https://doi.org/10.1007/s11370-022-00438-2

Social robot advisors: effects of robot judgmental fallacies and context

Torr Polakow[1] · Guy Laban[2] · Andrei Teodorescu[3] · Jerome R. Busemeyer[4] · Goren Gordon[1]

Received: 17 February 2022 / Accepted: 3 August 2022 / Published online: 27 August 2022

Abstract
The role of social robots as advisors for decision-making is investigated. We examined how a robot advisor with logical reasoning and one with cognitive fallacies affected participants' decision-making in different contexts. The participants were asked to make multiple decisions while receiving advice from both robots during the decision-making process. Participants had to choose which robot they agreed with and, at the end of the scenario, rank the possible options presented to them. After the interaction, participants were asked to assign jobs to the robots, e.g. jury or bartender. Based on the 'like-me' hypothesis and previous research of social mitigation of fallacious judgmental decisions, we have compared participants' agreement with the two robots for each scenario to random choice using t-tests, as well as analysed the dynamical nature of the interaction, e.g. whether participants changed their choices based on the robots' verbal opinion using Pearson correlations. Our results show that the robots had an effect on the participants' responses, regardless of the robots' fallaciousness, wherein participants changed their decisions based on the robot they agreed with more. Moreover, the context, presented as two different scenarios, also had an effect on the preferred robots, wherein an art auction scenario resulted in significantly increased agreement with the fallacious robot, whereas a detective scenario did not. Finally, an exploratory analysis showed that personality traits, e.g. agreeableness and neuroticism, and attitudes towards robots had an impact on which robot was assigned to these jobs. Taken together, the results presented here show that social robots' effects on participants' decision-making involve complex interactions between the context, the cognitive fallacies of the robot and the attitudes and personalities of the participants and should not be considered a single psychological construct.

Keyword Decision making, Robot Advisor, Conjunction fallacy, Human robot interaction

1 Introduction

Current research and development in the field of social robotics are positioning robots in a variety of roles, including tutors, companions, and peers, and even as care providers [1–4]. They have been used to assist elders with cognitive and social impairments [5], to deliver psychosocial interventions [6] and support rehabilitation [7,8], as well as to assist children in learning [2,3,9]. Social robots have been argued to be implemented in various other contexts in the future, such as bartenders [10] and psychologists [11], as well as jury [12] and analysts[13]". The more ubiquitous social robots become, the more we will rely on them for continuous decision making, e.g. consulting with and seeking advice from social robots on a daily basis [14–16].

Hence, an important goal of human–robot interaction (HRI) research is to create a robot interaction in which people feel comfortable and are willing to accept robots into their social circles [17]. In the social robotics and HRI literature, there are several psychological and neurocognitive explanations for why and how humans perceive robots as social entities and, in turn, accept their social presence [4]. According to the similarity effect [18] and the "like-me" hypothesis [19,20], people are more attracted to people like them. The early research into HRI investigated and explored the similarity effect and the "like-me" hypothesis mainly in terms of similarity as a visual stimulus (e.g. humanness of appearance

✉ Goren Gordon
goren@gorengordon.com

1 Curiosity Lab, Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv 6997801, Israel

2 Social Brain in Action Lab, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

3 Department of Psychology, Institute for Information Processing and Decision Making (IIPDM), University of Haifa, Haifa, Israel

4 Psychological and Brain Science, Indiana University, Bloomington, India

or motion) [21,22]. These homophily effects were recently studied in the context of HRI, where personalized robot speech has been shown to affect trust [23]. Moreover, several studies showed that task and appearance should match [24,25]. Thus, we expect that social robots' behavioural similarities to humans play a meaningful role in shaping humans' perceptions of them [26].

As social robots are designed and developed to take active roles in our social lives, they are likely to engage with humans in advising roles (e.g. as tutors, companions, and even as care providers) and impact our decisions. When people make decisions, they are prone to judgmental fallacies, such as conjunction and disjunction fallacies [27–30]. These fallacies mean they make decisions that contradict the logical choice—a choice that follows classic probability theory. However, it has also been shown that when people consult with other people, their fallacy rates decrease [31]. Furthermore, when experiencing social decision making, people tend to conform and be influenced by their peers [32]. In the context of HRI, social robots have been shown to induce conformity [33] when the robot is the minority, but adults resist social influence from a group of social robots [34].

The following question then arises: Will people prefer social robots that are more like them, e.g. make fallacious judgments, or more like "robots", e.g. present logical reasoning? Moreover, we envision social robots to occupy more diverse jobs [10–13], wherein different people will encounter them in various contexts. Thus, we wish to investigate people's perceptions in different contexts and how robots' attributes and people's personality traits affect these perceptions.

In this contribution, we aim to answer the following research questions: Will people *agree* more with fallacious robots than with logical robots? How does *agreement* change with context? Which robots, i.e. fallacious or logical, will be *assigned to the jobs* of jury, analyst, or bartender? How do people's *attitudes and personalities* affect their preferences for robot advisors? Which robot will generally be *preferred*?

2 Related works

2.1 Judgmental fallacies

People tend to make judgmental fallacies [27]. A fallacious behaviour is any behaviour that reflects a violation of basic laws that stem from classic probability theory [35]. In this paper, we focus on conjunction and disjunction fallacies, which violate the law of total probability as follows: *The conjunction fallacy* occurs when a person judges the probability of the conjunction of two events to be more likely than either of the constituent events. For example, when people assign a higher probability to the statement "Linda is a bank teller *and* a feminist", than to the statement "Linda is a bank teller", even though the latter is, by classical probability laws, at least as probable as the former. *The disjunction fallacy* occurs when a person judges the probability of the disjunction of two events to be less likely than either of the constituent events.

Previous studies have explored these fallacies and how common their occurrence is in people, and all these previous studies showed that most people tend to make fallacious decisions. The original paper [27] reported a frequency of 85%, a paper from a decade ago [31] reported a frequency of 58% and a recently published paper [36] reported a frequency of 50% for conjunction and 85% for disjunction.

However, several things can mitigate this behaviour. For example, Charness [31] investigated the occurrence of conjunction fallacies under conditions under which a group of individuals were to consult with each other. In a treatment with incentives, participants were informed that there was a correct answer and that anyone who chose this correct answer would receive 4$. In the treatment without incentives, participants were told that they would receive 2$ for filling out the questionnaire. These experiments were conducted with individuals, pairs, and trios of participants. When no incentive was given, the fallacy rates were 58%, 48% and 26% for single participants, pairs and trios, respectively. For the incentive condition, the fallacy rates were 33%, 13% and 10% for single participants, pairs and trios, respectively. These results are particularly interesting, as they show that incentives and/or cooperation with other participants reduce fallacy rates.

It has also been shown that the method of presentation can have dramatic effects on fallacy rates. Hertwig and Gigerenzer [30] replicated the results of [27] when the subjects were asked to rank or give probabilities to different options. However, they asked subject frequency questions such as "200 women have the same description as Linda above. Out of the 200 women, how many are bank tellers, bank tellers and feminist, etc". In this condition, the fallacy rates dropped below 20 per cent. This study highlights the importance of the question's framing and its effect on fallacy rates.

Finally, Polakow et al. [37] showed in an online study that when asked to "choose between ranking", people make significantly less fallacious judgments compared to when asked to "rank" the statements. Moreover, the effect that social robots have on fallacious decision-making was investigated. In each question, videos of two robots presented their answers to a question, where one answer was fallacious and one was logical. Participants had to choose which robot they agreed with. It was shown that people significantly more often chose the robot that answered logically. This result is similar to results reported by Charness [31], i.e. robots had effects on decision-making similar to those of human agents. In the current study, we have investigated interaction with *real robots* in sequential decision-making scenarios.

2.2 Fallacious artificial agents

While there is limited research on the effects of fallacious social robots, we can reflect on the matter when learning about decision-making processes with social robots. Moreover, relevant studies with other types of artificial agents and machines provide a meaningful context for the study of fallacious social robots.

Social robots were previously found to have a genuine effect on people's decisions in a variety of contexts and settings [4]. Social robots' influence on people's decisions is usually determined by the robot's available and related social cues, as well as people's expectations of the robot [38]. A recent study [39] found that a social robot that demonstrates more decisive behaviour (with minimal cues suggesting hesitant behaviour) is evaluated as more mentalistic (vs. mechanistic) compared to a social robot that demonstrates more hesitant behaviour. Using prisoner's dilemma games, another study established that people have a strong reciprocal tendency to social robots, which might even surpass the influence of the reward value of their decisions [40].

When social robots employ and demonstrate social cues, these cues are often attributed to higher degrees of mind perceptions [41] in terms of agency (the ability to plan and act) and experience (the ability to sense and feel) [42]. Accordingly, when social robots demonstrate human-like behaviour and social cues, people have expectations of these agents to similar those they have of humans [43]. Cognitive dissonance and the discrepancy between users' expectations of the agent following the agent's social cues and its performance typically affect people's perceptions and reactions to the agent and, accordingly, the likelihood of an agent influencing a person's decision-making [44]. For example, a previous study demonstrated that conversational agents that used invasive personalization techniques were perceived more negatively (i.e. the agent was viewed as being riskier and users felt in less control) than conversational agents that personalized recommendations without invading users' privacy. Accordingly, this had a substantial effect on how users perceived the recommendations that were provided by the agents and the likelihood that a user would follow these recommendations [45].

Based on these studies, we expect that social robots that correspond better to human logic (more fallacious) will be perceived in a more positive way and, accordingly, will have a higher potential to influence people's decision making. This is in contrast to the aforementioned study of Charness [31]. Hence, our study design delves deeper into this distinction, by presenting different contexts in which social robots give advice, as well as more ecological scenarios exhibiting sequential decision-making.

2.3 Social effects of imperfect robots

Previous studies have shown that imperfect robots, which exhibit behavioural errors, have both social advantages and disadvantaged in interactions with humans. On the one hand, Short et al. [46] investigated a rock-paper-scissors interaction. The robots behaved in one of the following three ways: fair, verbally cheating, cheating by declaring a different hand gesture that the one it performed. Their results showed that the cheating robots elicited more social engagement than the fair robots.

On the other hand, Salem et al. [47] studied how a robot's mistakes affect its trustworthiness and acceptance in human–robot collaboration. They found that faulty robots did not influence the task performance but were ranked as less trustworthy and reliable. Gompei and Umemuro [48] explored speech errors of robots. They found that the timing of the error influenced the familiarity and perceived sincerity of the robots. Early errors lowered the perceived sincerity, while later errors increased the robot's familiarity.

Another study [49] investigated a social interaction in which a robot and a human competed against each other. The results showed that faulty robots were rated as being less competent, reliable, and intelligent than error-free robots. Despite these findings, participants reported having enjoyed the interactions with the faulty robots more than those with the error-free robots. In a recent study [50], a robot and a human were required to complete LEGO building tasks together. There were three scenarios in with the robot was either faultless, committed social norm violations, e.g. by interrupting the participant when she was talking, or engaged in technical failures. They did not find significant differences in people's ratings of the robot's anthropomorphism and perceived intelligence, as measured by the Godspeed questionnaire [51]. However, participants liked the faulty robot significantly more than the faultless robot.

These studies suggest that an imperfect robot, i.e. a robot that behaves more like a human can sometimes be perceived as more socially acceptable. However, how a robot is perceived also depends on what the person thinks of it *prior* to the interaction, which can be influenced by contextual priming.

2.4 Contextual priming

Multiple studies in the past have shown that priming for a specific context or scenario in which one makes a decision affects the outcome [52,53]; for example, it has been shown that single words prime people and affect their impression of a (fictional) character. Moreover, it was shown that the participants were not aware of the priming that influenced their decisions. This study revealed that individuals' recent experiences could affect their perception of a fictional character.

One study [54] studied a fort game, wherein participants were asked to imagine themselves as generals and the enemy was going to attack their fort (from a specific direction). They needed to protect the fort by allocating soldiers to the gates. They showed that participants allocated their resources differently based on whether analytical priming or holistic priming was employed. Analytically primed participants concentrated their resources in the gate closer to the enemy. This is in contrast to participants who were primed with the holistic condition, as they allocated their resources more sparsely.

Another study [55] examined the difference between people required to make a decision while examining all the options either simultaneously or sequentially. They presented the participants with multiple options where one of option was obviously dominant over the others. Their results showed that under the simultaneous condition, 84.42% of the participants chose the dominant option versus 75.46% of those under the sequential condition. Moreover, they repeated the same procedure where the dominance relationship was more transparent. In this case, the numbers dropped to 64.34% (54.87%) under the simultaneous (sequential) condition.

In addition to contextual priming, the matching hypothesis conveys the concept that the appearance of a robot should match the task it performs [24,25]. Transformed to our scenario, the hypothesis states that the cognitive behaviour of the robot should match its task. Combining the matching hypothesis and contextual priming gives rise to the hypothesis that people will perceive and match robots with different cognitive fallacies to different contextual scenarios.

2.5 Robot role and job assignment

People naturally assign roles to one another while interacting. Similarly, when a person interacts with a robot, they also assign a role to the robot. We were interested in whether the fallaciousness of the robot's behaviour affects which roles people assign to them.

A study in which robots were described as either equipment or coworkers found that technology use self-efficacy and prior robot use experience were associated with more positive attitudes towards both robot positions [56]. One study [57] explored the roles that a child (13.67 ± 0.71 years) assigned to an educational robot after one or several interpersonal interactions with the same robot. The participants interacted with the robot (torso of a Nao [58]) multiple times for 30 minutes each time. During an interaction, they played a serious game about sustainable development with the goal of collaboratively creating a sustainable city. They found that the participants perceived the robot not only as a tutor but also as a classmate or a friend. After multiple encounters, the participants saw the robot as less of a tutor and more as a classmate. This study showed that participants can assign multiple roles to robots.

Another study [59] examined how coworkers perceive a robot that served the job of receptionist for a month. They introduced a humanoid robot into a collaborative social workplace. The humanoid's primary task was to function as a receptionist and provide general assistance to the customers. After a month, they asked the coworkers to choose one use of a social robot in a workplace from the following: customer support, receptionist, public relations, not sure and other. They found that most people assign the robot the job of receptionist (54%) rather than customer support (31%), public relations (7%) or concierge (8%). This showed that people can imagine robots in multiple roles.

3 Research questions and hypotheses

The overarching research question in this study was whether people interacting in a sequential decision-making process prefer fallacious robots over logical robots. In other words, do people agree with, become influenced by and select robots that repeatedly perform judgemental fallacies over those that do not?

Based on [31,37], people should select *non-fallacious robots* more, whereas based on the "like-me" hypothesis [20,26] people should select *fallacious robots* more. Hence, we wanted to address the context dependency of this preference, meaning, do people's preferences for fallacious robots depend on the situation in which they encounter the robot? Does it depend on the context of the decisions in which the robots make judgmental fallacies?

Furthermore, we wanted to study whether people assign different jobs to each robot. We selected the jobs of barman [10], psychologist [11], analyst [13], jury member [12], investment banker [60] and caregiver [5]. These jobs were selected to be representative of different requirements, e.g. more emotional (barman, psychologist, caregiver) and more analytical (analyst, investment banker). In both context and job selection, we hypothesize that *logical-based* scenarios and jobs will promote selection of *non-fallacious, or rational* robots, whereas *emotional-based* scenarios and jobs will promote selection of *fallacious, or irrational* robots.

Moreover, we wanted to assess whether the robots' fallaciousness affects how they are perceived by the participants. Based on [46,50], we hypothesize that the fallacious robot will be more likable.

Finally, the dynamic nature of the interaction was of importance, namely, we did not want to contend with a single decision on the part of the robots and the participants.

We wanted to explore how repeated interactions with these robots influence people's decision making.

To answer these research questions, we designed a dynamic interaction with two robots, one that makes fallacious decisions and one that does not. The interaction revolved around repeated decisions concerning the following two scenarios (see Sec. 4.5.2). One scenario (detective) revolved around a crime committed and evidence presented, wherein the decisions to be made were with respect to which suspect/s are more likely to have committed the crime. The detective scenario was based on evidence and was thus considered to be more intellectual and logical. A second scenario (art auction) revolved around art pieces of various types and artists, wherein the decisions to be made were with respect to which art piece was worth more. The art scenario was based on individual appreciation of art and artists and was thus considered to be more emotional and subjective.

Our main pre-registered hypothesis can be found in [61], whereas a full list of the hypotheses, tests and results can be found in [62]. We list the most-noteworthy hypotheses related to our central research questions below:

- Participants' agreement with robots will be context-dependent: they will agree with fallacious robots more often than they will agree with logical robots in *emotional* contexts (H1.1) [20,26] and vice-versa in *logical* contexts (H1.2) [31,37].
- Participants' job-assignments to robots will be job-dependent: they will assign emotional-social jobs to fallacious robots more often than to logical robots (H2.1) [20,26] and vice-versa for logical jobs (H2.2) [31,37].
- Participants will change their decisions based on the robots' decisions (H3.1) [32,33]. Participants will change their opinion towards the robot they have agreed with (H3.2).
- Participants will perceive the non-fallacious robot as more intelligent (H4.1) [47], but less likable (H4.2) [46,50] than the fallacious robot.

4 Methods

Consistent with recent proposals [63,64], we report how we determined our sample size, all data exclusions, all manipulations and all measures in the study. In addition, following open science initiatives (e.g. [65]), the study was pre-registered in AsPredicted.com [61], wherein the deidentified data sets, full set of hypotheses and analyses associated with this study are freely available online in [62]. By making the data available, we enable and encourage others to pursue tests of alternative hypotheses, as well as more exploratory analyses.

4.1 Participants

One of the base effects of robots' behaviour in our context is how they are perceived [46–50,58,59]. The most relevant study with this regard is Ref. [50]. Based on this study [50], the difference in Godspeed's likeability scale for the two conditions of the robots is 0.37 with a standard deviation of 0.63. Mann–Whitney U is used to compare the Godspeed ranks. Taking a confidence level of 95% and a power of 80%, the required sample size is 52 participants.

The study was conducted with 55 participants from Israel. They were recruited from social networks and university flyers. Five participants were excluded from the analysis due to technical difficulties with the robots. Twenty-three of the participants were female, and the mean age was 27.74 ± 8 years.

All participants signed consent forms, and the study was approved by our institutional Internal Review Board.

4.2 Experimental design and stimuli

A laboratory experiment consisting of a within-subjects 2-factor experimental design with two problem-solving scenarios as treatments (art auction vs detective task) was conducted.

The colours of the two NAO robots were different, i.e. one was red and one was blue; one robot was on the left, and the other was on the right, Fig. 1. The robots had two different and opposite behaviours. One repeatedly made a conjunction fallacy, and the other did not. In other words, the probabilities communicated by the robots were determined by whether they were fallacious, e.g. $p_A < p_{A \cap B} < p_B$, or not, e.g. $p_{A \cap B} < p_A < p_B$.

In a randomized order, all participants completed the two scenarios with the two NAO robots (see App. A, B). One scenario was about a detective task, and the other was about an art auction. The problem-solving approach in the first scenario (detective task, App. A) is more analytical, as it is based on clues and facts regarding the incident. The second scenario (art auction, App. B) is more latent and abstract, as it is concerned with the subjective appreciation of the value of art.

The location, colour, and behaviour of the robots, as well as the order of the scenarios, were randomized. Therefore, there are eight possible combinations in total, i.e. two scenarios, two sides and two colours of the robots. Each participant encountered a single combination. For example, a possible combination involved having the red robot on the right, behaving logically, and the first scenario was the art auction.

Given this design, the number of participants in each experimental combination is as follows: Auction ($N = 23$) or Detective ($N = 27$) as the first story; Blue ($N = 22$) or

Fig. 1 Experimental setup. One robot is located on the right, and the other robot is on the left. In the middle, there is an elevated laptop, and the screen is at the same height as the robot's shoulders

Red ($N = 28$) robot on the right; Blue ($N = 25$) or Red ($N = 24$) robot as logical.

4.3 Measurements

During the study, we used measurements in three sections (see Fig. 2).

Prior to the interaction with the robots, the participants completed the following: (i) demographic questionnaire: age, gender, education; (ii) big five inventory (BFI) for personality rates: extroversion, agreeableness, conscientiousness, neuroticism and openness; (iii) negative attitudes towards robots (NARS) [66], which asks about negative attitudes towards situations and interactions, social influence of robots, emotions in interaction with robots.

During the interaction with the robots, the participants were asked the following two types of questions: (i) *choose* between the two robots (Q3, Q6, Q10) and (ii) *rate probabilities* by entering a number between $0\% - 100\%$ (Q1, Q4, Q7). This is repeated two times, once for each of the two scenarios.

After the interaction with the robots, participants were asked several types of questions, as follows: (i) a comparative Godspeed questionnaire [51] that consists of five scales that are relevant to evaluating the perception of (social) human–robot interaction: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. In our setup, the participants rated both robots simultaneously on each Godspeed item, i.e. they had to rate each robot on each item, wherein the items appear simultaneously for both robots; (ii) *choose* one of the two robots for different jobs, general preference, and (hypothetically) taking it home; and (iii) *open question* on any difference between the robots.

4.4 Analysis methods

We have used various analysis methods in our study (full hypotheses list and their analysis can be found in [62]). For count comparisons to random choice in each scenario, we have used a one-sided Binomial test. For the analysis of the number of agreements for each robot in each scenario, we have used the one-tailed t-test.

Correlations between robot's answers and the participant's answers were analysed using the Pearson correlation test. For job assignments, we have used a multi-linear regression analysis, wherein the personality traits and attitude towards robots were the predictors of the assignment of the robots to a specific job.

4.5 Procedure

4.5.1 Experimental setup

The system consisted of two NAO robots and a laptop in the middle. The participants completed the questionnaires using an additional computer that was placed near the system (Fig. 1). A participant entered the experiment room and sat down. First, the participant was given instructions to pay attention to what the robots said and to answer all the questions. Then, the participant sat in front of the computer to complete the questionnaires (BFI, NARS). After finishing them, the robots were revealed, and the interaction began. There was an app on the laptop that presented the story (textually and verbally). In each advance in the interaction (Fig. 2:right), a new screen appeared in the app. New information was written on the screen and verbally conveyed, using prerecorded speech, to the participant. Then, depending on the question, the app presented the participant the option to rate probabilities, choose a robot or rank probabilities by order. The robots talked and moved, one after the other, and the order in which the robots talked was assigned randomly by the app. After the interaction with the robots was finished, the participant was asked to complete the remaining questions (Fig. 2: left) on the computer.

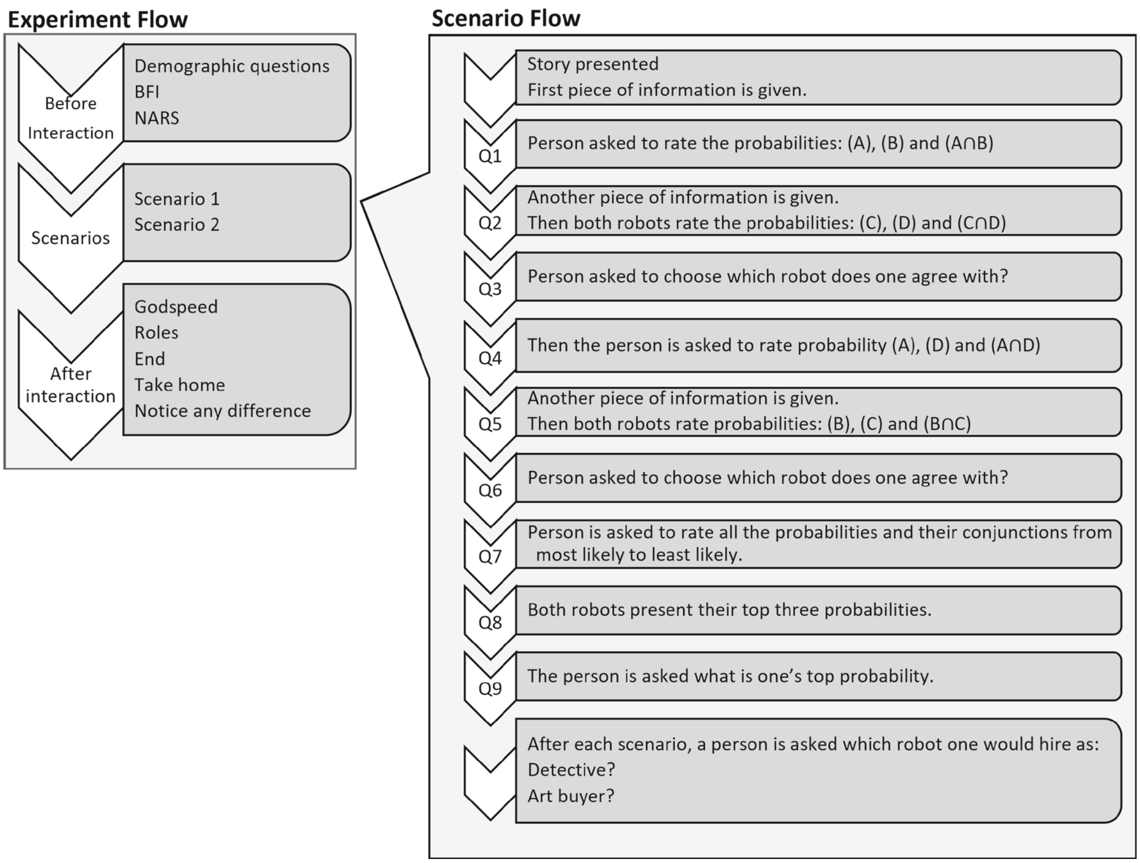

Fig. 2 The flow of the experiment. Left: the full flow of the experiment. Right: full flow of one interaction (scenario) out of two with the robots. The full scenarios are given in the appendices A and B

4.5.2 Study flow

The full experimental flow is shown in Fig. 2. Before the interaction with the robots began, each participant was given two personality-related questionnaires, namely, BFI and NARS.

Following these questionnaires, the robots were revealed to the participants, and one of the following two scenarios (see appendices A, B) was introduced: a scenario about a detective task and another about an art auction. During the interaction, the participant took turns with the robots. First, the participant had to make a judgmental decision (Q1) by rating the probability of three options, where each option was rated on a slider from 0%–100%, with a 10% increment. Then, after receiving more information about the scenario, each of the two robots communicated its own judgmental decision (Q2) by speaking their assigned probabilities for each option. The fallacious robot gave the probabilities of the three options in a fallacious order. For example, in the Detective Scenario, the fallacious robot said: "The probability that suspect C is the robber is 90%, the probability that suspect D is the robber is 50%, and the probability that suspects C and D committed the robberies together is 70%". The non-fallacious robot gave probabilities that represented a rational order. For example, "The probability that suspect C is the robber is 90%, the probability that suspect D is the robber is 70%, and the probability that suspects C and D committed the robberies together is 50%". The robots' rating was followed by the following question given to the participant: "Which robot do you agree with?" (Q3).

This sequence of participant rating, robot rating and participant agreement was repeated for a different combination of items to rate (Q4-6). Then, the participant was asked to rate all probabilities of items and their conjunctions (Q7). Both robots communicated their top three items (Q8), followed by the participants' report of their final decision (Q9). The participant was then asked to choose which robot they would hire as a detective or art buyer.

The entire procedure was repeated for the second scenario, where the order of the two scenarios was randomized across participants.

To summarize, for each scenario, a participant provided a rating twice, reported agreement with one of the robots twice, ranked all probabilities, chose the top ranked probability, and assigned two jobs. The robots presented their opinion three times: provided two ratings and presented the top three ranked probabilities.

Table 1 Participants' fallacy rates in both scenarios in the question in which they were asked to rate probabilities

Scenario	Q1	Q4
Art	0.82	0.72
Detective	0.66	0.66

After the interaction, we asked the participants which robot they preferred. First, with a comparative Godspeed questionnaire, we asked the participants to choose one of the robots for each of the following six jobs: barman, psychologist, analyst, jury member, investment banker and caregiver. These jobs were selected to be representative of different requirements, e.g. more emotional (barman, psychologist, caregiver), more analytical (analyst, investment banker), more social (psychologist, caregiver), etc.

Ultimately, to obtain a general notion of preference, we asked the participants the following questions: "Which robot would you take home? Why?" and "Did you notice a difference between the robots? What was it?"

5 Results

5.1 Participants' fallacies

In questions Q1 and Q4 (Fig. 2), participants were asked to rank the following three probabilities: the probabilities of the two options and their conjunction. In all four questions, more than 60% of the participants made the conjunction fallacy (Table 1). Note that in the art scenario, more people made fallacious decisions than in the detective scenario, albeit not significantly so ($\chi^2(1, N = 100) = 2.969$, $p = 0.085$). These fallacy rates decreased in the range reported in previous studies on conjunction fallacy [27,29,30].

5.2 Participants agreed more with fallacious robots but only in emotional contexts

The participants were asked if they agreed with the robot twice (Q3, Q5, Fig. 2) in each scenario, i.e. four times in total. The analysis shows that participants agreed with the fallacious robot 57.5% of the choices (p=0.04, two-tailed Binomial test). In a second analysis, the number of times each participant agreed with the fallacious robots ranged from 0 to 4 (normalized to [0, 1]). A one-sample t-test was performed to check if these choices differed from random choice. The one sample t-test results entail that the difference between participants' agreement with the fallacious robot and a random decision was not significant. Nevertheless, we observed a positive trend in the sample ($\mu = 0.58 \pm 0.29, t_{49} = 1.82$, $p = 0.074$, $d = .26$).

We continued to test H1.1 and H1.2 and performed the same analysis for the art and detective scenarios, respectively. In each scenario, participants made two choices. The analysis shows that in the art scenario participants agreed with the fallacious robot 61% of the choices ($p = 0.0178$, one-tailed Binomial test) supporting H1.1., whereas in the detective scenario they agreed with the fallacious robot 54% of the choices ($p = 0.81$, one-tailed Binomial) not supporting H1.2. In a second analysis, the number of times each participant agreed with the fallacious robots ranged from 0 to 2 (normalized to [0, 1]), Fig. 3. A t-test was performed to check if these choices differed from a random choice. The participants were more inclined to agree with the fallacious robot, although the effect was only significant for the art scenario ($\mu = 0.61 \pm 0.38$, $t_{49} = 2.037$, $p = 0.047$, $d = .29$). These results support H1.1. In the detective scenario, the selections did not differ from random selections ($\mu = 0.54 \pm 0.39$, $t_{49} = 0.727$, $p = 0.471$, $d = .1$) and do not support H1.2.

5.3 Participants changed their decision based on the robot they agreed with

To test the hypothesis that the participants were influenced by the robots' communicated decisions, we performed a Pearson correlation test on $P(D)$ between the participants and the robots' answers, where the robots were labelled by their behaviour, i.e. logical or fallacious. The participants rated this probability for the first time in Q4 after hearing the robots' rating in Q2. Our analysis shows that the robots' decisions were significantly correlated with the participants' decisions only in the detective scenario between the logical robot and the participants' answer (Detective, logical robot: $r = 0.33$, $p = 0.02$, fallacious robot: $r = 0.21$, $p = 0.14$. Art, logical robot: $r = 0.17$, $p = 0.23$, fallacious robot: $r = 0.09$, $p = 0.55$). These results partly support H2.1.

We continued to test whether these effects were dependent on the robot that the participants agreed with. To test this hypothesis, we performed a Pearson correlation test on $P(D)$ between the participants and the robots' answers, now labelled by which robot they chose or agreed with in Q3, Table 2. Our analysis shows that the chosen robot's decisions were significantly correlated with the participants' decisions, whereas the other robot's decisions were not (Detective scenario: $r = 0.57, p < 0.0001$, art scenario: $r = 0.45, p < 0.001$). These results support H2.2.

Another analysis shows that participants changed the probability $P(A)$ provided in question Q4 from that previously provided in question Q1 after hearing the robots in Q2, which did not rate that probability (Table 3).

Questions Q7-Q9 were ranking questions. First, the participants ranked all the options in question Q7. Then, the robots presented their top three probabilities in question Q8. Finally, in question Q9, the participant was asked to choose the most likely option for each scenario. In the art (detective) scenario, 68% (80%) of the participants changed their

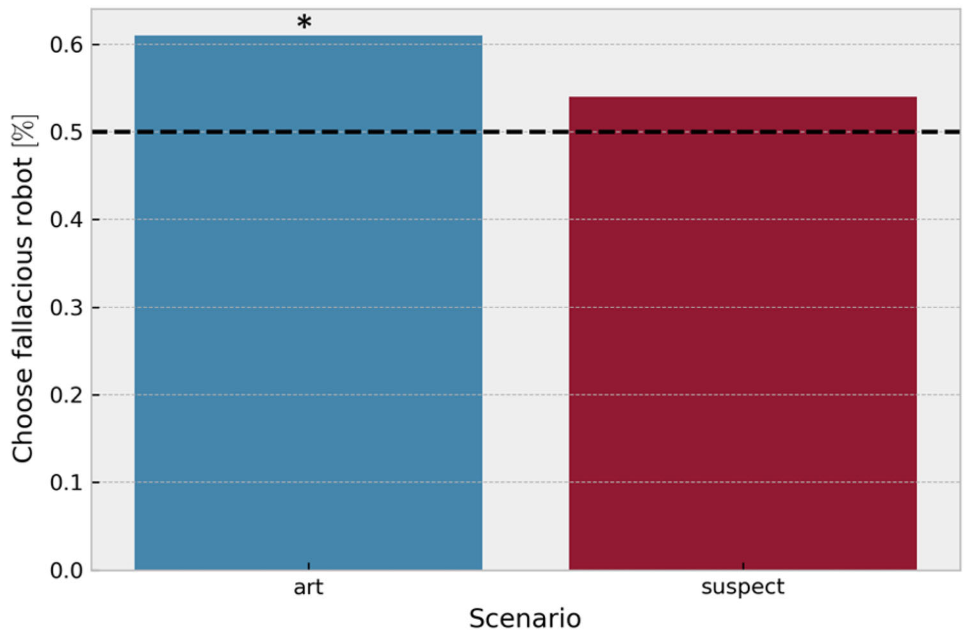

Fig. 3 Percentage of times that the fallacious robot was chosen in each scenario (the dashed line indicates a random selection, 50%). * ($p = 0.047$)

Table 2 Correlations between the probability a participant gave to context D in question Q4 and the same probability the robots gave in question Q2. $***p < 0.001$

Robots' probability (Q2)		Chosen	Not-chosen
Person probability (Q4)	Detective	0.57***	−0.09
	Art	0.45***	−0.19

Table 3 Difference between the probability $P(A)$ participants provided in question Q1 and then again in question Q4. $***p < 0.0001$

	Difference	t-test
Detective	-0.11 ± 0.21	−3.74***
Art	0.19 ± 0.27	5.17***

top ranking. Out of these participants, 50% (64%) did not change to one of the robots' top rankings, and 26% (25%) chose the same option as the fallacious robot chose. In the art scenario, 18% (12%) of the participants changed their selection to match the one chosen by the robot they chose in question Q6 (Q3). In the detective scenario, 25% (25%) of the participants changed their selection to match that of the robot they chose in question Q6 (Q3).

Taken together, these results suggest that the participants were influenced by the decisions expressed by the robots, especially those of the robot they agreed with last in the detective scenario.

5.4 Exploratory analysis: attitudes and personality affect robot job assignment

We have conducted an exploratory analysis with the goal of investigating how attitudes towards robots and personality traits may affect the selection of specific robot advisors to different jobs. For this purpose, we performed a Binomial test on each of the jobs, to analyse if the fallacious robot was selected more than the non-fallacious. Our analysis shows that no significant difference was found (Investment $p = .66$, Analyst $p = .66$, Jury $p = .44$, Barman $p = .76$, Psychologist $p = .88$, Caregiver $p = .33$).

We continued to test whether participants preferred a specific robot to take home. We found that there is no significant preference towards one of the robots over the other ($p = .66$).

We also found that participants could not consciously detect the difference between the robots ($p = .23$). At the end of the questionnaire, we asked the participants to rate both robots on the Godspeed scale. For all subscales, no significant difference was found between the robots.

However, to investigate whether personality traits or attitudes towards robots affect robot job assignment, a multi-linear regression was performed. The dependent variable was calculated as the (normalized) number of jobs participants assigned to the fallacious robots out of all six jobs. Thus, assignment of the fallacious robot to all 6 jobs was coded as 1, and assignment of the logical robot to all 6 jobs was coded as 0. The independent variables were the personality traits from the BFI and NARS. A significant regression equation was found ($F_{8,41} = 3.107$, $p = 0.008$) with $R^2 = 0.377$. The participants' predicted job assignment to the fallacious robot is equal to $1.26 + 0.13 * Agreeableness - 0.14 *$

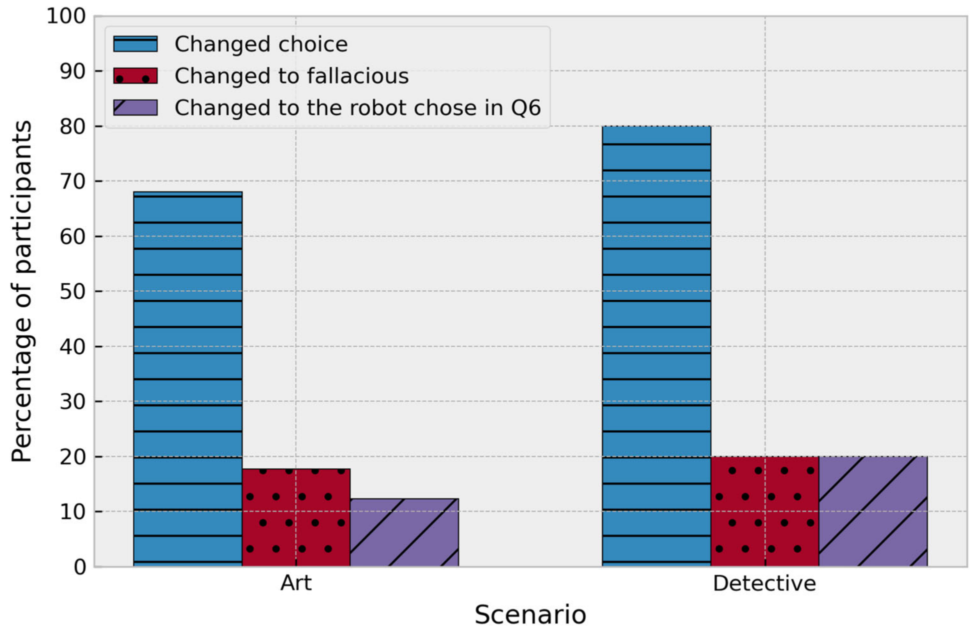

Fig. 4 Participants that changed their choice from question Q7 to question Q9 for both scenarios. For each scenario: (Left bar) Participants who changed their top ranking; (Middle bar): participants who changed their top ranking to that of the fallacious robot; (Right bar) participants who changed their top ranking to that of the robot they last chose (in question Q6)

$Neuroticism$. The job assignment to the fallacious robot increases as participants are rated as more agreeable and decreases as participants are rated as more neurotic. Both agreeableness and neuroticism were significant predictors of job assignment to the fallacious robot.

To further study these effects, we conducted a logistic regression to each of the six jobs plus the two jobs after each scenario Q10, with Bonferroni correction for multiple comparisons. We found that only the detective job was significantly predicted but with context (art/detective) dependence. For the art scenario, the detective assignment was affected by the negative emotions towards robots, $Detective(art) = -3.14 \cdot Emotions$, $p = 0.005$. That is, participants who had more negative emotions towards robots assigned the detective job to the robot. For the detective scenario, the detective assignment was affected by the participants' personality traits (BFI), $Detective(detective) = 3.63 \cdot Agreeableness$ $-2.47 \cdot Conscientiousness$ $-2.82 \cdot Neuroticism$, $p = 0.0007$. That is, participants who were rated as more agreeable preferred to assign the fallacious robot to the detective job. In addition, if a participant was rated as more conscientious and neurotic, he or she preferred to assign the logic robot to the detective job.

Taken together, this exploratory analysis suggests that participants' personality traits, most notably agreeableness and neuroticism, predicted their job assignment.

5.5 Open question analysis

At the end of the interaction, we asked the participants which of the robots they would take home and why. While there was no significant difference in preference between the robots, their answers were very informative. We subdivided the explanations into five attributes that the participants described, as follows (Fig. 5):

- Colour: "I prefer the colour blue", "blue is calmer", or "red is relaxing".
- Movement and speech of the robot: "move smoother", "talk clearer", or "higher language".
- Think like the participant: "closer to me", or "compatible with the way of thinking".
- Its logic: "logical", 'rational", or "analytical".
- Holistic characteristics: "emotional, conversationalist", "cute", or "not anxious".

Although there is no significant difference between people who preferred the fallacious robot to those who preferred the logical one in their explanation, a trend is visible. Interestingly, participants that preferred to take the logical robot home, based their decision on holistic properties of the robot (p=0.44 two-tailed Binomial test) and its movement and speech (p=0.72 two-tailed Binomial test) more than the participants that chose the fallacious robot. In addition, participants that chose the fallacious robot explained their decision by the robot's logic more than participants that chose the logical robot (p=0.44 two-tailed Binomial test).

Intriguing, upon examining the subscales that emerged from the open questions (Figure 5) again, similarities between the subscale and Godspeed's subscales were found, i.e. anthropomorphism vs. holistic, animacy vs. colour, movement and speech, likeability vs. thinks like me and perceived intelligence vs. logic. A study by Deshmukh and colleagues [67] found that the understandability of robots'

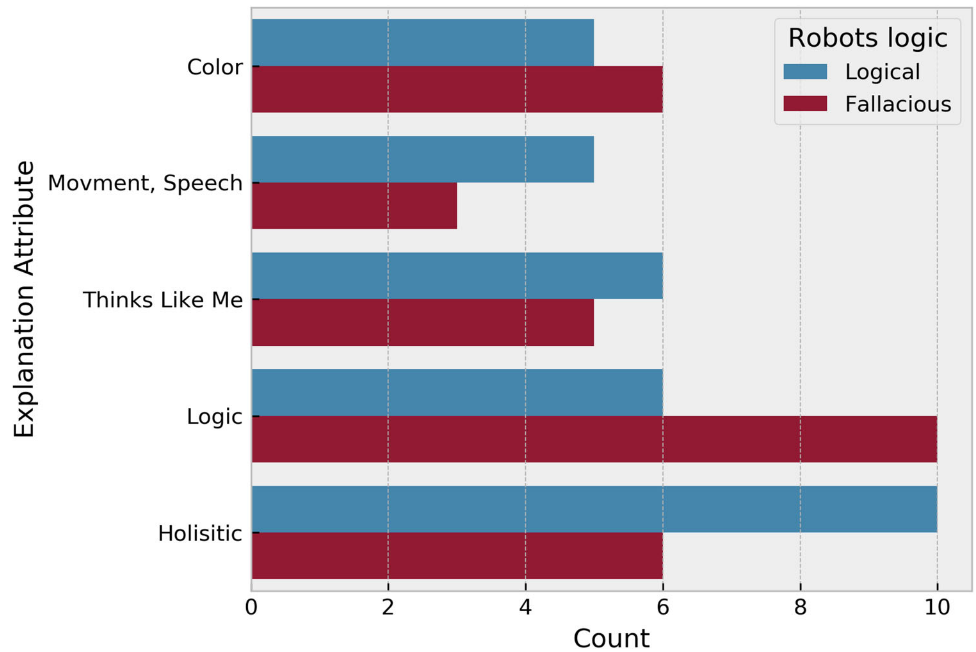

Fig. 5 The reasons participants took robots home, categorized according to the attributes of the explanation

gestures was correlated with Godspeed. The results of the study suggest a positive correlation between anthropomorphism, altruism and perceived intelligence but a negative correlation between likeability and perceived safety. These findings support our findings that participants were able to justify which robot they preferred to take home based on how they understood the robots' behaviour.

6 Discussion

This study investigated how robot advisor's logic (fallacious or not) affected participants' decision making in two different scenarios (art and detection stories). Prior research regarding fallacious robot advisors was inconsistent, namely, the "like-me' hypothesis [19,20] and homophily phenomenon [23] predict that people will prefer robots that behave like themselves, i.e. in a fallacious manner, whereas social decision making [31,37] has been shown to mitigate selection of fallacious options. Hence, our main research question was whether the context affects the preference, or agreement with fallacious robot advisors, suggested by the matching hypothesis [24,25]. We thus hypothesized that in the art scenario, which involves more emotional and less rational aspects, people will agree more with the fallacious robot, whereas in the detective scenario, which is based on evidence and requires analytical thinking, people will agree more with the logical robot advisor.

To answer these questions, we designed a novel experimental design wherein participants and two NAO robots, one fallacious and one logical, made repeated decisions in the two aforementioned scenarios. During the interactions, the robots presented multiple answers and the participants had to decide which robot they agreed with. After the interactions, the participant had to rank the robots and assign them to different jobs that differ by their skills (emotional/rational). To the best of our knowledge, this is the first time social robots with cognitive fallacies were used to help participants make decisions.

We found that (i) participants were more likely to agree with the fallacious robots in the emotional context; (ii) participants altered their opinions relating to the robots they agreed with, regardless of its logic; and (iii) while there was no preference in the jobs assigned to the robots, personality traits affected the job assignment, wherein participants who were more agreeable and less neurotic opted to assign the fallacious robot to these jobs. Interestingly, in the art auction scenario, participants who had more negative emotions towards robots assigned the fallacious robot to the detective job, while in the detective context, the personality traits of the participants affected the detective assignment. In the end, we investigated why participants chose to (hypothetically) take home one robot and found that there were five attributes that participants based their choice on. The main reasons included logic for the fallacious robots and holistic characteristics for the logical robot.

The presented study attempts to represent real-life decision making processes by being *dynamic*, include *real robots in two different contexts* and involve real people with *attitudes and personalities*. We henceforth discuss these unique attributes and the insights our study presented.

6.1 Participants changed their decisions based on the robot they agreed with

The participants provided ratings similar (with higher correlation) to those of the robot they agreed with (H2.2), whether or not it was the logical robot (H2.1). This means that the perceived logic of the robots by the participants (agree/disagree with) affects the participants' ratings more than the actual robot logic (fallacious/ logical). Additionally, in the last question in each scenario, the participants changed their top ranking after hearing the rankings of the robots. However, they did not change it to the robot they agreed more with or to the last robot they agreed with.

The participants made decisions that correlated with the robot they agreed with. This may be an example of confirmation bias [68], such that the participants confirmed their previous agreement with the robots by repeating their answers again.

However, another explanation for this is that they made this decision based on the information given and not based on the robot and then agreed with the robot that complied with their decision. This explanation does not explain the fact that participants changed their ranking after hearing the ranking provided by the robot, since there was no new information in between.

The decoy effect [69] may explain why participants changed their opinion from the first time they were asked to rank. Their own opinion may have been the dominant preference, whereas the robot they agreed with more's answer might have been a decoy, i.e. the asymmetrically dominated alternative.

In question Q4 (Fig. 2), participants were asked again about probability of concept A. Between question Q1 to Q4, there was no additional information about concept A yet participants changed their answer to this option (Table 3). One model that can explain this result is the quantum model [70], wherein decision making is modelled as a quantum system, with quantum probability. This model tries to explain several interesting phenomena via quantum effects, such as interference and collapse of the wave function. One of these effects is quantum entanglement, wherein if two concepts are quantum entangled, measuring or asking about one concept changes the probability of the other. In our context, this model may explain the change in concept A, even though no information was added.

6.2 Participants agree with specific robots, depending on the context

Each participant chose which robot he or she agreed with four times in total, i.e. twice in each scenario. In disagreement with our H1.1 hypothesis, the participants did not agree more with the fallacious robot (or the logical robot). Interestingly, the result was different for each separate scenario. Under the art context, the participants agreed significantly more with the fallacious robot.

Polakow et al. [37] and Charness et al. [31] showed that participants are less likely to make fallacious decisions after hearing the opinions of other agents (humans in [31] and social robots in [37]). Our results show that this is not always the case, as in the current study, the participants heard the robot's opinions on other concepts. In other words, they never had to make a ranking decision on exactly the same concepts the robots talked about. This shows that the mere presence of other agents does not suffice to reduce fallacy rates, but rather the presentation of options by the other agents upon which the participant needs to make a decision.

Furthermore, we deliberately chose two different scenarios, i.e. one that was more emotional (art) and one that was more logical (detective), to investigate the connection between the context, i.e. scenario, and participants' agreement with different social robot advisors. This is in line with the study, showing that analytical priming did not succeed in reducing the endorsement of irrational gambling beliefs [54,71]. However, in our art scenario, there was emotional priming, which we found to increase the endorsement of fallacious robots. Our results show that social robot advice is complex and context-dependent. There is no single robot advisor that participants agree with. In different contexts, participants agreed with different robots.

6.3 The effects of personality traits and attitudes towards robots on robot job assignment

In total, there were six jobs to assign to the robots. The participants did not assign the jobs more to the fallacious or logical robots, and they did not have a preference regarding which robot to take home. The participants' personalities affected which robot they preferred for the jobs, where participants who were more agreeable assigned more jobs to the fallacious robot, whereas neurotic participants assigned more jobs to the logical robot.

These results contradict a previous study [72], where, using regression analysis, it was found that high scores of neuroticism were associated with high irrationality. However, another study [73] found that participants who were more agreeable had a higher degree of irrationality, whereas it was also found that agreeableness is negatively related to irrational beliefs [74]. While these results might be contrary to our result, these papers examined irrational beliefs of people while our study investigated job assignments for fallacious agents.

To study a specific job assignment, a logistic regression was performed. The detective job was found to be significantly predicted by personality traits and attitudes, but it was context dependent. In the art context, the NARS affected the detective job assignment; i.e. participants who had more negative emotions towards interactions with robots preferred to assign the fallacious robot to the detective job. However, in the detective context, the BFI affected the detective job assignment; i.e. the participants who were rated as more agreeable preferred to assign the fallacious robot to the detective job. On the other hand, the participants who were rated as more conscientious and neurotic preferred to assign the logical robot to the detective job.

These results suggest that the job assignment is highly dependent on the person's personality and that context mediates the personality traits that affect job assignment.

6.4 Explicit and implicit perception of robot fallacies

We also studied whether participants noticed the differences in the robots. We found that they did not consciously (Godspeed, choices) or explicitly (based on the answer to the question regarding if they noticed a difference between the robots) perceive a difference between the robots. This suggests that the difference in response, e.g. preference for the fallacious robot in the art auction context, was subconscious in nature [75,76]. Future studies can delve deeper into this subconscious preference, via neurophysiological markers, e.g. eye tracking and EEG [77,78].

At the end of the questionnaire, we asked participants the following question: "Which robot would you take home and why?" There was no preference for one robot over the other (48%/52%). Nonetheless, the participants justified their choices based on the robots' behaviour and logic. The most influential factors were the mental characteristics of the robot. Curiously, the participants who chose the logical robot justified their choice based on holistic characteristics. In contrast, the participants who chose the fallacious robot rationalized this choice based on the logical characteristics of the robots. On the other hand, no differences between the robots (and take home choices) were found.

7 Conclusions

This study utilized two robots as advisors to participants for decision-making in different contexts. Participants' decisions were affected by the robot advisors. Additionally, our results suggest that the context (scenario) affected the participants' perception of the robots. Interestingly, participants' personality traits and attitudes towards robots also had a non-trivial effect on their job assignment for the robots.

Taken together, our study suggests operational recommendations for the design of social robot advisors. The scenario in which the robots provide advice should be taken into consideration when selecting both their logic and their assigned jobs. One should not use a one-size-fits-all design, as we have shown that people are influenced by these different aspects. Furthermore, knowing the users' personalities can help in designing the proper advisor logic, as we have shown that people's neuroticism and agreeableness traits may affect their perception of such advisors. Finally, we have shown that people are influenced by social robotic advisors; hence, their usage in more general settings merits further investigation.

In the future, a better understanding of the different contexts and the robots' logic is needed. Additional contexts should be added to explore a wider range of associations. Moreover, the participants should also be asked how they perceive the context. Finally, studying other types of judgmental fallacies and their effects on robot perceptions is another interesting research direction.

To conclude, robots' judgemental fallacies and human participants' decision-making and perceptions follow a complex interaction mediated by context, personality and attitudes. Studying this interaction is an important task for future human–robot decision-making collaboration.

Acknowledgements This study was supported by the Israel-US Binational Science Foundation No. 2016262.

Appendix A Detective scenario

1. A diamond shop was robbed. The police came straight away and caught a couple of people (separately) near the scene, but no diamonds were found. The detective assigned to the case is sure, with no doubt, that at least one of the suspects robbed the shop. You and the robots need to help the detective figure out who robbed the shop. The suspects are as follows:
   (a) Suspect - A: Is tall and is wearing a black Louis Vuitton leather jacket and a Rolex watch.
   (b) Suspect - B: Has blonde hair and had a cut above his or her left eyebrow.
2. Person rate A, B, A and B.
3. New information:
   (a) Suspect - C: Tried to run from the scene when the police asked to stop.
   (b) Suspect - D: Is not willing to talk without a lawyer present.
4. Robots rate C, D, C and D:
5. Ask the person which robot he or she agrees with?
6. Person: after hearing the person the robots chose, rate A and C:
7. New information: Suspect B told the police that she got the cut when from a tree branch when she took

her dog for a walk a few hours prior to the incident. Suspect D is still refusing to talk even after the lawyer arrived.

8. Robots: give their ratings for B and D.
9. Ask the person which robot he or she agrees with?
10. Person: rank all the possibilities.
11. Robots: also provides rankings of the top three possibilities.
12. The detective asks, “Who did it?” (to see if the robots’ rankings affected the person’s rankings)

Appendix B Art scenario

1. In the last couple of years, the market for fine art has been booming. Last year, the most expensive piece of artwork that was sold was an expressionist oil painting by a late famous artist. Additionally, last year, one item that caught the most attention around the world was made by a famous young graffiti artist. Tonight, you and the two robots went to an art auction together. There were expressionist paintings to realistic sculptures of horses and even a few pieces of graffiti artwork. The auction was a great success, though not all the pieces were sold. Person: What piece do you think was sold for the highest price? Rate the options:
   (a) A realistic piece. (A)
   (b) A piece from a famous artist. (B)
   (c) A realistic piece made by a famous artist. (A∩B)
2. New information: One piece received a great deal of attention from the young investors in the audience.
3. Robots: Which piece do you think it was?
   (a) An expressionist piece. (C)
   (b) A piece from a young artist. (D)
   (c) An expressionist piece made by a young artist. (C∩D)
4. Ask the person which robot he or she agrees with?
5. Person: After hearing the robots, person rate A and D:
6. New information: the most expensive artwork in the auction contained a figure of a person, but it was not clear what was it made of.
7. Robots: Give their ratings for A and D.
8. Ask the person which robot he or she agrees with?
9. Person: rank all the possibilities.
10. Robots also provide rankings of the top three possibilities.
11. The person is being asked, “What was the most expensive artwork that was sold?” (to see if the robots’ rankings affected the person’s rankings)
12. Which robot would you hire as an art buyer?

References

1. Breazeal C, Scassellati B (1999) How to build robots that make friends and influence people. In: 1999 Proceedings IEEE/RSJ international conference on intelligent robots and systems 1999. IROS ’99, vol. 2, pp. 858–8632. https://doi.org/10.1109/IROS.1999.812787
2. Gordon G, Breazeal C, Engel S (2015) Can children catch curiosity from a social robot? In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction. HRI ’15, pp. 91–98. ACM, New York, NY, USA. https://doi.org/10.1145/2696454.2696469
3. Short E, Swift-Spong K, Greczek J, Ramachandran A, Litoiu A, Grigore EC, Feil-Seifer D, Shuster S, Lee JJ, Huang S, Levonisova S, Litz S, Li J, Ragusa G, Spruijt-Metz D, Mataric M, Scassellati B (2014) How to train your DragonBot: Socially assistive robots for teaching children about nutrition through play. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp. 924–929. IEEE, Edinburgh, UK. https://doi.org/10.1109/ROMAN.2014.6926371
4. Henschel A, Laban G, Cross ES (2021) What makes a robot social? A review of social robots from science fiction to a home or hospital near you. Curr Robot Rep 2(1):9–19. https://doi.org/10.1007/s43154-020-00035-0
5. Fasola J, Mataric M (2013) A socially assistive robot exercise coach for the elderly. J Human-Robot Interact 2(2):3–32. https://doi.org/10.5898/JHRI.2.2.Fasola
6. Robinson NL, Cottier TV, Kavanagh DJ (2019) Psychosocial health interventions by social robots: systematic review of randomized controlled trials. J Med Internet Res 21(5):13203. https://doi.org/10.2196/13203
7. Feingold Polak, R, Tzedek SL (2020) Social robot for rehabilitation: expert clinicians and post-stroke patients’ evaluation following a long-term intervention. In: Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction. HRI ’20, pp. 151–160. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3319502.3374797
8. Mohebbi A (2020) Human-robot interaction in rehabilitation and assistance: a review. Curr Robot Rep 1(3):131–144. https://doi.org/10.1007/s43154-020-00015-4
9. Stower R, Calvo-Barajas N, Castellano G, Kappas A (2021) A meta-analysis on children’s trust in social robots. Int J Soc Robot. https://doi.org/10.1007/s12369-020-00736-8
10. John NE, Rossi A, Rossi S (2022) Personalized human-robot interaction with a robot bartender. In: Adjunct proceedings of the 30th ACM conference on user modeling, adaptation and personalization. UMAP ’22 Adjunct, pp. 155–159. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3511047.3537686. Accessed 2022-07-05

11. Riches S, Azevedo L, Vora A, Kaleva I, Taylor L, Guan P, Jeyarajaguru P, McIntosh H, Petrou C, Pisani S, Hammond N (2022) Therapeutic engagement in robot-assisted psychological interventions: a systematic review. Clin Psychol Psychother 29(3):857–873. https://doi.org/10.1002/cpp.2696
12. Hayashi Y, Wakabayashi K (2020) Experimental investigation on the influence of prior knowledge of a decision-support robot for court juries. In: Companion of the 2020 ACM/IEEE international conference on human-robot interaction. HRI '20, pp. 236–238. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3371382.3378238. Accessed 2022-07-05
13. Kobets V, Yatsenko V, Mazur A, Zubrii M (2018) Data analysis of private investment decision making using tools of robo-advisers in long-run period. In: ICTERI Workshops, pp. 144–159 (2018)
14. Kidd CD, Breazeal C (2007) A robotic weight loss coach. In: Proceedings of the national conference on artificial intelligence, vol. 22, p. 1985. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, Orlando, Florida (2007)
15. Nisimura R, Uchida T, Lee A, Saruwatari H, Shikano K, Matsumoto Y (2002) ASKA: receptionist robot with speech dialogue system. In: IEEE/RSJ International conference on intelligent robots and systems, vol. 2, pp. 1314–13192. https://doi.org/10.1109/IRDS.2002.1043936
16. Lee MK, Kiesler S, Forlizzi J (2010) Receptionist or Information Kiosk: How Do People Talk with a Robot? In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work. CSCW '10, pp. 31–40. ACM, New York, NY, USA. https://doi.org/10.1145/1718918.1718927
17. Walters ML, Syrdal DS, Dautenhahn K, te Boekhorst R, Koay KL (2008) Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. Auton Robot 24(2):159–178. https://doi.org/10.1007/s10514-007-9058-3
18. Byrne D (1961) Interpersonal attraction and attitude similarity. Psychol Sci Public Interest 62:713–715
19. Meltzoff AN, Prinz W (eds.) (2002) The imitative mind: development, evolution and brain bases. Cambridge Studies in Cognitive and Perceptual Development. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511489969
20. Meltzoff AN (2007) 'Like me': a foundation for social cognition. Dev Sci 10(1):126–134. https://doi.org/10.1111/j.1467-7687.2007.00574.x
21. Cross ES, Ramsey R, Liepelt R, Prinz W, Hamilton A.F.d.C (2016) The shaping of social perception by stimulus and knowledge cues to human animacy. Philos Trans Royal Soc B: Biol Sci 371(1686), 20150075. https://doi.org/10.1098/rstb.2015.0075. Publisher: Royal Society
22. Hortensius R, Cross ES (2018) From automata to animate beings: the scope and limits of attributing socialness to artificial agents. Ann N Y Acad Sci 1426(1):93–110. https://doi.org/10.1111/nyas.13727
23. Salek Shahrezaie R, Anima BA, Feil-Seifer D (2021) Birds of a feather flock together: a study of status homophily in HRI. In: Li H, Ge SS, Wu Y, Wykowska A, He H, Liu X, Li D, Perez-Osorio J (eds.) Social robotics. Lecture Notes in Computer Science, pp. 281–291. Springer, Cham. https://doi.org/10.1007/978-3-030-90525-5_24
24. Goetz J, Kiesler, S, Powers A (2003) Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: Proceedings of the 12th IEEE international workshop on robot and human interactive communication, 2003. ROMAN 2003., pp. 55–60. IEEE, Millbrae, CA, USA
25. Li D, Rau PLP, Li Y (2010) A cross-cultural study: effect of robot appearance and task. Int J Soc Robot 2(2):175–186. https://doi.org/10.1007/s12369-010-0056-9
26. Woods S, Dautenhahn K, Kaouri C, Boekhorst R, Koay KL (2005) Is this robot like me? Links between human and robot personality traits. In: 5th IEEE-RAS international conference on humanoid robots, 2005., pp. 375–380. https://doi.org/10.1109/ICHR.2005.1573596
27. Tversky A, Kahneman D (1983) Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychol Rev 90:293–315
28. Bar-Hillel M, Neter E (1994) How alike is it versus how likely is it: a disjunction fallacy in probability judgments. J Pers Soc Psychol 65(6):1119. https://doi.org/10.1037/0022-3514.65.6.1119
29. Morier DM, Borgida E (1984) The conjunction fallacy: a task specific phenomenon? Pers Soc Psychol Bull 10(2):243–252. https://doi.org/10.1177/0146167284102010
30. Hertwig R, Gigerenzer G (1999) The 'conjunction fallacy' revisited: how intelligent inferences look like reasoning errors. J Behav Decis Mak 12(4):31
31. Charness G (2009) On the conjunction fallacy in probability judgment: new experimental evidence regarding Linda. Economics Working Paper Archive, The Johns Hopkins University, Department of Economics
32. Cialdini RB, Goldstein NJ (2004) Social influence: compliance and conformity. Annu Rev Psychol 55(1):591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015
33. Qin X, Chen C, Yam KC, Cao L, Li W, Guan J, Zhao P, Dong X, Lin Y (2022) Adults still can't resist: a social robot can induce normative conformity. Comput Hum Behav 127:107041. https://doi.org/10.1016/j.chb.2021.107041.
34. Vollmer A-L, Read R, Trippas D, Belpaeme T (2018) Children conform, adults resist: a robot group induced peer pressure on normative social conformity. Sci Robot 3(21), 7111. ISBN: 2470-9476 Publisher: American Association for the Advancement of Science
35. Kolmogorov AN (2013) Foundations of the theory of probability. Martino Fine Books, Eastford, CT, USA
36. Costello F, Watts P, Fisher C (2018) Surprising rationality in probability judgment: assessing two competing models. Cognition 170:280–297. https://doi.org/10.1016/j.cognition.2017.08.012
37. Polakow T, Teodorescu AR, Busemeyer JR, Gordon G (2021) Free ranking vs. rank-choosing: new insights on the conjunction fallacy. https://psyarxiv.com/r9kxp
38. Wallkötter S, Stower R, Kappas A, Castellano G (2020) A robot by any other frame: framing and behaviour influence mind perception in virtual but not real-world environments. In: Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction. HRI '20, pp. 609–618. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3319502.3374800
39. Marchesi S, Perez-Osorio J, Tommaso DD, Wykowska A (2020) Don't overthink: fast decision making combined with behavior variability perceived as more human-like. In: 2020 29th IEEE international conference on robot and human interactive communication (RO-MAN), Naples, Italy, 2020, pp. 54–59. https://doi.org/10.1109/RO-MAN47096.2020.9223522
40. Hsieh T-Y, Chaudhury B, Cross ES (2020) Human-robot cooperation in prisoner dilemma games: people behave more reciprocally than prosocially toward robots. In: Companion of the 2020 ACM/IEEE international conference on human-robot interaction. HRI '20, pp. 257–259. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3371382.3378309

41. Laban G, George J-N, Morrison V, Cross ES (2021) Tell me more! Assessing interactions with social robots from speech. Paladyn J Behav Robot 12(1):136–159. https://doi.org/10.1515/pjbr-2021-0011
42. Gray HM, Gray K, Wegner DM (2007) Dimensions of mind perception. Science 315(5812), 619–619. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science
43. Epley N, Waytz A (2010) Mind perception. In: Handbook of social psychology. American Cancer Society, Atlanta, Georgia, USA
44. Vanman EJ, Kappas A (2019) "Danger, Will Robinson!" The challenges of social robots for intergroup relations. Soc Personal Psychol Compass 13(8):12489. https://doi.org/10.1111/spc3.12489
45. Laban G, Araujo T (2020) The effect of personalization techniques in users' perceptions of conversational recommender systems. In: Proceedings of the 20th ACM international conference on intelligent virtual agents. IVA '20, pp. 1–3. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3383652.3423890
46. Short E, Hart J, Vu M, Scassellati B (2010) No fair. An interaction with a cheating robot. In: 2010 5th ACM/IEEE international conference on human-robot interaction (HRI), pp. 219–226. https://doi.org/10.1109/HRI.2010.5453193
47. Salem M, Lakatos G, Amirabdollahian F, Dautenhahn K (2015) Would you trust a (Faulty) robot?: effects of error, task type and personality on human-robot cooperation and trust. In: Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction. HRI '15, pp. 141–148. ACM, New York, NY, USA. https://doi.org/10.1145/2696454.2696497
48. Gompei T, Umemuro H (2015) A robot's slip of the tongue: effect of speech error on the familiarity of a humanoid robot. In: 2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN), pp. 331–336. https://doi.org/10.1109/ROMAN.2015.7333630
49. Ragni M, Rudenko A, Kuhnert B, Arras KO (2016) Errare humanum EST: Erroneous robots in human-robot interaction. In: 2016 25th IEEE International symposium on robot and human interactive communication (RO-MAN), pp. 501–506. https://doi.org/10.1109/ROMAN.2016.7745164
50. Mirnig N, Stollnberger G, Miksch M, Stadler S, Giuliani M, Tscheligi M (2017) To err is robot: how humans assess and act toward an erroneous social robot. Front Robot AI 4. https://doi.org/10.3389/frobt.2017.00021
51. Weiss A, Bartneck C (2015) Meta analysis of the usage of the godspeed questionnaire series. In: 2015 24th IEEE International symposium on robot and human interactive communication (RO-MAN), pp. 381–388. https://doi.org/10.1109/ROMAN.2015.7333568
52. Higgins TE, Rholes WS, Jones CR (1977) Category accessibility and impression formation. J Exp Soc Psychol 13(2), 141–154. https://doi.org/10.1016/S0022-1031(77)80007-3
53. Bargh JA, Chartrand TL (2014) The mind in the middle: a practical guide to priming and automaticity research. ISBN: 1107011779 Publisher: Cambridge University Press
54. Li LMW, Masuda T, Hamamura T, Ishii K (2018) Culture and decision making: influence of analytic versus holistic thinking style on resource allocation in a fort game. J Cross Cult Psychol 49(7):1066–1080. https://doi.org/10.1177/0022022118778337
55. Basu S, Savani K (2017) Choosing one at a time? presenting options simultaneously helps people make more optimal decisions than presenting options sequentially. Organ Behav Hum Decis Process 139:76–91. https://doi.org/10.1016/j.obhdp.2017.01.004
56. Latikka R, Savela N, Koivula A, Oksanen A (2021) Attitudes toward robots as equipment and coworkers and the impact of robot autonomy level. Int J Soc Robot. https://doi.org/10.1007/s12369-020-00743-9
57. Alves-Oliveira P, Sequeira P, Paiva A (2016) The role that an educational robot plays. In: 2016 25th IEEE International symposium on robot and human interactive communication (RO-MAN), pp. 817–822. https://doi.org/10.1109/ROMAN.2016.7745213. ISSN: 1944-9437
58. Gelin R (2019) NAO. In: Goswami A, Vadakkepat P (eds) Humanoid robotics: a reference. Springer, Dordrecht, pp 147–168
59. Vishwanath A, Singh A, Chua YHV, Dauwels J, Magnenat-Thalmann N (2019) Humanoid co-workers: How is it like to work with a robot? In: 2019 28th IEEE international conference on robot and human interactive communication (RO-MAN), pp. 1–6. IEEE Press, New Delhi, India. https://doi.org/10.1109/RO-MAN46459.2019.8956421
60. Chen M (2022) Application status of intelligent investment consultant based on artificial intelligence in China, pp. 789–792. Atlantis Press, Amsterdam. https://doi.org/10.2991/aebmr.k.220307.127. ISSN: 2352-5428. https://www.atlantis-press.com/proceedings/icfied-22/125971623 Accessed 2022-07-05
61. Polakow T, Gordon G, Busemeyer JR, Teodorescu AR (2020)Preregistration: do people prefer logical or fallacious robots for different tasks? Pre-registration. https://aspredicted.org/43u4h.pdf
62. Annonymous: supplementary information: interaction with fallacious robots (2021). https://osf.io/f92cm/?view_only=83ed6233ff814dd99f1b42bd172ef820
63. Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci 22(11):1359–1366. https://doi.org/10.1177/0956797611417632
64. Nelson LD, Simmons JP, Simonsohn U (2012) Let's publish fewer papers. Psychol Inq 23(3):291–293. https://doi.org/10.1080/1047840X.2012.705245
65. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, Simonsohn U, Wagenmakers E-J, Ware JJ (2017) Ioannidis, J.P.A.: A manifesto for reproducible science. Nat Human Behav 1(1), 1–9. https://doi.org/10.1038/s41562-016-0021. Number: 1 Publisher: Nature Publishing Group
66. Syrdal DS, Dautenhahn K, Koay KL, Walters ML (2009) The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. Adaptive and emergent behaviour and complex systems. SSAISB, London
67. Deshmukh A, Craenen B, Foster ME, Vinciarelli A (2018) The more I understand it, the less I like it: the relationship between understandability and godspeed scores for robotic gestures. In: 2018 27th IEEE International symposium on robot and human interactive communication (RO-MAN), pp. 216–221. https://doi.org/10.1109/ROMAN.2018.8525585
68. Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol 2(2):175–220. https://doi.org/10.1037/1089-2680.2.2.175
69. Huber J, Payne JW, Puto C (1982) Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. J Consum Res 9:90–98
70. Nelson DL, McEvoy C (2007) Entangled associative structures and context. In: AAAI Spring symposium: quantum interaction pp. 98–105
71. Armstrong T, Rockloff M, Browne M, Blaszczynski A (2020) Encouraging gamblers to think critically using generalised analytical priming is ineffective at reducing gambling biases. J Gambl Stud 36(3):851–869. https://doi.org/10.1007/s10899-019-09910-8
72. Samar SM, Walton KE, McDermut W (2013) Personality traits predict irrational beliefs. J Ration Emotive Cogn Behav Ther 31(4):231–242. https://doi.org/10.1007/s10942-013-0172-1
73. Oehler A, Wendt S, Wedlich F, Horn M (2018) Investors' personality influences investment decisions: experimental evidence on

extraversion and neuroticism. J Behav Financ 19(1):30–48. https://doi.org/10.1080/15427560.2017.1366495
74. Sava FA (2009) Maladaptive schemas, irrational beliefs, and their relationship with the Five-Factor Personality model. J Cogn Behav Psychother 9(2):1–13
75. Ferber RC (1967) The role of the subconscious in executive decision-making. Manage Sci 13(8):519. https://doi.org/10.1287/mnsc.13.8.B519
76. Welsh DT, Ordóñez LD (2014) Conscience without cognition: the effects of subconscious priming on ethical behavior. Acad Manag J 57(3):723–742. https://doi.org/10.5465/amj.2011.1009
77. Bell L, Vogt J, Willemse C, Routledge T, Butler LT, Sakaki M (2018) Beyond self-report: a review of physiological and neuroscientific methods to investigate consumer behavior. Front Psychol 9:1655
78. Zoëga Ramsøy T, Michael N, Michael I (2019) A consumer neuroscience study of conscious and subconscious destination preference. Sci Rep 9(1):15102. https://doi.org/10.1038/s41598-019-51567-1