

Do We Blame it on the Machine?

Task Outcome and Agency Attribution in Human-Technology Collaboration

Haiyan Jia
Lehigh University
haiyan.jia@lehigh.edu

Mu Wu
California State University, Los Angeles
mwu33@calstatela.edu

S. Shyam Sundar
Penn State University
sss12@psu.edu

Abstract

With the growing functionality and capability of technology in human-technology interaction, humans are no longer the only autonomous entity. Automated machines increasingly play the role of agentic teammates, and through this process, human agency and machine agency are constructed and negotiated. Previous research on “Computers are Social Actors (CASA)” and self-serving bias suggest that humans might attribute more technology agency and less human agency when the interaction outcome is undesirable, and vice versa. We conducted an experiment to test this proposition by manipulating task outcome of a game co-played by a user and a smartphone app, and found partially contradictory results. Further, user characteristics, sociability in particular, moderated the effect of task outcome on agency attribution, and affected user experience and behavioral intention. Such findings suggest a complex mechanism of agency attribution in human-technology collaboration, which has important implications for emerging socio-ethical and socio-technical concerns surrounding intelligent technology.

1. Introduction

When humans routinely work and communicate with technology to complete tasks and solve problems, they constantly assess, consciously or non-consciously, the roles that they themselves and the technology play in the interaction. Given the well-known tendencies for fundamental attribution error in teamwork and the application of social rules to technology [1], the existing literature in communication and social psychology would suggest a self-serving bias among human users when they collaborate with technologies on tasks: When things go well, the user will take the credit; but when things go badly, they will blame the technology while distancing themselves from the failure.

Whether these propositions and prior findings hold true in the current scenario of human-machine collaboration, where machines become increasingly intelligent and assume new roles, is the focus of this study. Instead of carrying out simple tasks, intelligent agents get involved in more complex and important processes, such as problem solving and decision making, serving more as teammates than tools. As technology starts to play a more autonomous role, it is important to examine how it may affect humans' perceptions and assignment of responsibility for the decisions that have been made or the outcome of actions. This study explores two research questions:

- 1) Is the fundamental attribution error applicable in human-technology collaboration, when technology plays the role of a teammate and exhibits a certain level of intelligence or agency?
- 2) What factors influence users' attribution of agency, as well as user experience, perceptions and behaviors, in different task outcome scenarios?

The findings will provide insights into attribution biases and individual differences that impact collaboration with and acceptance of technology as teammates.

2. Literature review

The role of technology has gradually progressed from functional to assistive, and in the projection of becoming collaborative and autonomous. Co-evolving with this shift is the research spanning technology, process, people, legalization, etc. Beyond the technical aspects, ethical and moral challenges appear more pressing than ever. A recent study [2] administered a survey to 65 researchers around the world to develop a research agenda to explore the key research questions surrounding the theme of “Machines as Teammates.” In this list, *responsibility and liability* emerged as a primary research direction. Potential problems include: “Who is responsible, and who is to blame?”, “To what extent machines are responsible?”,

and “Will a human always be in charge and responsible for the result, or can this responsibility be transferred to a machine?”

As the technology evolves from tools to teammates, interaction with technology moves to “the establishment of human relationships with autonomous entities” [3]. Therefore, these research questions need to be addressed not only with technical or engineering solutions, but with an emphasis on human factors. Related research on human-computer interaction and human-robot interaction has shown that user expectations of the technology, and their perceptions of their relationship with the technology, are fundamental to shaping the interaction process, and are powerful enough to influence both their willingness to team up and interaction outcomes even without manipulating any actual collaborative interaction (e.g., [4], [5], [6]). Therefore, it is crucial to understand the psychological mechanisms of responsibility allocation in order to answer behavioral, technical and legal questions surrounding responsibility and liability issues in human-technology collaboration. In this study, we draw upon theory and literature on attribution bias and agency, as well as research on individual differences that impact human-technology interaction to explore how individuals respond to different outcome scenarios and attribute blame to themselves versus a technology teammate in a collaborative game.

2.1. Attribution bias

Attribution refers to the perception and understanding of the cause of events and of responsibility for the outcomes [7]. There are two types of attribution [8]: internal attribution (assigning the cause to certain internal factors such as ability and effort) and external attribution (assigning the cause to situational factors such as luck and environment). Previous studies have found that humans tend to “bask in reflected glory (BIRGing)”, which suggests that individuals tend to claim their association with a successful event or party [9]; in contrast, “cutting off reflected failure (CORFing)” suggests an opposite tendency of disassociation with a failed entity [10]. This phenomenon has been observed in multiple contexts, such as sports [9] and video games [11]. One underlying mechanism is that humans have a self-enhancement tendency to maintain their self-esteem [12], “taking personal responsibility for their desirable outcomes yet externalizing responsibility for their undesirable outcomes” ([13], p. 895). Propositions indicate that self-serving bias is especially pertinent under certain conditions, such as when an individual’s behavior is observed by others; when an individual has

a free choice to act; or when high ego involvement or high self-awareness is present [14].

The perception of responsibility is closely associated with agency: an individual decides to act in a certain way and therefore is responsible for the effects of the action [15]. This describes two key factors—causality and intention—which affect blame assignment in negative outcomes [16]. If and when a person possesses the agency, both the intention to act and the ability to cause effects, then this person is responsible for the outcome of their actions.

Based on the literature, we hypothesize that:

H1: Humans will exhibit a self-serving attribution bias such that they will attribute more agency and more responsibility to self for a desirable outcome, whereas they will attribute more agency and responsibility to external factors for an undesirable outcome.

2.2. Human and machine agency

However, such self-serving bias may manifest in very different ways in the context of human-machine interaction. One of the major factors is how an individual perceives the machine, whether as an intelligent or autonomous agent or merely as a tool or device. Experiments in the “Computer are Social Actors (CASA)” paradigm show that humans treat machines (e.g., a computer) similarly to a human [4]. It is argued that individuals tend to “overuse human social categories”, and thus “mindlessly” apply social rules and expectations to machines [17].

The implications of such mindless social responses on agency attribution are unclear. Scholars argue that such responses are limited in breadth and depth, and are premature cognitive commitments, rather than manifestation of a belief that machines are essentially human [17]. Indeed, if the technology is perceived as only providing simple information or fulfilling utilitarian tasks, for instance, then humans would perceive themselves as more (or solely) responsible for conducting the actions and for the outcome [18].

Contrary to these earlier findings, recent studies show that users would attribute certain levels of agency both to themselves and to the machine when interacting, communicating or collaborating with it. Specifically, in such interaction, *human agency* is conceptualized as the user’s ability to influence the nature and course of an interaction [19], whereas *machine agency* is the perceived level of autonomy and volitional control usually embedded in the functionality and other capabilities of the machine. Empirical evidence shows that humans perceive technology as agents especially when the latter demonstrates intelligence and engages in social

communication [20]. In certain scenarios, technology can even be perceived as a threat to human agency [21], thus giving rise to a tension between machine agency and human agency [22].

Perception of the technology as automatic agents influences whether and how much they may be considered responsible, and even blamed in undesirable situations. An experimental study has found that an artificial intelligence (AI) driver receives a higher level of blame than a human driver in a hypothetical car accident scenario, and the intention of blaming comes from the participants' perception of dissimilarity between AI and human drivers and the severity of the outcomes [23].

Such findings reflect the transactional nature of the collaboration between human users and intelligent machines. According to a recent psychological model of human-AI interaction from the perspective of the theory of interactive media effects (HAI-TIME model), the extent to which users take ownership of the outcomes of a collaboration depends on the degree to which they have realized their own agency, as well as perceived favorable social exchange and augmentation of self [22]. To the extent the outcomes of the interaction are favorable, users are likely to view themselves as equal partners, but to the extent they do not deliver on transactional benefits, the tension between the machine and user is likely to lead to a blame game.

This transactional model further complicates the hypothesis based on the fundamental attribution error (H1), as it is applied to human-technology collaboration. It suggests that:

H1a: Human teammates would acknowledge both human agency and machine agency if the collaboration outcome is desirable; and,

H1b: Humans are likely to attribute more agency, and therefore blame, to their technology teammates when the interaction outcome is undesirable.

2.3. Sociability

Individual characteristics often influence how they perceive the outcome of an event and attribute responsibility, and their interaction with technology. A review of prior literature [24] shows that certain factors such as innovativeness [25] and desire for control [26] have been considered as predictors of adoption of intelligent systems. Others have examined demographic factors such as age and gender: Younger generations display higher engagement (e.g., [27]) and more positive attitude (e.g., [28]); whereas gender influences conversational styles with avatars [29]. Personality traits are also found to determine psychological variables such as trust, which is central

to allowing individuals to collaborate with a variety of technology teammates [30].

Among the user personality dispositions, *sociability*, which is considered to influence one's social interactions with other people, has emerged as one of the factors that may also affect their interactions with technology. Scholars (e.g., [31]) have been applying the "CASA" paradigm to compare human-human social interactions and human-AI interactions and whether personality traits would appear stable.

Would individuals demonstrate similar social and communicative dispositions when interacting with humans and technologies? Would a sociable person be more likely to prefer having a machine as their teammate? For people with shyness, would the presence of a machine elicit the discomfort and inhibition that would occur in the presence of other humans? We propose to investigate *sociability*, a tendency to affiliate with others or need to be with others [32]. Sociability could be an important factor in predicting human interaction and collaboration with technologies. Psychology and group dynamics literature has found that sociability predicts teamwork across various tasks (e.g., [33]). At the individual level, sociability is greatly associated with cooperativeness [34]; at the collective level, it is associated with group productivity [35].

In the specific context of human-technology collaboration, it is important to explore whether sociability may be associated with an enhanced tendency to apply social rules and expectations to the technological teammates. If so, sociable users may exhibit social characteristics (e.g., politeness, reciprocity, proactivity, conscientiousness, communicability, manners, etc.), which are considered beneficial and facilitative to their collaboration with the technology [36] [17]. These social characteristics are likely to reduce attribution biases and enhance enjoyment of the collaborative experiences, leading to positive user experiences, perceptions and behavioral intention. Therefore, we hypothesize that sociability moderates the relationship between collaboration outcome and agency attribution. Specifically:

H2a: Human teammates with a high level of sociability will exhibit reduced attribution bias; and
H2b: Human teammates with a low level of sociability will exhibit increased attribution bias.

Scholars are calling for more research into the "relational dynamics through which people associate with these technologies, and, in turn, relate to themselves and others" [37]. As already mentioned, the assumption that individuals would perceive technology as a person, which warrants human treatment or attribution, was once rejected and challenged by observations of mindless social

interactions [17]. But, these findings need re-examination, as we have been witnessing rapid advancement of technological development, and more importantly, a shifting nature of technology [38]: technologies have become much more human-like [39], imbued with greater social cues such as voice [40] and gender cues [41], communicative, embodiment of cultural values [42], and integrated into home and personal spaces [43]. An understanding of users' responses to these increasingly intelligent, automated technologies, and how such human-technology interaction or collaboration may challenge perceptions of agency and humanness, has significant cognitive, social, technical, ethical, and legal implications.

3. Method

An experimental study was conducted to examine the effects of task outcome and sociability as a personality characteristic on agency attribution, perceptions, and behaviors in a wayfinding game. Participants ($N = 31$) were college students ($M = 19.42$, $SD = 1.39$) recruited from a large university. The majority of participants were Caucasian (64.5%, $n = 20$), followed by Asian (16.1%, $n = 5$), African American (9.7%, $n = 3$), Hispanic (6.5%, $n = 2$), and Arab (3.2%, $n = 1$). They were instructed to play a way-finding game called "Treasure Hunt" (to find three treasures hidden locally in the shortest period of time) with aid from a mobile app particularly designed for this experimental study to provide navigation assistance.

3.1. Procedure

Upon participants' arrival at the lab, they were asked to complete a pre-test questionnaire measuring individual difference variables including demographics, frequency of using a computer, frequency of using a mobile device, and frequency of using maps, and sociability.

After completing the pre-test questionnaire, participants were given an LG smart phone that had the "Treasure Hunt" app. A brief instruction on how to use the app was also provided before they start the game. During this process, participants were primed to think the app as an intelligent teammate that will work with the participants to complete the game successfully.

Once they received the smartphone, participants would leave the lab and work together with the app on the treasure hunt. During the treasure hunt, the app will communicate with the participants to locate the treasures and identify a fastest route to complete the

tasks. GPS tracking was activated on the smartphone to monitor their performance. After participants returned, they would receive a report informing them of their game performance ranking among all the players (either as the winner or as the loser of the game). The ranking was pre-assigned and irrelevant to their actual performance. After viewing the game result, participants filled out a post-test questionnaire measuring their experience and perceptions (i.e. pleasantness, and satisfaction), behavioral intention (i.e. intention to share their experience with others), as well as their outcome attributions to themselves and to the app.

3.2. Manipulation and measurement

The independent variable, *task outcome*, was manipulated by varying participants' game results. Instead of reporting their actual ranking, participants were randomly assigned to either a win condition or a loss condition. In the win condition, participants were congratulated and informed that they were "ranked #1 out of the 74 treasure hunters," in the loss condition, participants received a game performance report stating, "Not fast enough... You are currently ranked #68 out of the 74 treasure hunters."

Moderating and outcome variables were measured using self-report instrument. Unless otherwise specified, all items were measured on a seven-point scale, ranging from "Strongly Disagree" to "Strongly Agree." *Sociability* (Cronbach's $\alpha = .83$, $M = 5.36$, $SD = .81$) was measured using a nine-item scale [12], including items such as "I find it easy to get along with people," "I am very involved in the social world," "I find it easy to get along with people," and "I always seek social contacts." *Outcome attribution to the app* ($M = 4.77$, $SD = 1.31$) was measured by one item "the app deliberately acted in the given way" whereas *outcome attribution to self* ($M = 5.03$, $SD = 1.43$) was measured by one item "I felt that I had control over the game process." *Intention to share game experience* ($M = 4.97$, $SD = 1.49$) was measured by asking "I would like to talk about my experience with my friends." Using a seven-point semantic differential scale, *pleasantness* ($M = 5.35$, $SD = 1.79$) and *satisfaction* ($M = 5.13$, $SD = 1.83$) were measured by asking participants to rate their feelings after gameplay on these semantic differential items "Annoyed : Pleased" and "Unsatisfied : Satisfied," respectively. Last but not least, *frequency of using a computer* ($M = 6.42$, $SD = .89$), *a mobile device* ($M = 6.81$, $SD = .40$), and *maps* ($M = 3.90$, $SD = 2.04$) were measured by asking participants to indicate, on average, how often they use each item on a seven-point scale, ranging from "Never" to "All the time."

4. Results

4.1. Manipulation check

Manipulation check was conducted and confirmed that the task outcome manipulation was successful, $t(29) = -7.64, p < .001$. Participants in the win condition ($M = 6.06, SD = 1.65$) were more likely to view their performances as better than other players compared to those who were in the loss condition ($M = 2.20, SD = 1.08$).

4.2. Agency attribution

A series of two-way analyses of covariance (ANCOVAs) controlling for age, gender, and frequency of using a computer, a mobile device, and maps, were conducted.

For agency attribution, analyses showed no statistically significant main effect of task outcome on agency attribution either to the app, $F(1, 22) = 2.22, p = .15$, or to self, $F(1, 22) = 3.90, p = .06$. This near-significant effect suggests support to self-serving bias, but, overall, H1, H1a and H1b were not supported.

However, a significant main effect of sociability was found on attribution to the app, $F(1, 22) = 6.72, p < .05$. Participants who self-reported higher levels of sociability tended to ascribe more intention to the app than those with lower levels of sociability ($B = .70, SE = .27$).

In addition, as shown in Figures 1 and 2, results showed a statistically significant interaction effect between task outcome and sociability on outcome attribution to self, $F(1, 22) = 8.50, p < .01$, and a marginally significant effect on outcome attribution to the app, $F(1, 22) = 3.98, p = .058$. Specifically, participants who self-reported being less sociable tended to attribute less intention to the app when they had lost than when they had won. They were also more likely to attribute the outcome to self in the win condition than loss (H2b partially supported). In contrast, for participants with higher levels of sociability, task outcome appeared to have less significant, or even reverse effects, on outcome attribution to both app and self (H2a supported).

4.3. User experience and behavioral intention

Data showed that task outcome had significant main effects on pleasantness, $F(1, 22) = 4.44, p < .05$, and satisfaction, $F(1, 22) = 4.14, p = .05$. Participants in the win condition reported higher levels of pleasantness ($M = 6.16, SE = .57$) and satisfaction ($M = 5.95, SE = .52$) than those who were in the loss

condition (pleasantness: $M = 4.58, SE = .48$; satisfaction: $M = 4.31, SE = .61$).

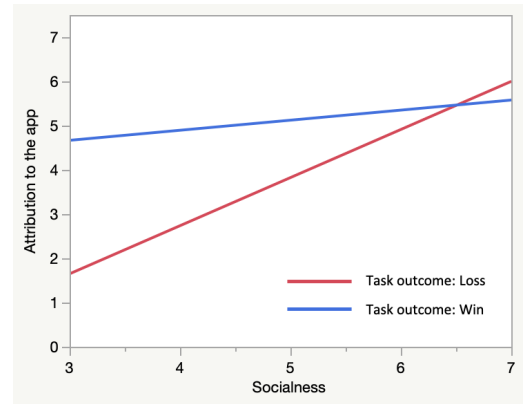


Figure 1. Interaction effect between task outcome and sociability on attribution to app

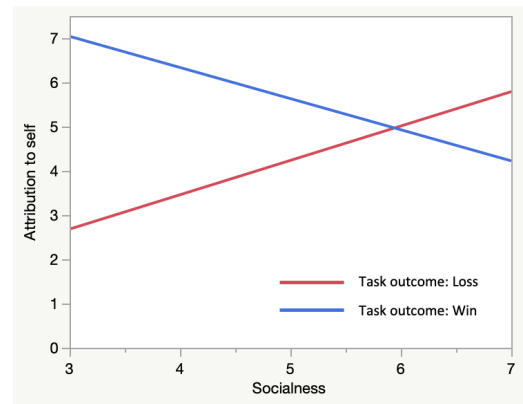


Figure 2. Interaction effect between task outcome and sociability on attribution to self

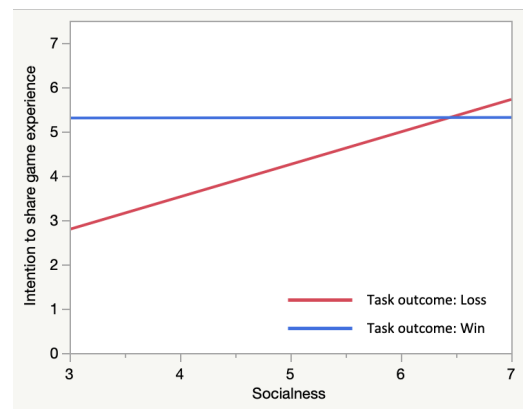


Figure 3. Interaction effect between task outcome and sociability on behavioral intention

Table 1. Results of the moderated mediation model

Independent Variable	Moderator Sociability	Mediators	Dependent Variables					
			Pleasantness		Satisfaction		Intention to share game experience	
			<i>B (SE)</i>	95% CI	<i>B (SE)</i>	95% CI	<i>B (SE)</i>	95% CI
Effect of task outcome at sociability ⁷ value of	4.56		.50 (.45)	-.29, 1.46	.49 (.49)	-.43, 1.52	.32 (.38)	-.40, 1.10
	5.44	Outcome attribution to the app	.19 (.24)	-.12, .80	.18 (.27)	-.18, .86	.12 (.18)	-.22, .52
	6.33		-.12 (.30)	-.63, .64	-.12 (.34)	-.67, .73	-.08 (.24)	-.65, .36
Index of moderated mediation			-.35, (.33)	-.98, .33	-.34 (.37)	-1.02, .41	-.22 (.29)	-.90, .28
Effect of task outcome at sociability ⁷ value of	4.56		2.16 (.99)	.50, 4.38	1.80 (.97)	.27, 3.98	1.31 (.75)	.19, 3.05
	5.44	Outcome attribution to self	.76 (.59)	-.21, 2.08	.64 (.55)	-.16, 1.93	.46 (.43)	-.11, 1.51
	6.33		-.63 (.87)	-2.36, 1.10	-.53 (.79)	-2.08, 1.09	-.38 (.54)	-1.42, .72
Index of moderated mediation			-1.57 (.82)	-3.21, -.13	-1.31 (.78)	-2.97, -.03	-.95 (.56)	-2.18, -.05

Note. Values of sociability are the 16th, 50th, and 84th percentiles.

In addition, the results showed a marginally significant interaction effect between task outcome and sociability on behavioral intention to share their experience with others, $F(1, 22) = 4.02, p = .058$. Participants who self-reported as less social were more likely to share their game experience with others if they had won, rather than lost, the game; the effect of outcome on behavioral intention was nonsignificant among those with higher levels of sociability (see Figure 3).

4.4. Agency attribution as mediators

To test the moderated mediation on user experience and behavioral intention, Hayes' PROCESS Macro (model 7) with 95% bias-corrected confidence intervals using 5,000 bootstrap resamples was employed [44]. The interaction effects between outcome and sociability were tested, and outcome attribution to the app and to self were examined as potential mediators. Age, gender, frequency of using a computer, a mobile device, and maps were entered as control variables.

The results show significant indirect interaction effects between task outcome and sociability on pleasantness, satisfaction, and behavioral intention. However, the moderated mediation through attribution to the app was not significant. The findings showed that participants with lower levels of sociability attributed the outcome to themselves, which in turn led to positive perceptions, and enhanced behavioral

intention. However, there were no significant indirect effects for individuals with higher levels of sociability (See Table 1).

5. Discussion

Our findings reveal that agency attribution in human-technology collaboration is more complex than previously known. When examining task outcome as a singular predictor, no difference was found in the levels of agency attributed to the technology and the users themselves. The lack of statistically significant main effects, however, may be because personality traits such as sociability moderate the influence of task outcome on how individuals attribute responsibility.

5.1. Sociable users recognize machine agency

Across different outcome scenarios, sociability appears to be associated with the tendency to attribute agency to technology. A closer look reveals that such a relationship is driven by the fact that those with lower levels of sociability did not blame the technology for the failure. Specifically, participants who were relatively low on sociability attributed much less agency to the technology when they lost the game, compared to the success scenario, and to participants who were more sociable, regardless of the task outcome.

For those with higher levels of sociability, perceptions of self and app agency remained relatively

stable, and relatively high, regardless of the task outcome. It appears that individuals who are inherently more social tend to view their technological teammates in more social ways as well. Both in success and in failure, sociable users are likely to consider the machine as intentional and agentic.

5.2. Low sociability contributes to attribution bias

Attribution bias only has appeared partially among participants with lower levels of sociability, when asked to assign a certain level of responsibility to themselves based on the task outcome. They ascribed more responsibility to themselves in the success scenario compared to the failure scenario. This finding could be explained by the positive correlation between sociability and self-esteem established in previous literature (e.g., [32]), such that the self-enhancement bias serves to maintain or enhance self-esteem [12].

More interesting, in the loss scenario, less sociable users displayed denial of agentic involvement from either entity. They might have viewed the teamwork as dysfunctional by attributing low levels of agency to both the machine and themselves.

5.3. Theoretical and practical implications

These findings provide useful insights, both theoretically and practically, into how individuals attribute agency, responsibility, and blame when interacting and collaborating with technology. They show the complexity of agency and blame attribution in human-technology collaboration. Specifically, our work adds to the existing literature with empirical evidence of how personality characteristics play an important role, and proposes a theoretical framework in which sociability moderates attribution outcomes by applying social rules and expectations to technological teammates, and changing agency and blame attribution.

Our research contributes to the understanding of blame attribution, which is an increasingly pressing socio-ethical and socio-technical issue, for instance, with the rapid development of autonomous vehicles (AVs). Despite great potential for safety and convenience, AVs receive skepticism and controversy due to the uncertainty of moral norms for these vehicles. Some researchers have named the phenomenon “blame attribution asymmetry” [45], such that AVs are blamed more harshly, and blame and responsibility are more likely to be attributed to the AVs. Blaming the AVs is observed experimentally and in real life, for harms that are actual or imagined

[46]. Our research shows the psychological mechanism through which users with certain personality characteristics (i.e. high sociability) apply social rules and expectations to the technology, attribute a high level of agency to it, and hold it responsible when things go wrong.

The self-serving bias widely identified in these studies (e.g., [46], [47]), that is, to attribute more responsibility of a car accident to the AVs than to self, was only partially present in our findings. Especially in our loss scenario, participants assigned the same, or slightly higher, level of responsibility to themselves. One plausible explanation is that the consequences of losing this way-finding game were minimal. Motivation to downplay one’s own fault is not comparable to car crashes that cause moderate to severe harm. Engineers and developers of autonomous and collaborative technology should consider the implications of these findings, and recognize the varying psychological responses to different risk scenarios and outcomes. Researchers should further explore other factors, such as shared control [48][49], varying levels of automation [50], and personal relevance in a loss scenario, which may also alter the blame attribution.

The findings also have implications for computer-supported and AI-enabled collaborative tools designed to gain user compliance for achieving desirable outcomes (e.g., health outcomes, civic engagement) through human-technology collaboration. Intelligent agents could be designed with features and cues that serve both functional and social needs. In order to enhance user experience and perceptions, these features and cues should be personalized to meet any individual user’s personality traits, communication styles, and collaboration preferences.

5.4. Future research directions

Future research should examine other key individual characteristics that influence agency attribution, such as locus of control. A wide range of personality traits such as introversion, agreeableness, and conscientiousness, which have been associated with trust, motivation, adoption, and engagement with technology (e.g., [51], [52], [30]), can also be considered.

Researchers should consider factors and analyses at the team level, to understand whether individuals perceive beyond their own agency but shared intentions and shared goals between themselves and their technological teammates. We call on scholars and research teams to test whether our findings are replicable in a variety of team structures and collaboration scenarios, with human users from

different demographics and cultural backgrounds, and involving a variety of task types and complexity, as well as different levels of outcome severity and personal significance.

6. Conclusions

Socio-ethical and socio-legal concerns that have arisen with the development of intelligent and automated technologies need to be addressed beyond their apparent technical aspects [53] [54]. It appears that the solution is not full automation or advanced functionality, but rather efficacious human-technology coordination and collaboration [55]. An interdisciplinary approach that examines and reimagines the different roles played by the technology, the users, the communicative and collaborative processes, and their societal implications is needed to make such a solution a reality.

7. Acknowledgement

An earlier version of this paper was presented as an extended abstract at the annual conference of the International Communication Association.

The first two authors, Haiyan Jia and Mu Wu, contributed equally to this work. They both contributed significantly to the conceptualization and execution of the study, including study and stimuli design, data collection, data analysis, report writing, and manuscript revisions. S. Shyam Sundar helped with the conceptualization and execution of the study. He also contributed to report writing and manuscript revisions.

8. References

- [1] C. Nass, J. Steuer, and E. R. Tauber, "Computers are Social Actors", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'94), 1994, pp. 72- 78.
- [2] I. Seeber, E. Bittner, R. O. Briggs, T. De Vreede, G. J. De Vreede, A. Elkins, R. Maier, A. B. Merz, S. Oestereib, N. Randrup, G. Schwabe, and M. Söllner, "Machines as Teammates: A Research Agenda on AI in Team Collaboration", *Information & Management*, Elsevier, 2020.
- [3] E. de Visser, R. Pak and T. Shaw, "From 'Automation' to 'Autonomy': The Importance of Trust Repair in Human-machine Interaction", *Ergonomics*, Taylor & Francis, 2018, pp. 1409-1427.
- [4] C. Nass, B. J. Fogg, and Y. Moon, "Can Computers Be Teammates?," *Int. J. Hum. Comput. Stud.*, Elsevier, 1996, pp. 669-678.
- [5] G. Matthews, J. Lin, A. R. Panganiban, and M. D. Long, "Individual Differences in Trust in Autonomous Robots: Implications for Transparency," *IEEE Trans. Hum. Mach. Syst.*, IEEE Systems, Man, and Cybernetics Society, 2020, pp. 234-244.
- [6] R. Zhang, N. J. McNeese, G. Freeman, and G. Musick, "An Ideal Human: Expectations of AI Teammates in Human-AI Teaming," *Proc. ACM Hum. Comput. Interact.*, Association for Computing Machinery (ACM), 2021, pp. 1-25.
- [7] H. H. Kelley and J. L. Michela, "Attribution Theory and Research," *Annu. Rev. Psychol.*, Annual Reviews, 1980, pp. 457-501.
- [8] F. Heider, "The Psychology of Interpersonal Relations", Wiley, Moscow, Russian Federation, 1958.
- [9] R. B. Cialdini, R. J. Borden, A. Thome, M. R. Walker, S. Freeman, and L. R. Sloan, "Basking in Reflective Glory: Three (Football) Field Studies", *Journal of Personality and Social Psychology*, APA Publishing, Washington, DC, 1976, pp. 366-375.
- [10] R. B. Cialdini and K. D. Richardson, "Two Indirect Tactics of Image Management: Basking and Blasting", *Journal of Personality and Social Psychology*, APA Publishing, Washington, DC, 1980, pp. 406-415.
- [11] E. P., Downs, and S. S. Sundar, "'We Won' Vs. 'They Lost': Exploring Ego-enhancement and Self-preservation Tendencies in The Context of Video Game Play", *Entertainment Computing*, Elsevier, 2011, pp. 23-28.
- [12] M. L. Snyder, W. G. Stephan, and D. Rosenfield, "Egotism and Attribution", *Journal of Personality and Social Psychology*, APA Publishing, Washington, DC, 1976, pp. 435-441.
- [13] J. Shepperd, W. Malone, and K. Sweeny, "Exploring Causes of the Self-Serving Bias", *Social and Personality Psychology Compass*, Wiley, 2008, pp. 895-908.
- [14] G. W. Bradley, "Self-Serving Biases in The Attribution Process: A Reexamination of The Fact or Fiction Question", *Journal of Personality and Social Psychology*, APA Publishing, Washington, DC, 1978, pp. 56.
- [15] M. Coeckelbergh, "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability," *Sci. Eng. Ethics*, Springer, 2020, pp. 2051-2068.
- [16] B. Weiner, "Judgments of Responsibility: A Foundation for a Theory of Social Conduct", Guilford Press, New York, NY, 1995.
- [17] C. Nass and Y. Moon "Machines and Mindlessness: Social Responses to Computers", *Journal of social issues*, Wiley, 2000, pp. 81-103.
- [18] G. Klein, D. Woods, J. Bradshaw, R. Hoffman, and P. Feltovich, "Ten Challenges for Making Automation a Team Player in Joint Human-Agent Activity", *IEEE Intelligent Systems*, IEEE Computer Society, 2004, pp. 91-95.
- [19] S. S. Sundar, "Agency and On-line Media", *Encyclopedia of Children, Adolescents, and the Media*, Sage Publications, Thousand Oaks, CA, 2016, pp. 635-636
- [20] H. Jia, M. Wu, E. Jung, A. Shapiro, and S. S. Sundar, "When The Tissue Box Says 'Bless You!': Using

- Speech To Build Socially Interactive Objects”, Proceedings of the 2013 Annual Conference on Human Factors in Computing Systems (CHI’13), 2013, pp. 1635-1640.
- [21] H. Jia, M. Wu, E. Jung, A. Shapiro, and S. S. Sundar, “Balancing Human Agency and Object Agency: An In-Depth Interview Study of The Internet of Things”, Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp’12), 2012, pp. 1185-1188.
- [22] S. S. Sundar, “Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAI)”, *J. Comput. Mediat. Commun.*, Oxford University Press, 2020, pp. 74–88.
- [23] J. W. Hong, “Why Is Artificial Intelligence Blamed More? Analysis of Faulting Artificial Intelligence for Self-Driving Car Accidents in Experimental Settings”, *International Journal of Human–Computer Interaction*, Taylor & Francis, 2020, pp. 1768-1774.
- [24] C. Rzepka and B. Berger, “User Interaction With AI-Enabled Systems: A Systematic Review of IS Research”, Proceedings of 2018 Annual International Conference on Information Systems, 2018.
- [25] F. Nasirian, M. Ahmadian, and O. K. D. Lee, “AI-based Voice Assistant Systems: Evaluating from the Interaction and Trust Perspectives”, *Adoption and Diffusion of Information Technology (SIGADIT)*, 2017.
- [26] I. Gaudiello, E. Zibetti, S. Lefort, M. Chetouani, and S. Ivaldi, “Trust as Indicator of Robot Functional and Social Acceptance. An Experimental Study on User Conformation to Icube Answers”, *Computers in Human Behavior*, Elsevier, 2016, pp. 633-655.
- [27] S. Shahid, E. Kraemer, and M. Swerts, “Child–Robot Interaction Across Cultures: How Does Playing a Game with a Social Robot Compare to Playing a Game Alone or with a Friend?” *Computers in Human Behavior*, Elsevier, 2014, pp. 86-100.
- [28] J. Martínez-Miranda, H. Pérez-Espinosa, I. Espinosa-Curiel, H. Avila-George, and J. Rodríguez-Jacobo, “Age-based Differences in Preferences and Affective Reactions Towards a Robot’s Personality During Interaction”, *Computers in Human Behavior*, Elsevier, 2018, pp. 245-257.
- [29] D. C. Derrick and G. S. Ligon, “The Affective Outcomes of Using Influence Tactics in Embodied Conversational Agents”, *Comput. Human Behav.*, Elsevier, 2014, pp. 39–48.
- [30] J. S. Elson, D. Derrick, and G. Ligon, “Examining Trust and Reliance in Collaborations Between Humans and Automated Agents,” Proceedings of the 51st Hawaii International Conference on System Sciences, University of Hawaii at Manoa, 2018.
- [31] Y. Mou and K. Xu, “The Media Inequality: Comparing the Initial Human-Human and Human-AI Social Interactions”, *Computers in Human Behavior*, Elsevier, 2017, pp. 432-440.
- [32] J. M. Cheek and A. H. Buss, “Shyness and Sociability”, *Journal of personality and social psychology*, APA Publishing, Washington, DC, 1981, pp. 330.
- [33] R. L. Anhalt, “Exploring the relationship between personality characteristics and effective teamwork behaviors”, *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 1995, pp. 3488.
- [34] L. Lu and M. Argyle, “Happiness and Cooperation”, *Personality and Individual Differences*, Elsevier, 1991, pp. 1019-1030.
- [35] J. E. Driskell, E. Salas, and S. Hughes, “Collective Orientation and Team Performance: Development of an Individual Differences Measure”, *Human Factors*, Sage Publishing, 2010, pp. 316-328.
- [36] A. P. Chaves and M. A. Gerosa, “How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design”, *Int. J. Hum. Comput. Interact.*, Taylor & Francis, 2020, pp. 729–758.
- [37] A. L. Guzman and S. C. Lewis, “Artificial Intelligence and Communication: A Human–Machine Communication Research Agenda”, *New Media & Society*, Sage Publishing, 2020, pp.70-86.
- [38] J. Reeves, “Automatic for the People: The Automation of Communicative Labor”, *Communication and Critical/Cultural Studies*, Taylor & Francis, 2016, pp. 150–165.
- [39] S. Zhao, “Humanoid Social Robots as a Medium of Communication”, *New Media & Society*, Sage Publishing, 2006, pp. 401-419.
- [40] A. L. Guzman, “Voices in and of the Machine: Source Orientation Toward Mobile Virtual Assistants”, *Computers in Human Behavior*, Elsevier, 2019, pp. 343–350.
- [41] W. R. Poster, “The Virtual Receptionist with a Human Touch: Opposing Pressures of Digital Automation and Outsourcing in Interactive Services”, *Invisible Labor: Hidden Work in the Contemporary World*. University of California Press, Oakland, CA, 2016, pp. 87-112
- [42] C. Marvin, “Reconsidering James Carey: How Many Rituals Does It Take to Make an Artifact?” *American Journalism*, Routledge Publishing, 1990, pp. 216-226.
- [43] L. Fortunati, “Robotization and the Domestic Sphere”, *New Media & Society*, Sage Publishing, 2018, pp. 2673-2690.
- [44] A. F. Hayes, “Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, Second Edition”, Guilford Press, New York, NY, 2018.
- [45] P. Liu, and Y. Du, “Blame Attribution Asymmetry in Human–Automation Cooperation”, *Risk Analysis*, Wiley-Blackwell, 2021.
- [46] M. Franklin, E. Awad, and D. Lagnado, “Blaming Automated Vehicles in Difficult Situations”, *Iscience*, Cell Press, Cambridge, MA, 2021.
- [47] T. Gill, “Blame It on The Self-Driving Car: How Autonomous Vehicles Can Alter Consumer Morality”, *Journal of Consumer Research*, Oxford University Press, 2020, pp. 272-291.
- [48] D. B. Shank, A. DeSanti, and T. Maninger, “When Are Artificial Intelligence Versus Human Agents Faulted for Wrongdoing? Moral Attributions After Individual

- and Joint Decisions,” *Information, Communication & Society*, Routledge, 2019, pp. 648–663.
- [49] E. Awad, S. Levine, M. Kleiman-Weiner, S. Dsouza, J. B. Tenebaum, A. Shariff, J. Bonnefon, and I. Rahwan, “Drivers are Blamed More Than Their Automated Cars When Both Make Mistakes,” *Nat. Hum. Behav.*, 2020, pp. 134–143.
- [50] E. Awad, S. Levine, M. Kleiman-Weiner, S. Dsouza, J. B. Tenebaum, A. Shariff, J. Bonnefon, and I. Rahwan, “Blaming Humans in Autonomous Vehicle Accidents: Shared Responsibility Across Levels of Automation”, arXiv:1803.07170 [cs.AI], Cornell University, 2018.
- [51] B. Caci and K. Dhou, “The Interplay Between Artificial Intelligence and Users’ Personalities: A New Scenario for Human-Computer Interaction in Gaming”, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, pp. 619-630.
- [52] A. Khalis and A. Y. Mikami, “Who's Gotta Catch'em All?: Individual Differences in Pokèmon Go Gameplay Behaviors”, *Personality and Individual Differences*, Elsevier, 2018, pp. 35-38.
- [53] M. Callier and H. Callier, “Blame It on the Machine: A Socio-Legal Analysis of Liability in an AI World”, *Wash. JL Tech. & Arts*, 1998, pp. 49.
- [54] B. Wagner, “Liable, But Not in Control? Ensuring Meaningful Human Agency in Automated Decision - Making Systems”, *Policy & Internet*, Wiley, 2019, pp. 104-122.
- [55] S. Nyholm and J. Smids, (2018). “Automated Cars Meet Human Drivers: Responsible Human-Robot Coordination and the Ethics of Mixed Traffic”, *Ethics and Information Technology*, Springer, 2018, pp. 335–344.