the system. The authors show that when the system reveals its uncertainties at L3 (e.g., when it indicates that the required information is missing), the operator’s trust in the system diminishes. Salem *et al.* [7] find that the revocability of an action has an effect on compliance. However, Wang *et al.*, [6] find no correlation between transparency and compliance. In this paper, we view compliance as an agreement rate with the agent’s request.

Impact on risk perception. We did not find studies seeking to measure a link between risk perception and transparency, although a few studies use a high/low risk situation in their experiment (e.g., [35]). Although Chancey *et al.* [16] find a non-significant effect of risk on trust, their results are unclear. In this study, using measures of risk perception enables us to understand the effect of risk perception upon compliance.

Impact on mental workload. Agent transparency provides operators with greater or lesser access to various types of information. A high level of agent transparency may produce an increased amount of information causing information overload, which affects the operators’ mental workload directly and may cause interface usability issues [36]. However, results concerning the link between transparency and mental workload

remain inconsistent [37] although they tend to show there is no detrimental impact on mental workload [10].

Impact on completion time. Findings differ from one study to another. Stowers *et al.* [28] report that the decision time is significantly longer in L3 than in L2, whereas Mercado *et al.* [30] and Wright *et al.* [31] find no differences between levels. Some authors have used a secondary task in addition to the decision task [31]. These authors do not find any difference in terms of secondary task performance or overall time to complete both tasks.

D. Research question and hypotheses

In this article, we seek to better understand how agent transparency may impact on two human cognitive processes, i.e., trust and risk perception, in a collaborative situation with robots. For Roundtree *et al.* [20], compliance is an indirect transparency factor to assess trust in human-machine interaction. Moreover, as emphasized in related works, agent transparency can be described according to two main dimensions (rTOh and rOFh) in Lyons’ framework [24] and at different levels (Chen *et al.*’s SAT framework [21]). In this paper, we propose to combine the frameworks of Lyons [24] and Chen *et al.* [21]. Our research question is as follows: how may the combined transparency levels of rOFh and rTOh impact the operators’ compliance, the trust they have in the autonomous agent, their risk perception, their mental workload, and their performance? We proposed five main hypotheses (Fig. 2):

Hypothesis 1 (H1a): Transparency concerning agent limitations in rTOh will decrease human compliance [20]

(H1b): Transparency concerning agent limitations in rTOh will decrease trust [20]

Hypothesis 2 (H2a): Transparency concerning situational risk in rOFh will affect human compliance [16][20].

(H2b): Transparency concerning situational risk in rOFh will affect perceived risk [16].

Hypothesis 3 (H3): rOFh transparency concerning a risk will moderate the effect of rTOh transparency and decrease compliance. [16].

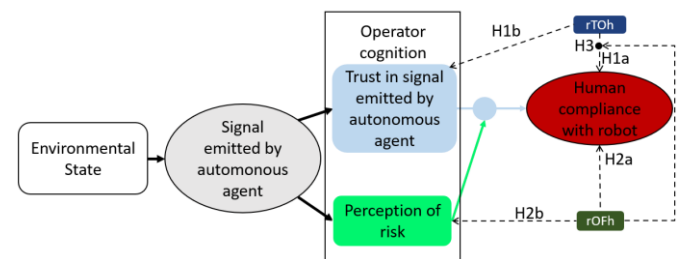


Fig. 2. Hypothesized effects of multidimensional transparency on trust in signal and risk perception upon compliance

III. METHOD

In order to investigate the research question, we set up a scenario in which a collaborative industrial robot (referred to as cobot) was an autonomous agent. This cobot emitted a signal

towards the human operator in the form of a request.

A. Participants

A total of 53 engineering students in the field of mechatronics or industry participated in our experiment. The gender distribution included 17 women and 36 men, the average age was 21.34 years ($SD = 1.67$ years), and no participant had vision problems affecting color perception (e.g., color blindness or dyschromatopsia). The recruiting campaign for the participants took place over 3 weeks through e-mails, posters, and oral communication at the beginning of class. The students were invited to participate in a one-hour study on the topic of human-cobot cooperation through a software package simulating an order preparation line. The students received a 10-euro voucher for their participation. All the participants signed a free and informed consent form containing the objective of the experiment, their right to withdraw from the experiment at any time, their right to ask to suppress their data, information on the use of their data, and the measures that would be used in the experiment.

B. Scenario

In the experimental scenario, participants had to supervise a simulated order preparation line composed of one conveyor belt and three preparation stations (see Fig. 3). In this context, the participants were decision-makers on resource sharing between two production workstations in which a cobot and a human teammate work together (they each fill a box with colored resources to complete a production order). Regularly, the cobot requested that the participants transfer resources from the human teammate's stock to its stock because it experienced a stock-out of a critical resource; in our study this occurred only on the "red cube" resource (Fig. 3). The cobot's request was based on the analysis of its own local stock quantity and the stock in the warehouse (i.e., whether there was enough of this type of resource in the warehouse) (underlined in Fig. 3). In some cases, however, the information about the stock in the warehouse might not be processed by the cobot due to potential myopia. Hence, the cobot's request could be faulty.

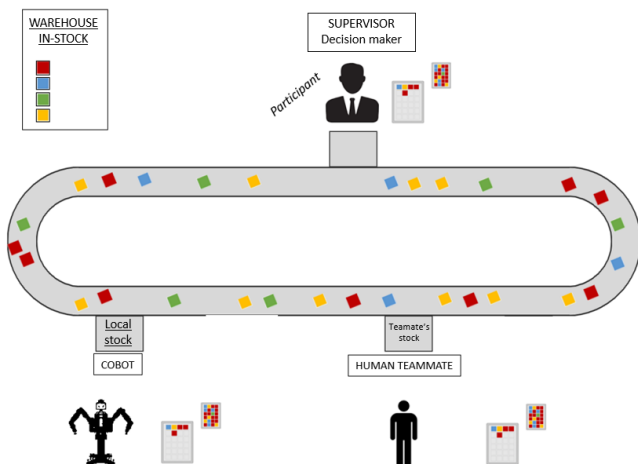


Fig. 3. Experimental context and role of the agent and human teammate

For example, the cobot could request help from the participant even if there was enough stock in the warehouse. Participants were invited to decide whether they would comply or not with the cobot's request, i.e., pick resources from the human teammate, which would imply interfering with the teammate's stock.

To answer this decision task, participants were placed into two different modalities of the rTOh transparency level:

- In the first modality (referred to as rTOh L2), the cobot provides a request and its rationale (i.e., it sends a possible stock-out alert), but it is not transparent about its myopia (i.e., it does not indicate whether it has checked the warehouse stock or not before making its request). Hence, participants could think that the cobot has checked the warehouse stock.
- In the second modality (referred to as rTOh L3), the same information is provided, but the cobot specifies that it has not taken the warehouse stock into account when making its request. The cobot is transparent on its myopia. Hence, participants were certain that the cobot had not checked the warehouse stock.

Moreover, to study the effect of adding rOFh transparency to rTOh transparency, we reproduced the previous two conditions by displaying information related to the capabilities of the human teammate (i.e., whether he is busier than the cobot on his production orders). The cobot did not use this information to make its request and to estimate the resources available on the preparation line but communicated it to the operator. This additional information is not related to the decision task but may be considered by participants as cobot interference on the human teammate's activity; the cobot's request could be correct, but the outcomes of the cobot's action could have a negative effect on the human teammate, due to a transfer of resources. This information is related to the perception of risk. In this experiment, this rOFh transparency shows the relative capacity factor of the teammate, indicating whether the human teammate is busier (value superior to 100%) or less busy (value inferior to 100%) than the cobot. The two modalities of rTOh transparency were combined with three modalities of rOFh transparency:

- a modality with no rOFh transparency (referred to as rOFh L0)
- a modality with positive feedback on the teammate's capacity factor (referred as to rOFh L1+). The cobot perceives that the teammate is less busy with his production order than the cobot itself is.
- a modality with negative feedback on the teammate's capacity factor (referred as to rOFh L1-). The cobot perceives that the teammate is busier with his production order than the cobot itself is.

Combining these levels of transparency with the two dimensions generated six experimental conditions (Table I). Each participant was confronted with the six experimental conditions twice; hence, each participant had to make twelve decisions. Situations were randomized to prevent a participant

from having the same two “rTOh” transparency situation in a row and to counterbalance a possible order effect. Each decision was independent from the others since it had no impact on the unfolding of the scenario. As participants were placed in short-term situations, they could not set up projections regarding the existing situation.

TABLE I
THE TRANSPARENCY LEVELS IN THE SIX EXPERIMENTAL CONDITIONS

Experimental Conditions	Transparency Level rTOh	Transparency Level rOFh
1	L2	L0
2		L1+
3		L1-
4	L3	L0
5		L1+
6		L1-

C. Material

To operationalize these experimental conditions, a software package was developed specifically for the experiment. This software showed an interface presenting the information communicated by the cobot to the supervisor (see Fig. 4). The cobot’s request about the resource transfer randomly popped up during the first 10 seconds of each session lasting 90 seconds. Participants could comply with the request by clicking on “accept”, and the cobot took resources in the teammate’s stock. Participants could reject the request by clicking on either “refuse” (i.e., they consider that the human teammate is too busy for the moment and the cobot cannot take resources) or “alternative” (i.e., the cobot should check the warehouse stock quantity) (Fig. 4 - top half). Therefore, accepting the request could possibly solve the cobot’s problem at the cost of potential negative interference on the teammate’s stock, and refusing the request did not help solve the cobot’s problem but permitted it to avoid negative interference. After making their decision, participants did not have access to the consequences of their choice.

The information provided by the cobot to the supervisor (Fig. 4 - bottom half) could differ depending upon the agent’s transparency levels. Fig. 4 shows the different display options concerning the warehouse in-stock indicator (relating to the rTOh dimension) and the relative capacity factor indicator (relating to the rOFh dimension). The black square represents the agent’s lack of transparency about the information gathering for reasoning for the rTOh dimension and its myopia (if rTOh = L2) and about the resources of the human teammate for the rOFh dimension (if rOFh = L0). In addition to this decision task, participants had to complete a secondary task. This secondary task involved filling in a grid by reproducing a visual pattern of cubes (Fig. 4 - top half) within a set time frame of 90 seconds.

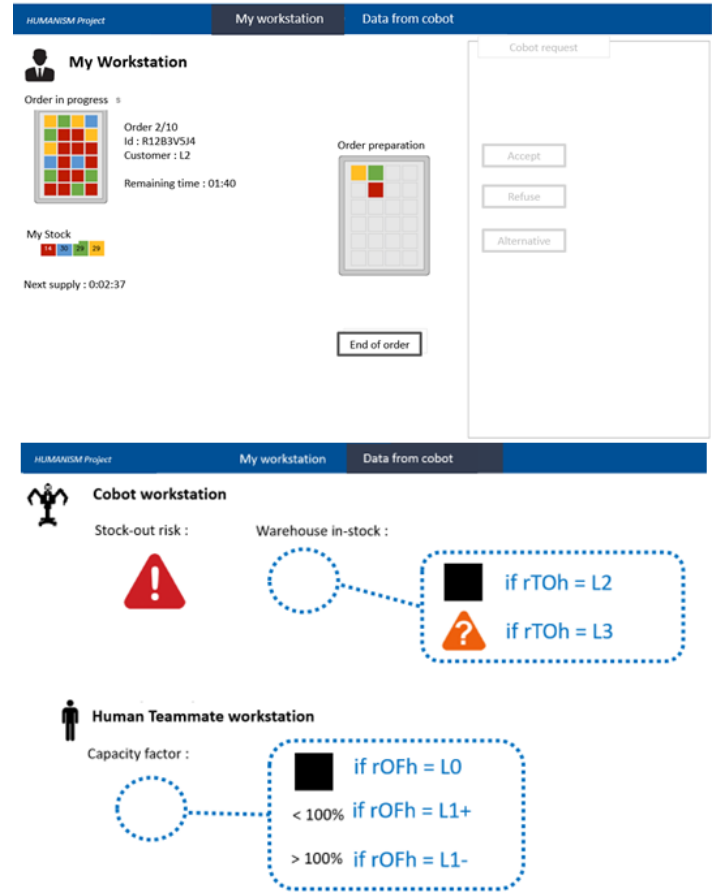


Fig. 4. The interface showing the filling task and the cobot’s request (top half) and the indicator display in line with transparency levels (bottom half). The stock-out alert is shown as a red exclamation mark.

D. Participants’ guidance.

Participants were first welcomed, and they signed the consent agreement. Next, the industrial context and the experimental scenario were presented using visual aids, to ensure they understood what the cobot was communicating. Participants were told that they were in charge of supervising the production line. Then, the rTOh transparency levels were presented, as described in section III.B. Especially for rTOh L3, participants were told that the cobot communicates about its myopia (i.e., about its potential error and incomplete information gathering), and that it is transparent about not having checked the warehouse in-stock. This possible limitation could result in an inaccurate request for help. The participants could not know for certain that the robot was faulty because they did not have access to the warehouse in-stock. Finally, the cobot’s rOFh transparency was explained. Participants were told that this information was not used by the cobot to make its request, but it was additional information that the robot communicates. This information could help assess the interference of the cobot’s transfer from the human teammate’s stock:

- For the rOFh L1+ level, participants were told that the

conditions were favorable for a transfer of resources.

- For the rOFh L1- level, participants were told that the conditions may interfere with the work of the human teammate.
- For the rOFh L0 level, participants were told that they did not have access to this information.

Before the experiment, we used a short questionnaire to control two possible and non-desired characteristics of this panel: the participants' familiarity with games such as Poker or Blackjack, since their decision was made with missing information, and a risk-taking propensity. These control variables were tested and showed no significant effect on the data collected. The participants' tendency to be trustful of technologies was also controlled through a 2-item scale but showed no significant effect on the data collected.

E. Data collection and processing

Collected data. The software package enabled the simulation of an industrial context and automatically collected the data after the participants had made each of the 12 decisions (see Table II). Compliance was coded as follows: **Comply**, when participants decide to "accept" the request from the robot; **Do not comply**, when participants use the "refuse" or "alternative" button. Trust was assessed by a questionnaire adapted from Jian *et al.* [38] and Yagoda and Gillian [39].

TABLE II
DATA COLLECTED IN THE EXPERIMENT (R STANDS FOR REVERSE ITEM)

Construct/Dependent Variable	Collected data	Scale/tool
Human compliance with robot	Accept, Alternative or Refuse	Accept was coded as compliance Alternative or Refuse was coded as non-compliance
Trust in the robot	Trust 1 : « I found the cobot understandable » Trust 2 : « I found the cobot to be unreliable » (R) Trust 3 : « I trust the cobot »	For the 3 items : 7-point Likert scale from « strongly disagree » to « strongly agree »
Perception of risk	Did you perceive a charge rate of the second agent higher than that of the cobot	4 possible answers : « Yes », « Very unlikely », « Unlikely », « No » Compared to the situation, the answer « Yes » or « Non » was scored as 2 if it was correct, or -2 if it was incorrect. Similarly, the answers « very likely » and « unlikely » were scored as 1 if correct and -1 if incorrect
Subjective mental workload	« How do you feel ? »	5-point Likert scale ranging from « I am bored » (1) to « I am overloaded » (5)
Performance	Completion time	/

Risk perception was assessed by a single item created for this experiment in order to analyze how participants perceived the possible interference of the cobot's request with the human teammate's activity. Due to a limited amount of time, the Instantaneous Self-Assessment (ISA) technique [40] was chosen to assess the cognitive workload of participants.

Statistical methods. We used R (R Core Team, 2012), and especially the lme4 package [41] ordinal logistic regression (function polr to study the effects on ordinal variables like Trust or ISA), and linear mixed effects analyses (function lmer to study the effects on continuous variables, such as completion time, task completion). Regarding the linear effects analysis, visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. For all the statistical models, we entered rTOh transparency and the rOFh

transparency as fixed effects (with interaction terms) into the full model. As a random effect, we had an intercept for participants. Regarding fixed effects, a stepwise model selection by AIC (stepAIC) was conducted. During each step, a new model was fitted, in which one of the terms of the model was eliminated and tested against the former model.

F. Experimental plan and exploratory questions

To answer the research question (Section II.D) and in view of the experimental protocol presented above, we adopted a mixed factorial design, with rTOh transparency (L2 or L3) and rOFh transparency (L0, L1+ or L1-) as within-subject factors. Fig. 5 represents the integration of experimental conditions and measures on the theoretical model.

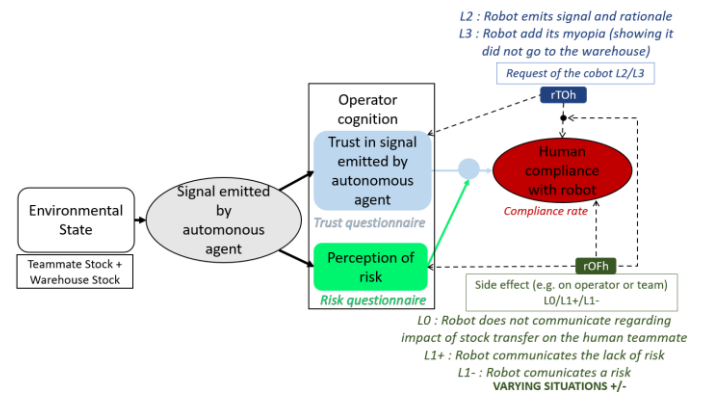


Fig. 5. Integration of the experimental conditions and measures on the theoretical model

IV. RESULTS

A. Effect of transparency upon compliance with the cobot

We first investigated the effect of rTOh and rOFh transparency levels upon human compliance with the cobot's requests. We conducted a logistic regression that reveals main effects of rTOh and rOFh.

TABLE III
LOGISTIC REGRESSION: EFFECT OF RTOH AND ROFH TRANSPARENCY UPON HUMAN COMPLIANCE WITH COBOT'S REQUEST

Characteristic	OR ¹	95% CI ¹	p-value
rTOh (ref L2)			
L3	0.14	-2.75, -1.24	<.001
rOFh (ref L1-)			
L0	8.47	1.46, 2.82	<.001
L1+	26.34	2.30, 4.25	<.001
rTOh * rOFh (ref L2 * L1-)			
L3 * L0	2.5	-0.11, 1.94	.08
L3 * L1+	7.29	0.51, 3.46	.008

¹OR = Odd Ratio, CI = Confidence Interval

Table III shows that compliance is significantly lower when rTOh transparency is at L3 when compared to L2 (OR = 0.14, $p < .001$). This decrease in compliance corresponds to when the agent shows that the request is based on partial observation of

the situation (rTOh L3) and confirms hypothesis H1(a).

We also observed that compliance was significantly lower when rOFh transparency indicated negative feedback, compared to the other levels of rOFh transparency. In effect, rOFh L1-, indicating possible important interference with the human teammate, resulted in lower compliance than rOFh L1+, indicating minimal interference with the human teammate (OR = 26.34, $p < .001$). Similarly, compliance was lower with rOFh L1- than with rOFh L0 (with no information on the human teammate) (OR = 8.47, $p < .001$). This result supports hypothesis H2(a).

A cobot that is not transparent about its myopia and that communicates negative feedback (rTOh L2*rOFh L1-) induces less compliance than a robot that is transparent about its myopia but does not communicate negative feedback (OR = 7.29, $p < .01$ for rTOh L3*rOFh L1+, and OR = 2.5, $p = .08$, for rTOh L3*rOFh L0). These effects support hypothesis H3.

These effects are shown in Fig. 6, where human compliance is always lower in rTOh L3 than in rTOh L2, regardless of the associated rOFh transparency level. We also observe that rOFh transparency has an important effect on the variations of compliance and moderates the effect of rTOh transparency. When no information is presented on the impact on the human teammate (rOFh L0), participants are predicted to be likely noncompliant with rTOh L2 (34%), and totally noncompliant with rTOh L3 when indicating a partial observability of the agent for its request (12%). However, when adding information related to the other operator, these tendencies to comply or not comply change slightly. With a positive situation indicating minimal impact on the other operator (rOFh L1+), participants are predicted to be rather compliant (86% with rTOh L2, 62% with rTOh L3). On the contrary, participants become totally noncompliant with the cobot (6% with rTOh L2, 1% with rTOh L3) when it is shown that the impact on the human teammate could be serious (rOFh L1-).

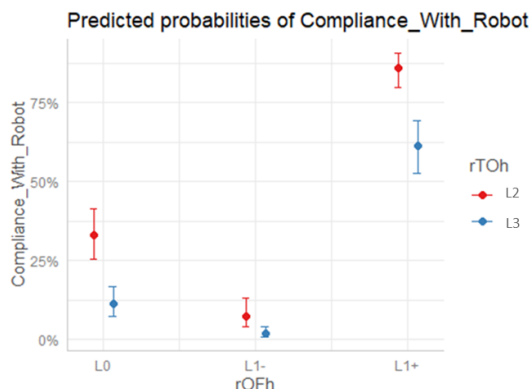


Fig. 6. Prediction of compliance with the cobot as a function of rOFh and rTOh transparency levels (error bars correspond to a 95% confidence interval, representing the uncertainty of the estimation)

B. Transparency on trust in robot and perception of risk

As asserted in the introduction, the two dimensions of

transparency are assumed to modify the two main factors of human compliance with the robot's request. rTOh transparency is supposed to play upon trust, whereas rOFh transparency may change risk perception. We sought to verify these relationships by studying the subjective responses from participants on the different questionnaires related to these two factors. With an ordinal logistic regression, we analyzed the effect of transparency on the additional question presented to participants, designed to estimate the level of risk perception (Table IV).

TABLE IV
ORDINAL LOGISTIC REGRESSION: RISK PERCEPTION

Characteristic	OR ¹	95% CI ¹	p-value
rTOh (ref L2)			
L3	0.79	0.51, 1.22	0.3
rOFh (ref L1-)			
L1+	1.73	1.12, 2.69	0.014

¹OR = Odds Ratio, CI = Class interval

We observed a significant effect of rOFh transparency. The risk was perceived as higher and more certain by participants in rOFh L1- than in rOFh L1+ conditions (OR = 1.73, $p = .014$), which confirms H2(b) (Table IV).

When conducting ordinal logistic regressions on the different items of our trust questionnaire, we also observed a significant effect of the rOFh transparency. Participants found the agent more understandable (OR = 1.57, $p = .009$), more reliable (OR = 1.66, $p = .004$), and more trustworthy (OR = 1.70, $p = .002$) when rOFh transparency is at level L1+ (i.e., with a minimal impact on the other production operator) compared to L0. Conversely, we did not observe any significant effect of the rTOh levels or the effect of their interaction (see Table V). Hence, results do not support H1(b).

TABLE V
ORDINAL LOGISTIC REGRESSION: EFFECT OF TRANSPARENCY
ON TRUST IN THE ROBOT

Characteristic	OR ¹	95% CI ¹	p-value
Understandable			
rTOh (ref L2)			
L3	0.86	0.65, 1.13	0.3
rOFh (ref L0)			
L1 -	1.30	0.92, 1.82	0.13
L1+	1.57	1.12, 2.21	0.009
Reliable			
rTOh (ref L2)			
L3	0.93	0.71, 1.23	0.6
rOFh (ref L0)			
L1 -	1.38	0.98, 1.93	0.064
L1+	1.66	1.18, 2.33	0.004
Trustworthy			
rTOh (ref L2)			
L3	0.82	0.62, 1.09	0.2
rOFh (ref L0)			
L1 -	1.19	0.85, 1.68	0.3
L1+	1.70	1.21, 2.40	0.002

¹OR = Odds Ratio, CI = Class interval

C. Relationships between compliance, trust, and perceived risk

We also examined whether compliance can be related to trust and perceived risks. Mann-Whitney analyses revealed that compliance and trust are significantly associated in the different questions on trust. Participants complying with the robot's request reported higher scores on the robot's understandability ($W=50846$, $p = .008$), reliability ($W=51383$, $p = .004$), and trustworthiness ($W=55450$, $p < .001$), in comparison with participants who did not comply. Similarly, there was also a significant association between compliance and risk perception. Participants complying with the robot reported the perception of a lower risk ($W=23378$, $p=.038$) than those not complying with the robot.

D. Transparency on dual task performance and mental workload

Completion time and dual task performance. Table VI shows a significant positive effect of rOFh L1+ on the time used to make a decision; the time needed is reduced by -3.92 seconds. For rTOh L3, the total time significantly increased (+ 5.34 seconds) in particular for the time needed between the analysis of indicators and the decision: + 4.27 seconds. There is a significant reduction of the performance obtained on the filling rate for the operational secondary task for rTOh L3.

TABLE VI
MIXED LINEAR MODEL OF PERFORMANCE (ON THE
SECONDARY TASK AND TIME TO MAKE A DECISION)

	Time between indicators analysis and decision	Total Time ~ rTOh	Order Completion ~ rTOh
Intercept	15.22 (1.22)***	52.33 (1.11)***	99.90 (0.89)***
rTOh (reference = L2)			
L3	4.27 (0.90)***	5.34 (0.82)***	-1.70 (0.71)*
rOFh (reference = L0)			
L1+	-3.92 (1.10)***		
L1-	-2.00 (1.10)		
AIC	5593.71	5716.97	4679.76
BIC	5621.06	5735.36	4697.57
Log Likelihood	-2790.85	-2854.49	-2335.88

(*** = $p < .001$, ** = $p < .01$, * = $p < .05$)

For rTOh L2, operators have a 99.9% filling rate, which is superior to the 98.2% filling rate observed when it is at L3. Conversely, we do not observe any significant effect of the interaction between rTOh and rOFh

Subjective mental workload. Variations in rTOh transparency and rOFh transparency did not show any significant effect on mental workload.

V. DISCUSSION

This study offers theoretical, methodological, and practical contributions. Those contributions help provide greater understanding of the effects of trust and risk perception upon the compliance of human operators with collaborative machines. This study also shows the possibility of manipulating two drivers of compliance with different levels and different dimensions of agent transparency.

From a theoretical perspective, the results show that human compliance with a myopic agent, reasoning with partial and limited information on the available resources, significantly changed according to how transparent the robot was regarding its myopia. Thus, a higher robot-TO-human transparency (rTOh L3) was linked with lower compliance with the robot's request than when the myopia was hidden to the operator (rTOh L2). This result is also corroborated by a lower performance, in terms of completion time or secondary task performance (production orders completion), when participants work with a robot using a rTOh L3 transparency than when collaborating with a robot with a rTOh L2 transparency. This result is contrary to recent analyses [10] but can be explained by the nature of the information transmitted by the rTOh L3 transparency (i.e., the robot shows faulty reasoning).

Moreover, compliance drastically varied when robot-OF-human transparency was added. The transparency regarding the positive situation of the other operator (as in condition rOFh L1+ corresponding to a minimal impact on this other operator) increased compliance in comparison with the no transparency condition (rOFh L0), whereas the transparency of a negative situation (rOFh L1-) decreased compliance in comparison with rOFh L0. The interaction between the rTOh and rOFh dimensions is significant when a risky situation is reported by the robot (rOFh L1-). It is congruent with Chancey et al.'s model as risk that moderates trust. This effect of robot-OF-human transparency on human compliance could be explained by how trust in the robot and risk perception are related to this rOFh transparency. We observed that a robot transparent about a positive situation (rOFh L1+) significantly led to a lower perception of the risk by the human operator than when the transparency is related to a negative situation (rOFh L1-). Transparency about the positive situation of the other operator also resulted in a higher trust of participants in the robot's reliability than when no information was shown on the other operator (rOFh L0).

These findings were therefore congruent with the model of Chancey *et al.* [16] (Fig.1) assuming that risk perception can be viewed as a moderator of human trust in the robot's reliability, and the moderating effect impacts human compliance with a robot. These authors were not successful in supporting this assumption (“*It is unclear from our results whether risk moderated the mediating effect of trust on compliance*”, p. 342). Our results brought stronger evidence, by showing that adding information (with rOFh transparency) about the risk of interfering with another operator can moderate the relation of trust and compliance with the collaborative machines. Conversely, our findings were inconsistent with Wang *et al.* [6], who do not find a correlation between transparency and compliance. This inconsistency can be explained by the type of transparency examined in the different studies. In Wang *et al.* [6], robot transparency concerned only the rationales and the explanation of the robot's decision. In our experiment, the transparency was more about the non-revocability (as suggested by Salem *et al.* [7]) and the consequences of the robot's decision on the activity of the human operator. However, the rTOh did

not have an impact on trust. We believe that the reason for this is that compliance (i.e., objective trust) is more context sensitive than subjective trust. Nonetheless, we found a relationship between the objective measure of trust and its subjective measure, which means that the behavioral measure is related to the attitudinal measure.

From a methodological perspective, the results of the study confirm that it is useful to distinguish the different types of information about which an autonomous agent can be transparent [10]. Information transparency must be analyzed not only in terms of levels, as proposed in the initial SAT model of Chen *et al.* [21], but also by considering the different transparency dimensions suggested by Lyons [24]. Some dimensions could have a more significant role in the quality of the decision making and could have greater influence than the other dimensions. This is why this paper analyzes the effect of transparency along two main dimensions, robot-TO-human and robot-OF-human transparency, to better understand the effect of transparency on trust in and compliance with an agent when combining these two dimensions. We believe it is necessary to consider these different dimensions separately, in particular when relating them to the operators' cognitive processes, as this may explain how trust and risk perception are formed in given situations.

From a practical perspective, this paper deals with the case of robots that may be unreliable in the context of Industry 4.0. This situation may commonly be found in industry, due to the robot's myopia or some technical problem of access to the information needed to correctly analyze a situation. Our study has explored this issue by highlighting how trust can be formed with a myopic agent, according to which rationale, and which elements of the context the agent communicates to the human operator.

In addition, the effect of expertise in the decision-making processes needs to be explored because the participants in our study were novices as they lacked experience in order preparation. Through the paradigm expert/novice, it would be worth examining populations familiar with and expert in industrial situations (e.g., production line operators in industry) [10].

Finally, this work opens interesting perspectives on the topic of Human-Autonomy Teaming and in the domain of Industry 4.0. The key issues concern designing the new interactions between the operators and cyber-physical systems of Industry 4.0 in order to improve the reliability of the decision-making process and avoid errors. Parasuraman and Riley [42] have shown that overtrust in and overcompliance with autonomous agents do not necessarily improve performance. Operators may not challenge the agents' choices and choose to blindly follow their propositions, which could lead to poor decision making, especially if an agent is not reliable. In the cooperation configurations found in Industry 4.0 where human operators supervise autonomous agents, two different behaviors may be observed: either the operators conduct an in-depth analysis of the situation, thereby increasing their activity level, or they carry out a superficial analysis and merely accept the

proposition made by the machine. The latter behavior is the complacency phenomenon [42]. Following some recent studies [43][44], we could imagine dynamically calibrating or repairing the trust of humans in autonomous agents, in a bi-directional transparency perspective between human and agent [45]. This dynamic trust calibration could be based on compliance, i.e., how humans provide feedback to the robot by accepting or refusing its requests. We would then play upon the robot transparency level, regarding not only the information related to the rationale and the analysis of the situation by the robot, but also the complementary information related to the perception of risk, that can interfere with trust in the agent. We could also vary the robot's etiquette and the way it presents the different types of information [46]. Such a dynamic process would help avoid inappropriate compliance with the robot and improve human-autonomy teaming.

VI. CONCLUSION

This study has focused upon human decision making following the request of a cobot. The effects of agent transparency have been examined through the transparency levels of Chen *et al.*'s model [21] and applied to transparency dimensions using Lyons's framework [24]. We have observed the effects upon human operators' compliance, trust in the cobot, and risk perception. In addition, congruent with the results of prior studies [10], we have not observed any increase in the participants' mental workload despite the increased transparency level. We have also found that risk perception can moderate trust. Further research might want to use a risk perception scale in order to have better comprehension of the role of risk perception in trust and compliance.

To promote the emergence of an Operator 4.0 in symbiosis with cyber-physical systems, optimizing human-machine cooperation needs to rely upon a high-quality dialogue. Therefore, it would be worth exploring in more detail the effect of other agent transparency levels applied to these dimensions upon the quality of the dialogue between human operators and autonomous agents. Greater understanding of these mechanisms would improve the dialogue and optimize cooperation between the two entities through the adaptability of the human-machine interfaces. Greater comprehension of transparency and its effects on trust and risk perception could be a way to dynamically calibrate trust in order to avoid overtrust or undertrust.

ACKNOWLEDGMENT

The authors wish to acknowledge the work of the engineer Iehann EVENO for the development of the software used in this experiment.

REFERENCES

- [1] E. Hoždić, « Smart factory for industry 4.0: A review », *International Journal of Modern Manufacturing Technologies*, vol. 7, no 1, p. 28-35, 2015.
- [2] F. Longo, L. Nicoletti, and A. Padovano, « Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context », *Computers & industrial engineering*, vol. 113, p. 144-159, 2017.

- [3] W. Derigent, O. Cardin, and D. Trentesaux, « Industry 4.0: contributions of holonic manufacturing control architectures and future challenges », *J Intell Manuf*, vol. 32, no 7, p. 1797-1818, oct. 2021
- [4] M. Adam, S. Gregor, A. Hevner, and S. Morana, « Design Science Research Modes in Human-Computer Interaction Projects », *AIS Transactions on Human-Computer Interaction*, vol. 13, no 1, p. 1-11, mars 2021
- [5] A. Rossi, K. Dautenhahn, K. L. Koay, and M. L. Walters, « How the Timing and Magnitude of Robot Errors Influence Peoples' Trust of Robots in an Emergency Scenario », in *Social Robotics*, vol. 10652, A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssele, and H. He, Éd. Cham: Springer International Publishing, 2017, p. 42-52. doi: 10.1007/978-3-319-70022-9_5.
- [6] N. Wang, D. V. Pynadath, and S. G. Hill, « Trust calibration within a human-robot team: Comparing automatically generated explanations », in 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), mars 2016, p. 109-116. doi: 10.1109/HRI.2016.7451741.
- [7] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, « Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust », in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, Portland Oregon USA, mars 2015, p. 141-148. doi: 10.1145/2696454.2696497.
- [8] F. Correia, C. Guerra, S. Mascarenhas, F. S. Melo, and A. Paiva, « Exploring the Impact of Fault Justification in Human-Robot Trust », p. 8, 2018.
- [9] S. Daronnat, L. Azzopardi, and M. Halvey, « Impact of Agents' Errors on Performance, Reliance and Trust in Human-Agent Collaboration », *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 64, no 1, p. 405-409, déc. 2020, doi: 10.1177/1071181320641091.
- [10] van de Merwe, K., Mallam, S., & Nazir, S. (2022). Agent Transparency, Situation Awareness, Mental Workload, and Operator Performance: A Systematic Literature Review. *Human Factors*, 00187208221077804. <https://doi.org/10.1177/00187208221077804>
- [11] J. Chen, S. Mishler, and B. Hu, « Automation Error Type and Methods of Communicating Automation Reliability Affect Trust and Performance: An Empirical Study in the Cyber Domain », *IEEE Transactions on Human-Machine Systems*, 2021.
- [12] T. Law and M. Scheutz, « Trust: Recent concepts and evaluations in human-robot interaction », *Trust in human-robot interaction*, p. 27-57, 2021.
- [13] A. Freedy, E. DeVisser, G. Weltman, and N. Coeyman, « Measurement of trust in human-robot collaboration », in 2007 International symposium on collaborative technologies and systems, 2007, p. 106-114.
- [14] J. M. Ross, *Moderators of trust and reliance across multiple decision aids*. university of Central Florida, 2008.
- [15] J. D. Lee and K. A. See, « Trust in Automation: Designing for Appropriate Reliance », *Hum Factors*, vol. 46, no 1, p. 50-80, mars 2004,
- [16] E. Chancey, J. Bliss, Y. Yamani, and H. Handley, « Trust and the Compliance-Reliance Paradigm: The Effects of Risk, Error Bias, and Reliability on Trust and Dependence », *Human Factors The Journal of the Human Factors and Ergonomics Society*, vol. 59, p. 333-345, apr. 2017
- [17] McNeese, N. J., Demir, M., Cooke, N. J., and Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors*, 60(2), 262-273
- [18] R. E. Stuck, B. J. Tomlinson, and B. N. Walker, « The importance of incorporating risk into human-automation trust », *Theoretical Issues in Ergonomics Science*, vol. 0, no 0, p. 1-17, oct. 2021
- [19] P. A. Hancock, D. R. Billings, K. Schaefer, J. Chen, E. D. Visser, and R. Parasuraman, « A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction », *Hum. Factors*, 2011, doi: 10.1177/0018720811417254.
- [20] K. A. Roundtree, M. A. Goodrich, and J. A. Adams, « Transparency: Transitioning From Human-Machine Systems to Human-Swarm Systems », *Journal of Cognitive Engineering and Decision Making*, vol. 13, no 3, p. 171-195, sept. 2019, doi: 10.1177/1555343419842776.
- [21] J. Y. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia, and M. Barnes, « Situation Awareness-Based Agent Transparency », *ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE, ARL-TR-6905*, apr. 2014.
- [22] A. Bhaskara, M. Skinner, and S. Loft, « Agent Transparency: A Review of Current Theory and Evidence », *IEEE Transactions on Human-Machine Systems*, vol. 50, no 3, p. 215-224, june 2020
- [23] F. Rajabiyazdi and G. A. Jamieson, « A Review of Transparency (seeing-into) Models », in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), oct. 2020, p. 302-308
- [24] J. B. Lyons, « Being Transparent about Transparency: A Model for Human-Robot Interaction », presented at 2013 AAAI Spring Symposium Series, mars 2013.
- [25] M. R. Endsley, « Toward a Theory of Situation Awareness in Dynamic Systems », *Hum Factors*, vol. 37, no 1, p. 32-64, mars 1995
- [26] A. S. Rao and M. P. Georgeff, « BDI agents: From theory to practice. », in *Icmas*, 1995, vol. 95, p. 312-319.
- [27] J. Y. C. Chen, S. G. Lakhmani, K. Stowers, A. R. Selkowitz, J. L. Wright, and M. Barnes, « Situation awareness-based agent transparency and human-autonomy teaming effectiveness », *Theoretical Issues in Ergonomics Science*, vol. 19, no 3, p. 259-282, may 2018, doi: 10.1080/1463922X.2017.1315750.
- [28] K. Stowers, N. Kasdaglis, M. Rupp, J. Chen, D. Barber, and M. Barnes, « Insights into Human-Agent Teaming: Intelligent Agent Transparency and Uncertainty », in *Advances in Human Factors in Robots and Unmanned Systems*, Cham, 2017, p. 149-160. doi: 10.1007/978-3-319-41959-6_13.
- [29] A. R. Selkowitz, S. G. Lakhmani, C. N. Larios, and J. Y. C. Chen, « Agent Transparency and the Autonomous Squad Member », *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no 1, p. 1319-1323, sept. 2016, doi: 10.1177/1541931213601305.
- [30] J. E. Mercado, M. A. Rupp, J. Y. C. Chen, M. J. Barnes, D. Barber, and K. Procci, « Intelligent Agent Transparency in Human-Agent Teaming for Multi-UxV Management », *Hum Factors*, vol. 58, no 3, p. 401-415, may 2016
- [31] J. L. Wright, J. Y. C. Chen, M. J. Barnes, and P. A. Hancock, « The Effect of Agent Reasoning Transparency on Complacent Behavior: An Analysis of Eye Movements and Response Performance », *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no 1, p. 1594-1598, sept. 2017, doi: 10.1177/1541931213601762.
- [32] J. B. Lyons, K. S. Koltai, N. T. Ho, W. B. Johnson, D. E. Smith, and R. J. Shively, « Engineering trust in complex automated systems », *Ergonomics in Design*, vol. 24, no 1, p. 13-17, 2016
- [33] T. Helldin, « Transparency for future semi-automated systems », PhD diss., Orebro University., 2014.
- [34] A. Kunze, S. J. Summerskill, R. Marshall, and A. J. Filtness, « Automation transparency: implications of uncertainty communication for human-automation interaction and interfaces », *Ergonomics*, vol. 62, no 3, p. 345-360, mars 2019, doi: 10.1080/00140139.2018.1547842.
- [35] S. Loft and al., « The Impact of Transparency and Decision Risk on Human-Automation Teaming Outcomes », *Hum Factors*, p. 00187208211033445, 2021.
- [36] J. Nielsen, « Usability inspection methods », in *Conference companion on Human factors in computing systems*, 1994, p. 413-414.
- [37] J. Y. C. Chen, « Agent Transparency », in *Smart and Intelligent Systems*, CRC Press, 2021.
- [38] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, « Foundations for an Empirically Determined Scale of Trust in Automated Systems », *International Journal of Cognitive Ergonomics*, vol. 4, no 1, p. 53-71, mars 2000
- [39] R. E. Yagoda and D. J. Gillan, « You want me to trust a ROBOT? The development of a human-robot interaction trust scale », *International Journal of Social Robotics*, vol. 4, no 3, p. 235-248, 2012.
- [40] A. J. Tattersall and P. S. Foord, « An experimental evaluation of instantaneous self-assessment as a measure of workload », *Ergonomics*, vol. 39, no 5, p. 740-748, may 1996, doi: 10.1080/00140139608964495.
- [41] D. Bates, M. Maechler, and B. Bolker, « lme4: Linear mixed-effects models using S4 classes [Software] », Retrieved from CRAN. R-project. org/package=lme4, 2012.
- [42] R. Parasuraman and V. Riley, « Humans and Automation: Use, Misuse, Disuse, Abuse », *Hum Factors*, vol. 39, no 2, p. 230-253, june 1997
- [43] K. Okamura and S. Yamada, « Adaptive trust calibration for human-AI collaboration », *PLOS ONE*, vol. 15, no 2, p. e0229132, febr. 2020
- [44] K. Akash, G. McMahon, T. Reid, and N. Jain, *Human Trust-based Feedback Control: Dynamically varying automation transparency to optimize human-machine interactions*. 2020.
- [45] Holder, E., Huang, L., Chiou, E., Joon, M., & Lyons, J.B. (2021). Designing for bi-directional transparency in Human-AI-Robot-Teaming. *Proceedings of the Human Factors and Ergonomics Society's 2021 International Annual Meeting*.
- [46] J. von Meyer, C. A. Miller, P. A. Hancock, E. de Visser, and M. C. Dorneich, « Politeness in Machine-Human and Human-Human Interaction », 2016.