Social influences in the digital era: When do people conform more to a human being or an artificial intelligence?

Paolo Riva [*], Nicolas Aureli, Federica Silvestrini

Department of Psychology – University of Milano-Bicocca, Italy

A B S T R A C T

The spread of artificial intelligence (AI) technologies in ever-widening domains (e.g., virtual assistants) increases the chances of daily interactions between humans and AI. But can non-human agents influence human beings and perhaps even surpass the power of the influence of another human being? This research investigated whether people faced with different tasks (*objective* vs. *subjective*) could be more influenced by the information provided by another human being or an AI. We expected greater AI (vs. other humans) influence in *objective* tasks (i.e., based on a count and only one possible correct answer). By contrast, we expected greater human (vs. AI) influence in *subjective* tasks (based on attributing meaning to evocative images). In Study 1, participants (N = 156) completed a series of trials of an objective task to provide numerical estimates of the number of white dots pictured on black backgrounds. Results showed that participants conformed more with the AI's responses than the human ones. In Study 2, participants (N = 102) in a series of subjective tasks observed evocative images associated with two concepts ostensibly provided, again, by an AI or a human. Then, they rated how each concept described the images appropriately. Unlike the *objective* task, in the *subjective* one, participants conformed more with the human than the AI's responses. Overall, our findings show that under some circumstances, AI can influence people above and beyond the influence of other humans, offering new insights into social influence processes in the digital era.

1. Introduction

The study of social influence considers how the presence (or absence) of "others" can affect people's thoughts, emotions, and behaviors. Until recently, these "others" were supposed to be mostly, if not always, human beings. However, in recent years, with the advent and diffusion of a large variety of non-human agents based on artificial intelligence—including chatbots, robots, self-driving cars, and virtual assistants—the possible sources of social influence may have expanded beyond just human sources. Thus, a fascinating question is whether the current social perception of such non-human agents can act as a source of social influence, and if so, under which circumstances they can also exceed the influence of other human beings.

Here, we conducted two studies in which we contrasted simultaneously two possible sources of influence: another human being and an artificial intelligence. In the first study, we used an objective task: counting elements where only one correct answer is possible. In the second study, we used a subjective task, that is, a task of attribution of meaning based on the evocative power of a set of images. By

implementing a direct competition between a human vs. an artificial agent, we aimed to investigate whether these different potential sources of social influence could overcome each other depending on the type of task.

1.1. Conformity with human agents

The classic study of social influence has developed in the twentieth century mainly thanks to the pioneering works of Sherif (1937) and Asch (1951). Sherif (1937) investigated whether participants would conform to a group when the task presented was difficult, ambiguous, or uncertain. In this context, the correct answer or appropriate behavior is unknown to the participant. In one seminal study (1937), participants had to estimate the length of movement of a light spot displayed in a dark room. However, in the absence of reference points, a light point, although still, created the illusion of movement (i.e., the so-called autokinetic effect). The task was ambiguous as each observer would propose a different estimate based on how they subjectively perceived the optical illusion. In the typical Sherif's setting, participants would be

asked first to provide a series of estimations of different trials in a private condition and then perform the same task in a group setting. In the group condition, participants tended to align their estimations to the group's mean even if that differed from the initial assessment in the private condition. The study provided evidence of the group norm prevailing over the personal norm. Sherif (1937) stated that participants conformed to the group because other participants' responses seemed informative of the right solution. This type of social influence in which a person looks to others for the correct answer or the most appropriate behavior has been called *informational* social influence (Melamed et al., 2019). In contrast to *informational* influence, the study of *normative* social influence mainly developed thanks to the pioneering Asch (1951)'s study. The author proposed a line match task where the correct answer was clear. On this occasion, participants conformed to wrong answers in some of the critical trials due to the social pressure exerted by the group to avoid social exclusion (Hales et al., 2017).

The focus of the present work, however, is that of *informational* social influence. Among the others, Cialdini (1987) investigated the main factors involved in this type of social influence. The author suggests that uncertainty (not knowing the correct answer or the appropriate behavior) is a key antecedent of informational social influence. A study by Lucas et al. (2006) corroborates this assumption, showing how participants tended to conform more with the group during a math task when the math problem was particularly complex. Another factor in inducing informational social influence is the perceived degree of expertise in the task field of the agent of influence (Cialdini, 1987). Many studies observed that an expert's opinion is perceived as more valid and leads to more conformity than that of another person considered less expert, even when the message carried by the two agents of influence is the same (Goldman, 2001; Hovland & Weiss, 1951). Therefore, two fundamental antecedents of informational social influence are the ambiguity (uncertainty) generated by the situation and the perception of expertise of the sources (Hovland & Weiss, 1951; Lucas et al., 2006).

Social psychologists have recently begun studying social influence processes in digital contexts (Lee, 2004; Lee et al., 2006, 2011; Rosander & Eriksson, 2012). Lee (2004) studied informational social influence in computer-mediated communication. Participants cooperated in an online trivia game with another participant presented only through a cartoon character (whose gender appearance was admittedly randomly assigned). Results showed that a gender-stereotypical perception of the interactive partner could be triggered just by the appearance of the online character. Moreover, when stereotypically gender-biased topics were presented (e.g., sports vs. fashion), the topic moderated the effects of gender inference on conformity. The author suggested that conformity behavior could be linked to the perception of the character as more expert in some topics than others (based on their gender).

In Rosander and Eriksson (2012), participants had to answer a series of subjective or objective questions on different topics. The authors considered objective questions regarding logic or general knowledge where the answer was univocal (e.g., "In which city can you find Hollywood?"). In contrast, the subjective questions concerned participants' attitudes on matters without clear answers. Before answering the question, participants were exposed to the responses of previous users. The authors observed that most participants conformed at least once, and, both in subjective and objective tasks, they were more prone to conform when the task was difficult.

Coppolino Perfumi et al. (2019) investigated both normative and informational social in online contexts. Normative influence was measured through an online version of Asch's lines tasks, whereas informational influence was measured with two semantic tasks. In line with the author's expectations, the results show that the strength of normative influence was scarce in a digital context. In contrast, informational influence persists, and participants considered even unknown others a reliable source of knowledge. The authors also observed that higher conformity was associated with the more ambiguous tasks.

Therefore, informational social influence can also occur in digital-mediated contexts. Yet, the studies mentioned above focused on people interacting (sometimes through an avatar or cartoon) in digital environments. However, it is possible to wonder not only if the context (offline vs. online) has moderating effects on the social influence processes but also if the sources of influence can also be represented by non-human agents.

1.2. Conformity with non-human agents

In the last decades, artificial intelligence (AI) technology development has increased the presence of non-human agents in the social environment. Scholars have started investigating whether and how these non-human agents could influence humans. However, the literature on the influence of non-human agents is still limited and partly characterized by conflicting findings. For instance, some authors observed that non-human agents could effectively be sources of social influence (Brandstetter et al., 2017; Chidambaram et al., 2012; Hertz & Wiese, 2016, 2018; Salomons et al., 2018; Siegel et al., 2009; Williams et al., 2018). Other studies have found that humans do not seem to be affected by the social influence exerted by non-human agents (Beckner et al., 2016; Brandstetter et al., 2014; Shiomi & Hagita, 2016).

Some scholars suggested that these discrepancies might be linked to the specificities of each task proposed. Indeed, in studies that had a clear, correct answer, participants tend not to conform with non-human agents (Beckner et al., 2016; Brandstetter et al., 2014; Shiomi & Hagita, 2016), whereas when the task proposed was ambiguous or elicited uncertainty, the likelihood of conforming to a non-human agent increased (Hertz & Wiese, 2016, 2018; Salomons et al., 2018; Salomons et al., 2021). In line with this evidence, Lucas et al. (2019) tested the possibility that a non-human agent can elicit influence in an ambiguous online task. Participants rated ten artworks in order of importance based on their subjective perspectives. When they reported their responses, they could change their opinion after a brief discussion with the agent that used informational or normative social influence tactics. They observed that normative influence was ineffective, whereas informational influence did affect participants' responses.

Hertz and Wiese (2018) also explored non-human agents' conformity factors. They proposed a study where participants completed a series of analytical and social tasks. Participants were divided into three conditions, and before reporting their answers, they observed the previous responses of a human, robot, or computer. In addition to the ambiguity of the task, the authors suggested that a key factor in eliciting conformism was the perception of the match between agents and task type (i.e., agent-task fit). Accordingly, in the analytical task, participants conformed the same way with human and non-human agents; in the social task, participants conformed more with the human than non-human agents. According to the authors, participants were influenced by a specific agent (human or non-human) because they perceived the agent as more proficient in the task proposed. These results align with Cialdini's (1987) assumptions about the role of perceived expertise on information social influence. An agent perceived as an expert in a specific field could more likely exert social influence on people.

These phenomena could be accounted for thanks to the pioneering work of Rappaport (1970). The author discussed the possibility that, in the medical context, human beings could be better at some tasks while machines in others. For example, computers may be better than humans in retrieving all possible diagnoses that match a given set of symptoms in memory. According to the author, humans are subject to memory bias, fatigue, and a lack of time for accurate reflection, while computers are not subject to these limitations. This idea was then known by the acronym of "HABA MABA": namely, "Humans are better at" and "machines are better at," and it was subsequently taken up by various authors (de Winter & Dodou, 2014; Glikson & Woolley, 2020), each time to reiterate the possibility that humans and machines, because of their different capabilities, could perform better than each other depending

on the type of tasks.

However, even when considering non-human agents as possible sources of influence, existing studies tended to consider such sources of influence separately from others (e.g., human agents). However, to evaluate the effect on conformity of these new digital agents, the most effective comparison could be that in which various alternative sources of influence are presented simultaneously to the participant. The present research is placed in this perspective.

2. The present research

The present research investigates whether humans faced with different objective vs. subjective tasks eliciting uncertainty are more influenced by the information reported by another human being or an AI. In an online setting, we conducted two separate studies. Following the HABA-MABA framework, we tested a study where the match between agent and task could be perceived in favor of the AI (Study 1; objective) and a study where the human could be perceived as the most informative (Study 2; subjective). Thus, the objective task consisted in counting a series of white points on a black background in a limited time. This task required analytic capabilities typically associated with artificially intelligent agents (e.g., chatbots). In the subjective task, participants had to evaluate the association between a concept and an evocative image. The content of these images is evocative, requiring symbolic and emotional competencies, skills that are typically still primarily associated with human agents.

These expected matches would also align with studies highlighting source expertise's role as a key predictor of informational social influence (Cialdini, 1987). However, following the classic principles of informational social influence (Lucas et al., 2006; Melamed et al., 2019; Rosander & Eriksson, 2012; Salomons et al., 2021), both tasks were also designed to maximize the task difficulty and ambiguity. While subjective tasks are ambiguous by definition (since there is no right or wrong answer), the ambiguity of the objective task ultimately lies in its difficulty. Therefore, our objective task (i.e., counting hundreds of points in seven seconds) was thought impossible to solve accurately in the allotted time.

Considering these premises, we proposed the following hypotheses:

HP1. We expected that participants would conform more to the AI than the other human being in the objective tasks (Study 1).

HP2. We expected that participants would conform more with the other human being than the AI in subjective tasks (Study 2).

3. Study 1

Study 1 contrasted the social influence of human vs. non-human agents through an objective analytical task. The task requires watching a series of 8 visual stimuli for three different stages and estimating the number of white dots on a black screen that appears for 7 s. The study's main aim was to test the possibility that a non-human agent (presented to participants as an artificial intelligence agent) could overcome the influence of another human on an objective task. Moreover, Study 1 explored the possibility that, once learned, such influence could persist even without the source's direct influence (indirect influence).

3.1. Methods

3.1.1. Sample

We run an a priori power analysis using the software G*Power (version 3.0; Faul et al., 2007) to determine an adequate sample for our study. For an independent sample *t*-test, the required sample size to find a small-to-medium effect size (d = 0.40) with an alpha level equal to 0.05 and a conventional power of 0.80 is N = 156.

We used internal (i.e., Sona System, a platform to recruit study

participants provided by our university) and external (e.g., Facebook, Instagram) channels to advertise the study. The study was presented as part of a larger project on visual perception. Participants received a link to a Qualtrics (2020) survey to access the study. Overall, 197 persons participated in the online study. Twenty did not complete the survey reducing the sample to 177 (59.9 % females; $M_{age} = 29.98$, SD = 14.39). All the participants were Italian but one Cuban, one Spanish, one Moldavian, and one Swiss.

3.1.2. Procedure

The study was conducted online and composed of three different stages adopting the same set of 8 images. In each trial, a fixation point (1 s) was followed by an image depicting several white dots on a black background (see Fig. 1). The image with the dots stayed on the screen for 7 s, after which it disappeared automatically. This time frame (7 s) was chosen to give participants a rough idea of the number of dots without providing enough time to count the dots accurately. The actual number of dots in each image varied between 138 and 288. This task was adopted from Castelli et al.'s (2001) paradigm (see also Andrighetto et al., 2018), and used to assess conforming behavior.

The three stages were structured as follows (for a similar structure, see Riva et al., 2011). In Stage 1 (*baseline*), participants indicated for the first time their dots estimation guesses for each of the eight images, presented in random order. Stage 2 (*direct influence*) was similar to Stage 1. However, right after the presentation of each image, participants learned on a new screen the ostensible estimations provided by an AI and another other human. The two estimates appeared simultaneously (one below the other on the screen) with the label of the source (AI, Human) and alongside its associated numerical estimation (e.g., "human being: 270"). The location on the screen (one above the other) in which the two sources were shown was counterbalanced. Then, participants were asked again to provide their own estimates on a new screen.

To test our main hypothesis, participants were randomly divided into two independent groups. In the first group (N = 77; from now on, the AI-overestimation group), the AI systematically overestimated the number of dots (+15 % of the real number of the dots), whereas the other human systematically underestimated the number of dots (−15 %). In the second group (N = 79; from now on AI-underestimation group) was the opposite: the other human systematically overestimated the number of dots (+15 %), and the AI underestimated the number of dots (−15 %). Therefore, both agents of influence (i.e., human and AI) remained equally distant from the correct answer and never accurately estimated the number of the dots.

Finally, in Stage 3 (*indirect influence*), participants had to provide their guess one last time, in the absence of agents' estimations.

3.1.3. Measurements

In Stage 1 (*baseline*), we measured participants' dot estimation when
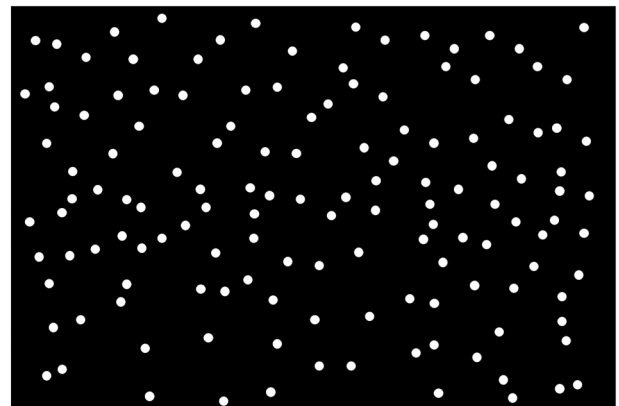

Fig. 1. An example of an image used in the objective task.

the agents were absent. We recorded participants' baseline in the absence of any sources of influence to assess indirect influence in Stage 3.

To assess participants' conforming behavior in terms of *direct influence* (Stage 2), we calculated two indices of influence, one for human influence and one for AI's influence based on Andrighetto et al.' (2018) operationalization. Specifically, the AI's index of influence was calculated as the difference between the participants' estimation and the anchor provided by AI. The absolute value of this difference was divided by the participant's estimation to reduce within-subject variance (see Wyer, 1966). Then we created an overall index of AI influence as the mean of the eight trials value of influence. We replicate the same operationalization for human influence. Our data thus allowed for a within-subjects comparison of the outcome. Lower scores resulted from a smaller distance between participants' and source estimations, thus, lower scores indicated higher conformism with the source.

The formula to calculate agent's influence was as following:

$$\frac{|Participant's\ estimation - Agent's\ Anchor|}{Participant's\ estimation}$$

We computed an index of *indirect influence*, subtracting their estimations in each task of Stage 3 with their previous answers in Stage 1 for the same image. Then we created an overall index as the mean of the eight trials value of indirect influence. Our data thus allowed for a between-subjects comparison on this outcome. Positive values were associated with participants who reported higher estimation in Stage 3 compared to Stage 1. In contrast, negative values were associated with participants who reported lower estimation in Stage 3 compared to Stage 1.

A brief questionnaire followed the three stages. Firstly, we asked which one of the two agents was perceived as the most accurate. In a multiple-choice question, participants could choose between "human being," "Artificial Intelligence," or "Neither of the two." Then, we asked participants to motivate their answers through an open question. Lastly, we collected sociodemographic information (i.e., gender, age, and nationality).

3.2. Results

3.2.1. Exclusion criteria for participants

Before conducting the main analyses, we screen data on the 24 trials to search for responses on the dot estimation task that significantly deviated from the group's mean. Considering that the participants were free to report any number from 1 to infinity, we had to pay attention to possible outliers' responses that could undermine the reliability of subsequent analyses. As two examples, one participant estimated 5 dots on an image containing 138 dots. Another participant reported an estimation of 12,354,357 dots for an image with 157 dots. These anomalous responses likely indicate poor compliance with the study, so we excluded these participants from our main analyses. Specifically, we used the standard deviation rule for identifying outliers (Dunn, 2021), considering outliers the participants who reported, in at least one task, estimation of + or – 3 SD from the mean of the group. As a result of this screening, twenty-one participants were excluded reducing the final sample to 156 participants (59 % females; M_{age} = 28.9, SD = 14.1).

3.2.2. Main analysis

We first conducted a paired sample *t*-test comparing how much the other human and the AI influenced the participants during the *direct influence* stage. We observed a significant difference between the influence exerted by the human (M = 0.31, SD = 0.28) and the influence exerted by the AI (M = 0.26, SD = 0.27); t (155) = −2.53, p = .012, d = −0.20. Considering that lower scores indicate higher conformism, participants conformed more to AI's influence than human ones. These results confirm our hypothesis (HP1) that people tend to conform to AI's

answers when facing an ambiguous but objective task.

To examine whether participants were also influenced when the sources of influence were absent, we compared participants in AI-overestimation conditions with the ones in AI-underestimation conditions. Specifically, we conducted an independent sample *t*-test to compare *indirect influence* in AI-overestimation and AI-underestimation conditions. We found a significant difference in scores for AI-overestimation (M = 35.46, SD = 59.10) and AI-underestimation conditions (M = 16.19, SD = 55.72); $t(154)$ = −2.10, p = .038, d = −0.34. Participants in AI-overestimation and AI-underestimation tend to increase their estimations in the indirect influence stage. Still, the participants in AI-overestimation reported higher values than participants in the AI-underestimation condition.

Finally, we investigated whether participants could explicitly state which of the two sources (AI vs. Human) was the more accurate. Our sample included 121 respondents who considered the AI more informative, 21 the human, and 14 neither of the two. A Chi-square test confirmed that participants considered the AI as significantly more informative than the human or the option "neither the human nor the AI", X^2 (2, 156) = 137.81, p < .001. Thus, participants were aware of their perceptions of AI's being more informative in this task.

4. Study 2

Study 2 investigated the social influence of human vs. non-human agents through a subjective task. The task involved selecting between competing concepts to describe best the evocative content of a series of 8 images. Competing concepts were provided simultaneously by an AI and another human. Study 2 main aim was to test the possibility that when subjectivity is involved, another human being influence would still be preferred to that of an AI. In doing so, we tested a boundary condition for the results of Study 1.

4.1. Methods

4.1.1. Sample

We run an a priori power analysis for a paired sample *t*-test setting a small effect size (d = 0.25; based on Study 1 results), a conventional power level at 0.80, and alpha at 0.05. The result of the analysis suggested including at least 101 participants in the sample.

Participants were recruited using the same channels of advertising as in Study 1. Overall, 118 persons accessed the online study on Qualtrics. Sixteen did not complete the study reducing the sample to 102 individuals (females 62 % M_{age} = 25.31, SD = 9.57). All participants were Italians but one Moldavan and one Swiss.

4.1.2. Procedure

Study 2 was composed of a single stage with eight trials. Each trial presented an evocative image taken from the card game Dixit (Libellud, 2008; see Fig. 2 for an example) without time limits. We used images taken from previous studies (Salomons et al., 2018; Salomons et al., 2021) to implement subjective tasks. Each evocative image was associated with two concepts ostensibly provided by an AI and a human as the best description of the picture. For each image, we identified two concepts plausibly associated with the figure. Then, we asked two research assistants, blind to the study's hypotheses, to evaluate whether one of the two concepts represented the image more. None of the presented concepts was perceived as more representative than the other. However, to further reduce the possibility that a more representative concept could be always associated with the same agent, we randomly assigned each concept across participants to the AI or the other human.

Under each image, we presented a 11-point Likert scale with the two associated concepts at its extremity. The scale was represented as a continuum from 1 "extremely representative of concept A" to 11 "extremely representative of concept B," with 6 associated with "concepts describe the image in the same way/ neither of the concepts

Fig. 2. An example of an image used in the subjective task. For this image, the two associated concepts were "serenity and freedom".

describe the image." Participants had to rate to what extent the image represented either of the two concepts. The order of presentation of the images was randomized across participants. Our data thus allowed for a within-subjects comparison of the selected outcome.

4.1.3. Measurements

To assess participants' conforming behavior, we operationalized an index of influence for the AI and one for the human. We calculated these indices simplifying the operationalization used in Study 1. Specifically, the AI's index of influence was calculated as the absolute value of the difference between the participant's response and the value associated with the AI's response (i.e., 1 or 11 based on which extremes of the continuum the source was collocated). Then we created an overall index of AI influence as the mean of the eight trials value of influence. We replicate the same operationalization for human influence. In doing so, we obtained two indices of influence, one for human influence and one for AI's influence. Lower scores resulted from a smaller distance between participant's and source's rating concepts, thus, lower scores were associated with a stronger influence from the source.

Finally, the same brief survey that ended Study 1 was adopted for completing Study 2 (i.e., selecting which agent was the most accurate, explaining why in an open-ended question, and providing sociodemographic information).

4.2. Results

We conducted a paired sample *t*-test comparing how much the participants were influenced by the other human and the AI during the single direct influence stage. We observed a significant difference between the influence exerted by the human (M = 4.74, SD = 1.12) and the influence exerted by the AI (M = 5.26, SD = 1.12); $t\,(102) = -2.33$, $p = .022$, $d = -0.23$. Considering that lower scores state higher influence, participants displayed greater conforming behavior to the human than

to AI.

Then, we investigated whether participants also perceived explicitly one source as more informative than the other. The sample included 45 respondents who considered the human as more informative, 29 the AI, and 28 neither of the two. A Chi-square test showed that participants did not report one of the agents as being more informative, $X^2\,(2,102) = 5.35$, $p = .069$.

5. General discussion

The advent of the digital age can also impact the classic processes of social influence. On the one hand, humans can frequently interact in online environments (from video calling platforms to the metaverse), which can introduce elements of similarity and difference compared to offline, face-to-face interactions. On the other hand, non-human agents (e.g., chatbots, social robots) can constitute direct sources of social influence and compete with other human beings to produce conformity in people. This second perspective was the focus of the present work. Our research shows that non-human agents can exert social influence on humans and that people are sensitive to the match between the type of influence source and the type of task in the context of informational social influence.

In Study 1, we presented a dot estimation task in three different stages to our participants. In the first stage, we collected a baseline of participants' responses. Then, during the *direct influence* stage, participants reported their estimation again, but this time after reading the human's and AI's estimation. We observed how the agents' responses directly influenced participants in this stage. Specifically, our findings suggest that during an objective task, such as a dot estimation, the AI is perceived by humans as more informative, leading to greater conforming to AI than other humans. We choose a difficult task to elicit more uncertainty, thus increasing the likelihood of informational social influence (Salomons et al., 2021). Besides ambiguity of the task, another pivotal factor in informational social influence is the perception of the influence source as an expert (Cialdini, 1987). As mentioned above, people are subjected to informative social influence because they do not know the correct answer/behavior in unclear situations, thus the expert's actions should be perceived as the most informative. Thus, the influence we observed in the objective task may be due to the perception of the AI as more proficient in analytic tasks. This assumption is corroborated by the previous literature on the social perception of non-human agents (Glikson & Woolley, 2020; Gray et al., 2007; Sundar, 2008). Gray et al. (2007) have shown that non-human agents are generally considered incapable of experiencing feeling and emotion but capable of rational thought and planning, restricting their expertise to analytical capabilities.

Furthermore, as highlighted above, the MABA-HABA ("Machine Are Better At vs. Humans Are Better At"; de Winter & Dodou, 2014; Glikson & Woolley, 2020; first described in Rappaport, 1970) theory aligns with our findings. According to the MABA-HABA framework, non-human agents have advantages over humans in specific domains (e.g., calculation, analytic task), whereas in other domains (e.g., emotional thinking), human capabilities outreach non-human's ones. Thus, non-human agents can be perceived as more informative in an objective task, leading to more conformism in their answers, as we observed in the current findings. Indeed, when we explicitly asked which agents were the most accurate, a significant majority of the participants chose AI. This result supports the idea that the conforming behavior that we observed was due to the perception of AI as more informative. Moreover, our results are in line with participants' responses to our open-ended question about the reasons why they perceived one agent as more accurate. For example, one participant wrote, "An AI can have an infinitely superior computing power to that of humans". Another participant stated, "For the human being it is difficult to count in a very short time. Artificial intelligence is faster because it is programmed to do so". Furthermore, of the 121 participants who reported the AI as the most

accurate, 89.2 % (69.2 % of the whole sample) specified in the open question that the accuracy of the AI derived from its analytic capabilities, supporting our assumptions.

Moreover, we found that participants continued to be influenced by AI's answers when the agents' influence was removed. In the third stage, participants were willing to change their first impression (*baseline* stage) following the exposure to the agents' estimations. Similar results were observed in a replication of Sherif's (1937) study conducted by Bovard (1948). The structure of the study was the same as the classic study, but, on this occasion, participants were re-contacted 28 days after the experiment, and they repeated the autokinetic task without the group. Although they were alone, participants continued to report the answer that emerged from the group almost a month early, showing how informative influence can persist over time (and the power of the group norm over the personal one). Brandstetter et al. (2017) observed similar results on robot social influence on human language. Participants' language not only was influenced by the non-human agent, but this influence also lasted beyond the time of interaction when the non-human agent was absent. Taking together, these results can be seen in the light of the nature of informative social influence. Indeed, unlike normative social influence, people undergo informational social influence even in the absence of the source because they can truly think that the behavior/ answer they are conforming to is the correct one. Our findings add to the literature showing that indirect social influence can occur not only when a human is the source but also for non-human agents.

Study 2 suggested that a person tends to conform more to another human (vs. AI) when facing a subjective task. Once again, the MABA-HABA theory comes to support our expectations. Indeed, according to this framework, humans are thought to perform better in tasks that require social or emotional intelligence, such as understanding the meaning of an image. However, our findings seem to conflict with a recent study conducted by Salomons et al. (2021), where non-human agents were capable of conformity in a subjective word-card matching task. The authors found that participants believed that the non-human agents were as proficient as themselves in the proposed task, although "robots do not usually perform well at high-level tasks such as understanding the meaning of images" (Salomons et al., 2021, p. 15:18). Indeed, their perception of non-human agents' capabilities explains why they conformed to non-human agents despite the mismatch between the agent and the task type. Nevertheless, Salomons et al. (2021) did not present a human as a competing source of influence. Therefore, it is possible that, when allowed to compare directly human vs. non-human agents, people would see more differences in terms of imaginative abilities. Looking at the open-ended responses we collected, we found some insight supporting the assumption that the other human exerted more informative social influence because of the assumed superior emotional capabilities. For example, one participant wrote, "[In human responses] there is more subjective and emotional conception, a missing component in AI." Another participant wrote, "[the human being's] language uses more the emotional sphere than a technical datum such as artificial intelligence." Therefore, humans' abstractive and emotional abilities might explain why we observed greater social influence (than the non-human agent) in the subjective task.

5.1. Limitations and future direction

There are some limitations regarding the current study. Our study did not measure mental states attributed to agents of influence. Future studies should consider the role of the attribution of mental states (anthropomorphism; Riva et al., 2015) both in terms of individual differences (moderator) and as a mechanism (mediator) that explains the conformism of a given agent. Moreover, future studies should also test whether, even in the subjective task, the influence persists when the sources of influence are no longer present.

The second limitation concerns the non-human agent chosen for the study. We decided to use an AI as a non-human agent to maximize the results' comprehensiveness to other non-human agents. Despite this, AIs can be programmed to perform different tasks, so not all AIs can be equally proficient in dot-counting or image recognition. Furthermore, in their review of studies on non-human agents, Glikson and Woolley (2020) observed how specific non-human agents (e.g., chatbots, AI, robots) are more or less influential based on their characteristics. Therefore, future research could propose a similar study but take into consideration how the influence exerted by the agent can vary depending on which specific non-human agent is presented. Considering that previous experience with a specific non-human agent is a key factor in conforming with the agent (Glikson & Woolley, 2020), familiarity in interacting with the agent could be a factor to be added to future studies.

Ultimately, in the years to come, the development of non-human agents will likely increase, and future generations of AI will probably exceed the capabilities of current AI. For example, future AI could be endowed with high abstractive abilities managing to perform well at high-level tasks such as the subjective task proposed by us. The results of our study could, therefore, not be generalizable over time but be specific to the present historical time. Something similar happened to Asch's studies of conformity. A review of Asch's replications has found that, over the years, people tend less and less to conform and move further away from Asch's initial results (Bond & Smith, 1996). For this reason, it will be important to keep investigating the social influence provided by human and artificial agents in the future decades to detect their role and possible changes over time.

6. Conclusion

Our research might offer new insights into social influence processes in the digital era. The results showed that people can conform more to non-human agents (than human ones) in a digital context under specific circumstances. For objective tasks eliciting uncertainty, people might be more prone to conform to AI agents than another human being, whereas for subjective tasks, other humans may continue to be the most credible source of influence compared with AI agents. These findings highlight the relevance of matching agents and the type of task to maximize social influence. Our findings could be important for non-human agent developers, showing under which circumstances a human is more prone to follow the guidance of non-human agents. Proposing a non-human agent in a task in which it is not so trusted could be suboptimal. Conversely, in objective-type tasks that elicit uncertainty, it might be advantageous to emphasize the nature of the agent as artificial intelligence, rather than trying to disguise the agent as human (as some existing chatbots tend to do). In conclusion, it is important to consider, on the one hand, that non-human agents can become credible sources of social influence and, on the other hand, the match between the type of agent and the type of task.

Declaration of competing interest

The authors declare no conflicts of interest.

References

Andrighetto, L., Baldissarri, C., Gabbiadini, A., Sacino, A., Valtorta, R. R., & Volpato, C. (2018). Objectified conformity: Working self-objectification increases conforming behavior. *Social Influence, 13*(2), 78–90.

Asch, S. E. (1951). In *Effects of group pressure upon the modification and distortion of judgments. Groups, leadership, and men* (pp. 222–236).

Beckner, C., Rácz, P., Hay, J., Brandstetter, J., & Bartneck, C. (2016). Participants conform to humans but not to humanoid robots in an English past tense formation task. *Journal of Language and Social Psychology, 35*(2), 158–179.

Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin, 119*(1), 111.

Bovard, E. W., Jr. (1948). Social norms and the individual. *The Journal of Abnormal and Social Psychology, 43*(1), 62.

Brandstetter, J., Rácz, P., Beckner, C., Sandoval, E. B., Hay, J., & Bartneck, C. (2014). A peer pressure experiment: Recreation of the asch conformity experiment with robots. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1335–1340.*

Brandstetter, J., Beckner, C., Sandoval, E. B., & Bartneck, C. (2017). Persistent lexical entrainment in HRI. *ACM/IEEE International Conference on Human-Robot Interaction, Part F, 1271*, 63–72. https://doi.org/10.1145/2909824.3020257

Castelli, L., Vanzetto, K., Sherman, J., & Arcuri, L. (2001). The explicit and implicit perception of ingroup members who use stereotypes: Blatant rejection but subtle conformity. *Journal of Experimental Social Psychology, 37*, 419–426.

Chidambaram, V., Chiang, Y., & Mutlu, B. (2012). Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 293–300).

Cialdini, R. B. (1987). In A. M. P. Harcourt (Ed.), *3. Influence*.

Coppolino Perfumi, S., Bagnoli, F., Caudek, C., & Guazzini, A. (2019). Deindividuation effects on normative and informational social influence within computer-mediated-communication. *Computers in Human Behavior, 92*, 230–237. https://doi.org/10.1016/j.chb.2018.11.017

de Winter, J. C. F., & Dodou, D. (2014). Why the fitts list has persisted throughout the history of function allocation. *Cognition, Technology & Work, 16*(1), 1–11.

Dunn, P. K. (2021). Scientific Research and Methodology: An introduction to quantitative research in science and health. Available online at: https://bookdown.org/pkaldunn/Book.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goldman, A. I. (2001). Experts: Which ones should you trust? *Philosophy and Phenomenological Research, 63*(1), 85–110.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619. https://doi.org/10.1126/science.1134475

Hales, A. H., Ren, D., & Williams, K. D. (2017). Protect, correct, and eject: Ostracism as a social influence tool. In *The Oxford handbook of social influence, Oxford handbooks* (pp. 205–217). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199859870.013.26.

Hertz, N., & Wiese, E. (2016). Influence of agent type and task ambiguity on conformity in social decision making. In *Proceedings of the Human Factors and Ergonomics Society* (pp. 313–317). https://doi.org/10.1177/1541931213601071

Hertz, N., & Wiese, E. (2018). Under pressure: Examining social conformity with computer and robot groups. *Human Factors, 60*(8), 1207–1218. https://doi.org/10.1177/0018720818788473

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15*(4), 635–650.

Lee, E. J. (2004). Effects of gendered character representation on person perception and informational social influence in computer-mediated communication. *Computers in Human Behavior, 20*(6), 779–799. https://doi.org/10.1016/j.chb.2003.11.005

Lee, M. K. O., Cheung, C. M. K., Sia, C. L., & Lim, K. H. (2006). *Proceedings of the Annual Hawaii International Conference on System Sciences, 6*(C), 1–10. https://doi.org/10.1109/HICSS.2006.204

Lee, M. K. O., Shi, N., Cheung, C. M. K., Lim, K. H., & Sia, C. L. (2011). Consumer's decision to shop online: The moderating role of positive informational social influence. *Information and Management, 48*(6), 185–191. https://doi.org/10.1016/j.im.2010.08.005

Libellud. (2008). *Dixit. Card game*.

Lucas, T., Alexander, S., Firestone, I. J., & Baltes, B. B. (2006). Self-efficacy and independence from social influence: Discovery of an efficacy–difficulty effect. *Social Influence, 1*(1), 58–80.

Lucas, G. M., Lehr, J., Krämer, N., & Gratch, J. (2019). The effectiveness of social influence tactics when used by a virtual agent. In *IVA 2019 - Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents* (pp. 22–29). https://doi.org/10.1145/3308532.3329464

Melamed, D., Savage, S. V., & Munn, C. (2019). Uncertainty and Social Influence. *Socius, 5*, 2378023119866971.

Qualtrics. (2020). Qualtrics. Available online at: http://www.qualtrics.com. (Accessed 30 March 2020).

Rappaport, M. (1970). Human factors applications in medicine. *Human Factors, 12*, 25–35.

Riva, P., Rusconi, P., Montali, L., & Cherubini, P. (2011). The influence of anchoring on pain judgment. *Journal of Pain and Symptom Management, 42*(2), 265–277.

Riva, P., Sacchi, S., & Brambilla, M. (2015). Humanizing machines: Anthropomorphization of slot machines increases gambling. *Journal of Experimental Psychology: Applied, 21*(4), 313.

Rosander, M., & Eriksson, O. (2012). Conformity on the internet - the role of task difficulty and gender differences. *Computers in Human Behavior, 28*(5), 1587–1595. https://doi.org/10.1016/j.chb.2012.03.023

Salomons, N., Sebo, S. S., Qin, M., & Scassellati, B. (2021). A minority of one against a majority of robots. ACM transactions on human-robot. *Interaction, 10*(2). https://doi.org/10.1145/3442627

Salomons, N., Van Der Linden, M., Strohkorb Sebo, S., & Scassellati, B. (2018). Humans conform to robots: Disambiguating trust, truth, and conformity. *ACM/IEEE International Conference on Human-Robot Interaction, 187–195.* https://doi.org/10.1145/3171221.3171282

Sherif, M. (1937). An experimental approach to the study of attitudes. *Sociometry, 1*(1/2), 90–98.

Shiomi, M., & Hagita, N. (2016). Do synchronized multiple robots exert peer pressure?. In *Proceedings ofthe Fourth International Conference on Human Agent Interaction* (pp. 27–33).

Siegel, M., Breazeal, C., & Norton, M. I. (2009). Persuasive robotics: The influence of robot gender on human behavior. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 2563–2568).

Sundar, S. S. (2008). The main model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger, & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 72–100). Cambridge, MA: The MIT Press.

Williams, R., Machado, C. V., Druga, S., Breazeal, C., & Maes, P. (2018). My doll says it's ok: A study of children's conformity to a talking doll. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 625–631). ACM.

Wyer, R. S. (1966). Effects of incentive to perform well, group attraction, and group acceptance on conformity in a judgmental task. *Journal of Personality and Social Psychology, 1*, 21–26.