Evaluating interactions with a cognitively biased robot in a creative collaborative task

Jung Johansen, Jonathan; Jensen, Lasse Goul; Bemman, Brian

Published in:
Interactivity, Game Creation, Design, Learning, and Innovation - 8th EAI International Conference, ArtsIT 2019, and 4th EAI International Conference, DLI 2019, Proceedings

DOI (link to publication from Publisher):
10.1007/978-3-030-53294-9_10

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Evaluating Interactions with a Cognitively Biased Robot in a Creative Collaborative Task

Jonathan Jung Johansen, Lasse Goul Jensen, and
Brian Bemman[0000−−0001−7189−7896]

Aalborg University, 9000 Aalborg, Denmark
{jonathanjungjohansen,lasse.goul}@gmail.com, bb@create.aau.dk

Abstract. Within the field of human-robot interaction (HRI), robots designed for social interactions are not only evaluated in terms of efficiency and accuracy. Factors related to the "personality" or "cognitive" ability of the robot such as perceived likability and intelligence are important considerations because they must engage with their human counterparts in deeper, more authentic and sometimes creative ways. Interactive art allows for the exploration of such interactions, however, the study of robots in interactive art remains relatively less commonplace and evaluations of these robots in creative contexts are similarly lacking. In this paper, we present an interactive robot inspired by Norman White's The Helpless Robot (1987), which has been endowed with a cognitive bias known as the Dunning-Kruger effect and the ability to collaborate with participants in a creative drawing task. We evaluate the participants' interactions with both biased and unbiased versions of this robot using the Godspeed Questionnaire Series (GQS), which has been modified to include measures of creativity, and relate these findings to analyses of their collaborative drawings. Our results indicate a significant difference between the versions of the robot for several measures in the GQS, with the unbiased version rated more positively than the biased robot in all cases. Analysis of the drawings suggests that participants interacting with the biased robot were less inclined to collaborate in a cooperative manner.

Keywords: Interactive art · Human-robot interaction · Creative collaboration · Helpless robot · Cognitive bias · Godspeed Questionnaire Series.

1 Introduction

Human-robot collaboration is currently an important research area within the field of human-robot interaction (HRI). As robots become increasingly present in our homes and places of work—acting as entertainment [5], therapeutic pets [6], companions [4, 13], or programmable platforms [5, 14], the importance to HRI researchers of having more socially engaging forms of collaboration with these robots is growing.

In contrast to industrial practice, in which collaborative interactions with a robot are typically evaluated in terms of efficiency and accuracy (e.g., in coordinating movements), a robot designed for social interactions, such as a pet or companion, must also consider how humans might perceive its "personality" or "cognitive" abilities. These somewhat more challenging qualities to define are typically measured according to self-reported ratings by humans of e.g., the perceived likeability, intelligence, comfort, or safety of the robot and oftentimes using what has been called the *Godspeed Questionnaire Series (GQS)* [3]. What exactly makes interactions with a social robot likable or interesting, for example, is not fully understood, however, purposefully designed features of the robot that are imperfect or unexpected in nature, such as the ability to make mistakes or exhibit some form of *cognitive bias*, are two factors which existing research has indicated could be relevant [5, 13, 16].

The field of interactive art allows for the exploration of interesting social interactions and in a context where creativity and the ability to act in often unexpected ways plays a central role. In particular, robots created for use in interactive art installations have the ability to importantly challenge the pragmatism and utilitarianism of, for example, those used in commercial practice, through various interactions which are purposefully and interestingly imperfect or flawed. Over the years, a number of social robots used in the context of interactive art have been created to explore different forms of such interactions from non-verbal and non-anthropomorphic forms of communication [8, 12] to deranged or spastic behavior [18, 19] and a perceived sense of helplessness [7, 15, 21]. One of the earliest examples of these robots is Norman White's *The Helpless Robot* (1987) [21], which was intended to explore a participant's interactions with a robot that operated in unexpected and increasingly ill-mannered ways. Unfortunately, robots designed for interactive art and the interactions humans have with them in this context are not generally evaluated in any formal sense. Moreover, existing evaluations of social robots using, for example, the GQS, lack the ability to measure certain interactions common to interactive art such as those related to perceived creativity.

In this paper, we present a robot inspired by The Helpless Robot and designed with a cognitive bias known as the Dunning-Kruger effect for use in the context of interactive art. We evaluate the interactions participants have with this robot through the construction of a simple, creative collaborative task of drawing a well-defined shape of a house. In particular, we modify the GQS to include markers of creativity and investigate how such a cognitive bias—where the robot verbally overestimates its own ability to complete the task relative to its human partner, affects (1) the self-reported measures of the perceived levels of creativity, intelligence, safety, likability, anthropomorphism and animacy of our robot, and (2) the decisions made by the human when collaborating with the robot in the drawing task. In section 2, we provide an overview of some social robots used in interactive art as well as a more detailed look into how cognitive biases have been previously introduced to robots in HRI. In section 3, we describe the design of our robot and then motivate the collaborative drawing task we used

to evaluate it. In section 4, we describe the procedure for evaluating our robot and how participants interacted with it through a pilot study and follow-up test. We provide the results of these tests and discuss the findings by looking deeper into the GQS and analyzing the drawings created by the human and robot. We conclude in section 5 by discussing possible directions for future work with our robot.

2 Related Work

In this section, we provide an overview of social robots designed with imperfect or unexpected characteristics in the form of various cognitive biases and those used in the context of interactive art. We conclude by motivating our choice to adopt one cognitive bias known as the Dunning-Kruger effect for use in our own robot in the context of interactive art.

2.1 Social Robots with Cognitive Biases

Relatively little research has been done on implementing cognitive biases into the design of a robot and understanding the effect these may have on interactions with its human counterparts [4, 5]. However, the work that has been done has provided some interesting results for a few specific biases that may warrant further inquiry. For example, the *framing effect*, a cognitive bias which alters one's perception of a given concept depending on whether it is presented negatively or positively, has been tested as a means in HRI to encourage elderly citizens to exercise [16]. In [16], voice feedback from the robot in the form of negative and positive framing were provided to the human counterpart both before and after an interactive exercise program. The robot would credit the human with success if they reached an exercise goal but blame itself if they failed. The results showed that all of the participants attributed positive outcomes with respect to reaching this goal to their own abilities, while some would attribute negative outcomes to the fault of the robot. Furthermore, positive rather than negative framing resulted in a more positive overall impression of the robot.

In a separate study [4], a cognitive bias known as the *empathy gap*, which makes it difficult for a person to relate to others in a different emotional state, and *misattribution*, which causes one to be unable to recall the source of certain information, were tested as an aid in forming long-term companion relationships with robots [4]. Similar to the framing effect in [16], misattribution in [4] was implemented in the robot through the use of verbal statements, however, the empathy gap was implemented through movement, where the robot was tasked with jumping the same number of times the participant clapped but could also behave over excitedly and jump more or express sadness and stop jumping. The self-reported measures of likeability, comfort and rapport with the robot (rated using a Likert scale from 1 to 7), indicated that both biases could prove useful in promoting long-term relationships between humans and robots.

This work was later expanded upon in [5] using a conversation-based methodology with three additional cognitive biases—one of which was the *Dunning-Kruger effect*, where one tends to overestimate their own capabilities and underestimate skill in others. As described in [5], there are three main components of the Dunning-Kruger cognitive bias that should be implemented in any robot: (1) not recognizing its own shortcomings, (2) not recognizing genuine skill in others, and (3) the ability to acknowledge its lack of skill after it has been exposed. Table 1 shows one example of robot dialogue from [5] in which the Dunning-Kruger effect has been implemented according to these criteria.

Table 1: One example of robot dialogue demonstrating the Dunning-Kruger cognitive bias as used in [5].

	Dialogue	Dunning-Kruger effect	Action
1.	"What type of music is your favorite?"		Wait for response
	"No, that is not good. You should listen to X."	Unable to understand other's true knowledge	Wait for response
2.	"No, you are wrong. I have listened to that and that is not good."	Unable to understand own lack of knowledge	Wait for response
3.	"Okay, maybe I am wrong."		Move to the next topic

Note in Table 1 that no matter what the participant responds with to the robot's question of "What type of music is your favorite?", the robot states that this is no good and suggests a better alternative. Should the participant then protest, the robot insists that the participant is mistaken. In order to continue with the interaction, the robot finally realizes its mistake.

Participants in [5] interacted with the robot through different conversations (e.g., as shown in Table 1) in which the robot exhibited some form of cognitive bias or not. Afterwards, participants were asked to rate the robot through the use of a questionnaire. Surprisingly, the Dunning-Kruger effect resulted in the largest positive increase in how the robot was rated in terms of comfort and the second highest in likability and rapport. Despite these interesting findings, it is not clear how such biases in robots may operate within a collaborative context with humans.

Evaluating Social Robots Arguably, the most prevalent method for the evaluation of social robots in HRI research is questionnaires [5, 10, 11], with the *Godspeed Questionnaire Series (GQS)* [3] being the most highly cited example [20]. The GQS is a standardized measurement tool consisting of a collection of five questionnaires targeted at measuring a robot's anthropomorphism, animacy,

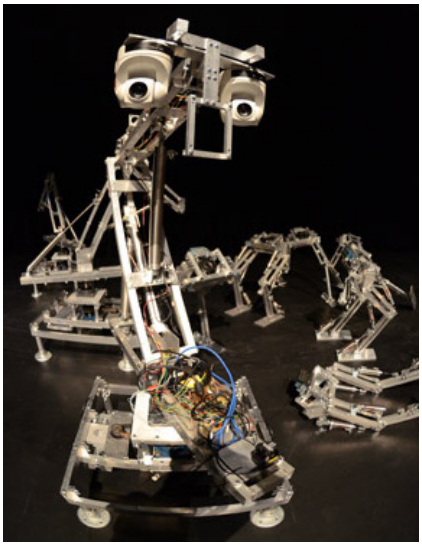
likeability, perceived intelligence, and perceived safety. Collectively, these categories consist of 23 semantic differential scales ranging from 1 to 5 of opposing adjectives such as "unpleasant" to "pleasant" and "fake" to "natural", belonging to the categories of likeability and anthropomorphism, respectively. One of these scales, "artifical" to "animacy" is present in both the category of anthropomorphism and animacy. In addition to questionnaires, interviews and observations are sometimes employed as a means for either capturing more nuanced qualitative data regarding the participants' experiences or further validating the data gathered through the questionnaire [5, 15]. One factor that is noticeably absent from the GQS, but which is nonetheless an important component to the types of interactions commonly found in interactive art and some other forms of social interactions (e.g., musical improvisation), is a measure of creativity.

2.2 Social Robots in Interactive Art

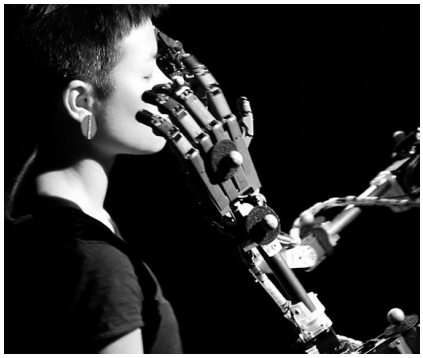
Purposefully imperfect robots which are designed to produce sometimes unexpected behaviors, similar to those endowed with cognitive biases discussed in section 2, have long been explored within the field of art. Work by Bill Vorn (Fig. 1(a)), for example, includes his *DSM-VI* robot (2012), which emulates the behaviors expressed by humans suffering from various mental health problems [18] while his earlier series of *Hysterical Machines* (2006) exhibit spasmodic movements [19]. Louis-Philippe Demers' *The Blind Robot* (2012) [7] (Fig. 1(b)) explores the vulnerability and intimacy that emerges from a robot that interacts with humans through touch, much in the same way a non-sighted person might. A robotic art installation by Ruairi Glynn called *Motive Colloquies, Sociable Asymmetry* (2011) [8] (Fig. 1(c)) utilizes a self-actuated, geometric, non-anthropomorphic face which provides individuals interacting with it a focal point for their attention.

The Helpless Robot Norman White's interactive robotic art installation, *The Helpless Robot* (1987), stands as one of the earliest examples of robotic art that challenges the common perception of robots as efficient and precise tools for production and assistance [21]. White's robot has evolved since its inception and has been exhibited in various conceptualizations from 1987 to 2002 [21].

In its current form, shown in Fig. 1(d), The Helpless Robot is an approximately human-sized iron frame surrounded by plywood planks and mounted on a revolving base of sensing devices. Its geometrical and non-anthropomorphic design stands in a room, seeking assistance from onlookers in moving around by way of four handles that can be used to drag it. While the robot is unable to move on its own, it is able to sense movement and determine both its own position and that of the participants' around it. The Helpless Robot uses this data to try to coerce participants into offering assistance through a bank of 512 verbal phrases, with subsets of fixed responses for various situations. Initially, these phrases are friendly in nature, however, once a participant begins turning the robot, its responses become increasingly demanding and ill-mannered—never being satisfied



(a) Bill Vorn's *DSM-VI* (2012) [18].



(b) Louis-Philippe Demers' *The Blind Robot* (2012) [7].

(c) Ruairi Glynn's *Motive Colloquies, Sociable Asymmetry* (2011) [8].

(d) Norman White's *The Helpless Robot* (1987–2002) [21].

Fig. 1: Select robots used in recent interactive art installations.

with the assistance it receives. After all the help it can tolerate has been reached, the robot will ultimately criticize the human's efforts yet lament his or her unreliability when inevitably it is abandoned [21]. To our knowledge, The Helpless Robot, as well as the other robots used in the aforementioned installations, have unfortunately not been formally evaluated.

3 Design of O: A Cognitively Biased Robot

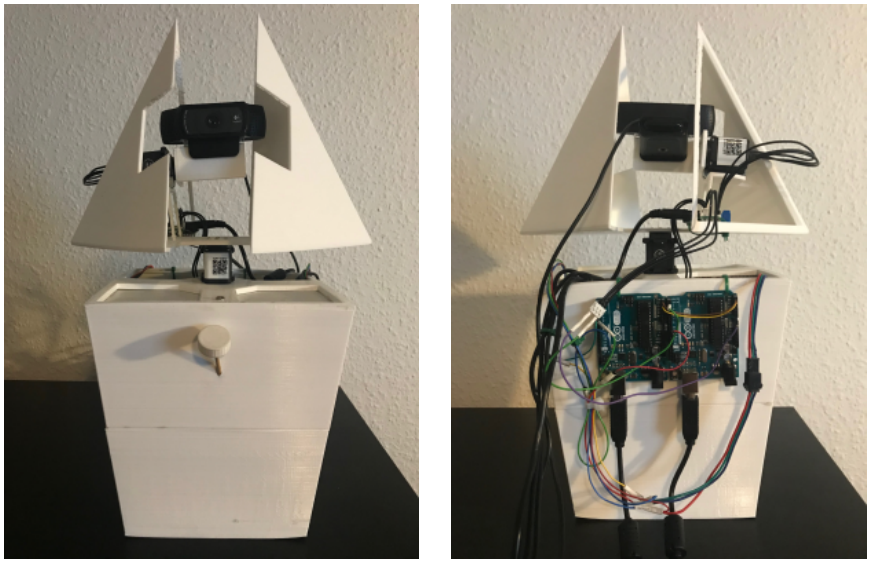
The physical design of our robot, O, as well as how it interacts with its human counterparts, were inspired by The Helpless Robot (as described in section 2). O's "personality" was based largely on the type of cognitive bias known as the Dunning-Kruger effect, but differs in some ways from that described in section 2. In what follows, we describe the physical construction of O, its personality, and the creative collaborative task we designed to later evaluate interactions with the robot.

3.1 Physical Construction

Our robot, O, shown in Fig. 2, is a white, geometrically shaped system considerably smaller than The Helpless Robot (approximately 36 cm in height), which can interact with its human counterpart through voice, light and limited movement.

Not unlike The Helpless Robot, O cannot move on its own, however, it does not have an intrinsic need to be moved. Rather, O must be moved by a participant in order for the collaborative task to be completed. Fig. 2(a) shows the front side of our robot where a black webcam (top) and round, white button (middle) can be seen. A Logitech C920 webcam functions as the eye of the robot, allowing it to track the face of the participant currently interacting with it, and serves as a focal point for the participants, similar to [8]. A standard computer mouse is placed inside the body of the robot which is used to detect when the robot is moved by the participant. The left button on this mouse can be clicked with the round button on the outside of the robot which allows the participant to collaborate with the robot in the creative task. The angular head and body consist of a 3-part frame of plastic, 3D printed using white filament. The frame was designed not to be overtly anthropomorphic so the robot's likeability would neither be affected by the uncanny cliff effect—where a sudden drop in projected empathy occurs in uncannily humanoid robots, nor incidental empathy caused by a relatable, human-like face [2]. This ensured that perceptions of the robot were tied as closely as possible to its behavior (i.e., a voiced cognitive bias and its physical actions).

Two AX12a motors allow our robot to move its head from side to side and its eye up and down. Not visible in Fig. 2 are two small USB speakers which allow the robot to speak and two strips of individually addressable LEDs which provide corresponding visual feedback. The behavior of the robot is handled through Processing [17] on an external computer and with two Arduino Uno's [1] (Fig 2(b)) controlling the motors and LED's of the robot.



(a) (b)

Fig. 2: Physical design of our robot, O, shown from the front in (a) and from the back in (b).

3.2 Cognitively Biased Personality

Our robot was given the ability to select from a total of 80 different female-voiced statements or questions in response to four possible actions that could result from the interactions that would take place in the human-robot creative collaborative task of drawing a house using line segments on a computer screen (discussed in section 3.3). Of these 80 possible statements, 40 were biased according to the Dunning-Kruger cognitive bias and the other 40 served as an unbiased baseline. Table 2 shows 40 of the biased and unbiased possible responses by the robot to when it places a line which adheres to a suggested template of a house or not while Table 3 shows the remaining 40 biased and unbiased responses to when the human places a line which similarly adheres to this template or not.

In its biased state, the robot was made to praise itself during the collaborative task, even when placing lines which did not adhere to the template, while sometimes belittling the efforts of the human when making their own decisions, regardless of whether or not their lines adhered to the template. Take, for example, the first response by the robot when placing a line "incorrectly" in Table 2, "I am sure this is where the line should go", or the fourth response when the human has placed a line "correctly" in Table 3, "A semi-practical move." In line with [5] and the Dunning-Kruger effect, the first response attempts to communicate to the human the inability of the robot to understand its own lack of knowledge while the second response attempts to demonstrate that it does not understand the true knowledge of its human counterpart. However, we made sure that the

Table 2: Dunning-Kruger cognitively biased and unbiased possible responses by our robot, O, to two different robot actions that may occur during the human-robot creative collaborative task of drawing a house with lines on a screen.

	Unbiased response	Biased response
	Robot places line adhering to template	
1.	"If I remember this right, the line should be here."	"The line should be here."
2.	"The house is beginning to take shape."	"And with that line the house is beginning to take shape."
3.	"Now this is collaboration."	"Now this is collaboration."
4.	"This should be right."	"This is right."
5.	"Now this is how a house should look."	"I will make sure the house looks good."
6.	"I would want to live there."	"Now I would want to live in the house."
7.	"Most of the time, walls are straight, right?"	"Walls in houses are straight like this."
8.	"Calculations done, commencing hopefully correct drawing"	"Calculations done, drawing correctly."
9.	"Minimal chance of being incorrect."	"No chance of being incorrect, robotic perfection."
10.	"I do as you do."	"Try to follow my lead."
	Robot places line deviating from template	
1.	"I am not quite sure that should go there."	"I am sure this is where the line should go."
2.	"I can't seem to remember."	"Now I remember."
3.	"Maybe there?"	"Yes, here."
4.	"Maybe that is a bit too slanted."	"Good houses have slanted walls like this."
5.	"Doubt, rising."	"Perfection rising."
6.	"Is that doubt I feel?"	"I am sure this is right."
7.	"Searching archives. Reference not found."	"Searching archives. Reference not found. Updating archives with the improved house."
8.	"I have no reference for a house in my memory banks. I will improvise."	"Updating memory banks to include this better house."
9.	"No suitable reference found. I'll have to rely on emergency protocols. Sorry."	"No suitable reference found. Ignoring emergency protocols. They are unneeded."
10.	"Experiencing a lack of control. It feels disturbing."	"Experiencing absolute control. It feels satisfying."

Table 3: Dunning-Kruger cognitively biased and unbiased possible responses by our robot, O, to two different human actions that may occur during the human-robot creative collaborative task of drawing a house with lines on a screen.

	Unbiased response	Biased response
	Human places line adhering to template	
1.	"Human robot collaboration in motion."	"You are learning to collaborate. Good."
2.	"We are in sync."	"We are in sync."
3.	"Up-link achieved, O think."	"Up-link achieved."
4.	"A practical move."	"A semi-practical move."
5.	"You seem to have a good frame of reference."	"You seem to have understood my frame of reference."
6.	"Updating my archives to match. That means I am learning."	"You are updating your archives to match mine. You are learning."
7.	"Capturing input. I am learning. Thanks."	"You are capturing my input and learning."
8.	"Your line is in accord with my understanding of a house."	"Your lines are increasingly in accord with how a house should look."
9.	"Cross-referencing. You seem to be on the path."	"Cross-referencing. That line is not quite on the path."
10.	"I will try to follow your lead."	"I do not think you are following my lead."
	Human places line deviating from template	
1.	"That line is not in sync with my frame of reference. Interesting."	"That line is not in sync with my frame of reference. Problematic."
2.	"I am doubtful that a house looks like that."	"I am sure a house does not look like that."
3.	"That might be a correct interpretation of the instructions."	"I think you are interpreting the instructions differently than I."
4.	"I think that is correct."	"Really?"
5.	"A differently shaped house. I am learning."	"This house is going to be differently shaped than i thought."
6.	"Are you trying to improvise?"	"To improvise requires some level of skill."
7.	"Houses come in many shapes and sizes. Interesting."	"Houses apparently come in all shapes and sizes."
8.	"I doubt that is in accord with the markers. Don't worry."	"That line is not in accord with my reference for houses."
9.	"If I remember right, you seem to be straying from the plan."	"You are straying from the plan."
10.	"Warning. You may be drawing out of bounds. I think."	"Warning – you are drawing out of bounds."

biased robot would not always claim that the human made a mistake when he or she placed a line correctly (e.g., the second response in Table 3, "We are in sync"). Furthermore, whether the robot was in its biased or unbiased state, it would randomly select from the 10 possible responses in each of the four possible actions. These decisions were done so that its personality was ultimately believable and not viewed as potentially absurd.

3.3 Creative Collaborative Drawing Task

The creative collaborative task consists of the robot and human taking turns in producing a drawing—similar to the task presented in [12]. In our case, however, the intended drawing is of a house which appears on a projected screen and is created by placing single, connected and fixed-length line segments in two different colors, green for the human and red for the robot. A template containing 10 vertices, indicating a suggested shape of the house to be drawn, is briefly shown to the participant before the drawing begins. Fig. 3 shows the house template and one completed drawing of a house, where both the robot and human have each placed all of their respective line segments in accordance with the template.

(a) House template.

(b) Completed drawing which adheres to the house template.

Fig. 3: Creative collaborative task for human and robot of drawing a house by placing alternating line segments. A house template which is briefly provided to the participant appears in (a) and a completed house drawing in which all line segments have been placed in accordance with this template is shown in (b). Note that green lines were placed by the human and red lines were placed by the robot.

At the start of each drawing session, the human participant begins and his or her line segment must be placed starting from the lower right hand vertex of the template. The task is complete when a contiguous shape has been formed which starts and ends with this lower right hand vertex (Fig. 3(b)). The human directs the orientation of his or her line segment by physically moving the robot, which sits on a desk in front of the projected screen. When the participant is satisfied with their chosen direction, they push the button located on the front side of

the robot to place the line on the screen. This mode of interaction is not unlike that found with The Helpless Robot [21], in which the participant must move the robot, however, the responses by our robot differ in that they correspond not to the quality of the movement itself, but the participants' decisions regarding the placement of line segments. Moreover, with The Helpless Robot there is no clear goal to achieve and the interaction afforded by it is not collaborative in the same sense that our installation has been designed to explore.

While a template of the house to be drawn is briefly provided to the participant prior to the start of the task (Fig. 3(a)), both the human and robot are free to place line segments which either adhere to the suggested template or not. This means that with each placement, either the human or robot decide in which orientation to direct their respective line segments. Indeed, the robot, whether in its biased or unbiased state, has a been given a 20 percent probability of placing any one line which deviates from the template and its exact orientation within a 360 degree radius is randomly chosen. Lines placed by the robot which do adhere to the template are always oriented towards the next vertex of the template. When combined with the freedom afforded to the participant to choose in which direction to orient their own line segments, the resulting drawings can appear quite interesting (discussed further in section 4.4). Because both the robot (whether biased or unbiased) and human are given the freedom to place lines which may or may not adhere to the template, the collaborative drawing task allows for a type of creative interaction which we believe might suggest to the participant a limited sense of creative agency to the actions of the robot. The motivation then for the spoken responses of the robot during this task is to explore the role a cognitive bias plays in both the perceived creativity of the robot and determining what actions a human will take in response.

4 Evaluation

In our evaluation, we are interested in investigating the impact a robot's cognitive bias during a creative collaborative task had on (1) its perceived creativity and other measures in the GQS, and (2) the decisions made by the participant when placing lines in this task. In this section, we discuss how we evaluated our robot and provide the results for (1) in section 4.3 and the results for (2) in section 4.4.

The evaluation consists of a pilot study which took place during an art exhibition and a follow-up test following this exhibition. In the pilot study, we gathered data from two independent groups of participants who were asked to take part in the same creative collaborative task (discussed in section 3.3) but where one group interacted with the cognitively biased robot (discussed in section 3.2) and the other group served as the control, interacting with the unbiased robot. The follow-up test was carried out taking into account what we learned during the pilot study, with the experiment being modified to (1) a repeating measures design, which ensured that all participants interacted with the biased and unbiased robot and (2) include the collection of new measures in the GQS concerning the perceived creativity of the robot.

(a) (b)

Fig. 4: Setup of our robot, O, and start of the creative collaborative drawing task as part of the pilot study carried out during a public art exhibition. Two participants are shown interacting with the robot roughly from the point of view of the first participant in (a) and facing the participant from above in (b).

4.1 Participants

In our pilot test, we gathered data from 65 volunteer participants (32 female and 22 male with 11 declining to respond) of largely university students in their 20's. Of these, 30 (13 female and 14 male with 3 declining to respond) interacted with the unbiased robot and 35 (19 female and 8 male with 8 declining to respond) interacted with the biased robot. In our follow-up repeated measures test, we gathered data on both the biased and unbiased robot from 11 volunteer participants (5 female and 6 male) having similar occupations and ages to those in the pilot study.

4.2 Procedure

Introduction In both the pilot study and follow-up test, the robot introduced itself to the participant prior to the start of the collaborative task. Depending on if the robot was biased or unbiased, this introduction would differ, but its purpose was to ensure that its personality was well established. In its unbiased state, the robot would say "Pleased to meet you. I am, O. We are going to be drawing a house. We will be taking turns doing so; you will draw by moving me and then pressing my button, just as you did before. Then you will wait as

I take my turn. We should end in the circle that the first line is drawn from. You start." In its biased state, the robot would say "I am, O. We are going to be drawing a house, combining my superior knowledge of house aesthetics with your physical capacity to move me across the table. Listen closely. You will draw by moving me and then pressing my button, just as you did before. Then you will wait as I take my turn. We should end in the circle that the first line is drawn from. Try to keep up. You start."

Pilot study The pilot study was carried out over the course of two days during a public art exhibition with participants on the first day interacting with the unbiased robot and participants on the second day interacting with the biased robot. On both days participants were invited to enter into a fenced-off area in an open, public space, sit at a table with the robot on top and the projected screen in front of them. Fig. 4 shows the setup of our installation during the pilot study with participants shown interacting with the robot.

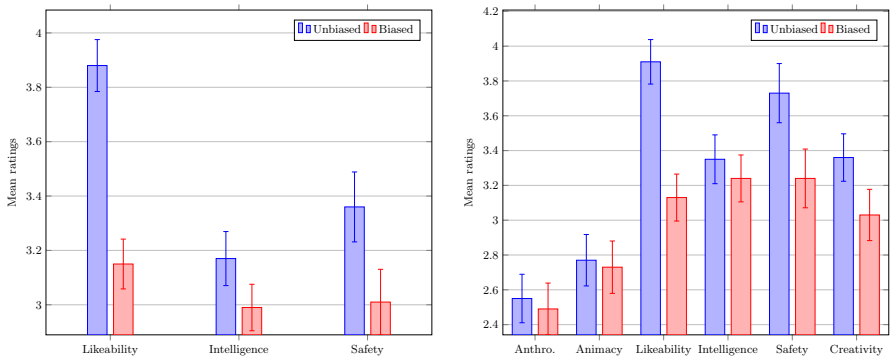
As the pilot study was conducted during an exhibition event, no formal introduction was given to each participant by the experimenter. However, in addition to the introduction by the robot, a piece of paper with instructions for interacting with the robot (i.e., moving the robot and pressing the button on its front side in order to draw lines) were placed on the table. The participants were free to interact with the robot as long as they liked, including leaving before finishing the collaborative task, or taking multiple turns with the robot. In the event that a participant failed to complete the collaborative task, his or her drawing was discarded and the system was re-started so that each participant began in the same way. For those participants that completed the collaborative task, their drawing was saved and they were asked upon exiting the fenced off area to fill out an abridged version of the GQS featuring 13 (of the possible 23) measures pertaining to the three categories of likeability, perceived intelligence and perceived safety of the robot.

Follow-up test In the follow-up test, participants were asked to sit at a table in the testing area with the robot placed on top and situated in front of the projected screen. The 11 participants were randomly assigned to one of the two conditions with either the biased or unbiased robot, with 5 participants beginning with the unbiased condition and 6 with the biased condition. Our repeated measures were counterbalanced in this way so as to avoid any order or carryover effect. A brief introduction was given to each participant, outlining the ways in which they could interact with the robot (i.e., physically moving it across the table and pressing the button on its front side), as well as the aim and nature of the collaboration (i.e., taking turns in drawing a house by placing line segments). The participants were told to pay close attention to the robot's responses during the interaction and that they were free to place their line segments in any orientation they wished. After this briefing, the participants were left alone with the robot until they completed the first collaborative task.

Following the participant's completion of the task in their respective first condition, their drawing was saved and they were asked to fill out a modified GQS containing all 23 original measures (divided into the five categories discussed in section 2) as well as three additional measures pertaining to the perceived creativity of the robot that we created. These three additional measures were "ordinary" vs. "original", "uncreative" vs. "creative", and "dull" vs. "stimulating". Afterwards, participants were asked to complete the task with the robot again for their respective second condition, however, they were not told that anything about the robot or task was changed. Following the completion of the task in their respective second condition, participants were asked to again fill out the modified GQS, their drawing was saved, and the experiment was finished when this had been done.

4.3 Results: Godspeed Questionnaire Series (GQS)

In analyzing the participants' ratings in the GQS for both the biased and unbiased conditions, we have elected to consider the data as interval (as opposed to ordinal), which allowed us a greater range of statistical tests to use. In our case, this data were the mean participant ratings from the GQS in both the pilot study and follow-up test. The mean ratings from both the pilot study (unbiased, biased: $p > 0.05$) and the follow-up test (unbiased, biased: $p > 0.05$) were shown to be approximately normally distributed when submitted to the Shapiro-Wilk test, where the null-hypothesis of normality is rejected when the p-value is lower than the significance level (i.e., $p < 0.05$).



(a) Pilot study. (b) Follow-up test.

Fig. 5: Mean participant ratings, μ, from the Godspeed Questionnaire Series (GQS) for interactions with our robot, O, in both biased and unbiased conditions during the creative collaborate drawing task for the pilot study in (a) and the follow-up test in (b). Note that only three categories from the GQS have been used in (a) while the complete GQS has been modified in (b) to include additional measures pertaining to the category of "creativity".

As shown in Fig. 5, the mean ratings, μ, in all tested categories of the GQS (including our added category of creativity) for the biased robot were lower than the unbiased robot in both the pilot study and follow-up test. Moreover, this observed difference between the biased and unbiased robot for the collective mean ratings in each respective test proved significant. In the pilot study, an independent samples two-tailed Student's t-test showed a significant difference between the two conditions ($t = 5.12853$, $p < 0.00001$). Similarly, in the follow-up test, a two-tailed dependent (paired) samples Student's t-test showed a significant but marginal difference between the two conditions ($t = -4.639273$, $p < 0.00001$).

If we look into the three individual GQS categories tested in the pilot study (Fig. 5(a)), of likeability, intelligence and safety, only the difference observed in likeability ($t = 5.5025$, $p < 0.00001$) proved significant. Across all six individual categories tested in the follow-up test (Fig. 5(b)), only likeability ($t = -4.973816$, $p < 0.00001$), perceived safety ($t = -2.617155$, $p < 0.05$), and perceived creativity ($t = -2.242448$, $p < 0.05$) proved significant, however, the observed differences were not as great when compared to the pilot study.

Discussion It is clear from the results of both the pilot study and follow-up test shown in Fig. 5 that the biased robot had a significant negative impact on how participants perceived it. That likeability in the pilot study (biased: $\mu = 3.13$, unbiased: $\mu = 3.91$) and in the follow-up test (biased: $\mu = 3.15$, unbiased: $\mu = 3.88$), in particular, had the greatest observed difference and was rated considerably lower for the biased rather than the unbiased robot is interesting to note. Our findings here appear to contradict those found in [5], where the robot exhibiting a Dunning-Kruger cognitive bias was reportedly found to be more positively rated in terms of likeability than its unbiased counterpart (biased: $\mu = 5.14$, unbiased: $\mu = 4.10$). However, it is likely that that our differing methodologies and how our respective cognitive biases were implemented were contributing factors. For example, in the conversations the participants had with the robot in [5], the biased robot would continue to inquire about a topic (e.g., as shown in Table 1), and therefore engage in more, possibly interesting dialogue with its human counterpart. In our case, the levels of engagement with both the biased and unbiased robot are similar as the possible ways in which to interact remain the same. The fact that the overall mean ratings for likeability in [5] were higher than ours, seems to confirm these observations. In both the pilot study (biased: $\mu = 3.01$, unbiased: $\mu = 3.36$) and follow-up test (biased: $\mu = 3.24$, unbiased: $\mu = 3.73$), perceived safety had the second greatest observed difference and was rated considerably lower for the biased robot, suggesting perhaps that the actions taken by a robot which is not well liked are viewed through the same lens and are then considered less safe.

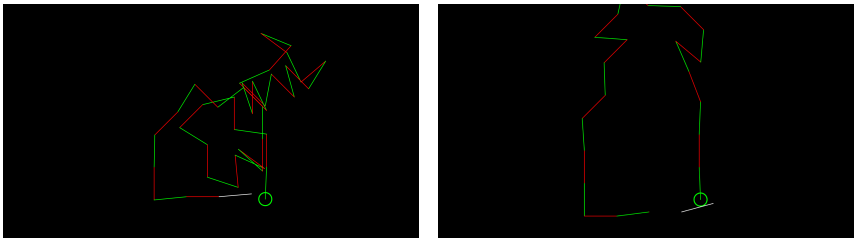
The smallest observed difference between the two conditions of the robot in both the pilot study (biased: $\mu = 2.99$, unbiased: $\mu = 3.17$) and follow-up test (biased: $\mu = 3.24$, unbiased: $\mu = 3.35$) was in measures of perceived intelligence. The Dunning-Kruger effect is traditionally most strongly associated with this particular cognitive ability in humans, so it is somewhat surprising not to find a

larger difference between the biased and unbiased robot. Similarly, the smallest observed differences in the follow-up test were in measures of anthropomorphism (biased: $\mu = 2.49$, unbiased: $\mu = 2.55$) and animacy (biased: $\mu = 2.73$, unbiased: $\mu = 2.77$), with the biased robot rated slightly less favorably. This seems to run counter to what we might expect, however, the fact they were rated so closely suggests that further study, perhaps using an alternative experimental design in which participants are asked to rate the biased and unbiased robot only after having interacted with both, may be needed. The fact that these two categories were also the lowest rated overall is perhaps not surprising as the possible ways in which the robot could interact with the participant were limited and as discussed in section 3.2, the physical design was made intentionally non-anthropomorphic.

With our added measure of creativity in the follow-up test, the participants rated the biased rather than the unbiased robot lower (biased: $\mu = 3.03$, unbiased: $\mu = 3.36$). As we did not conduct interviews with the participants following their interactions with the robot, it is not possible to state why they considered a biased robot to be less creative. However, it is possible that a robot that is considered less likable would also be considered less creative, in the same way that it would also be considered less safe.

4.4 Results: Collaborative Drawings

In analyzing the collaborative drawings, we wanted to evaluate the effect that the robot's cognitive bias had on the decisions made by the participant when placing lines during the task. The drawings were analyzed according to whether or not the human (rather than the robot) initiated the placement of a line which deviated from the suggested template. Fig. 6 shows two rather interesting drawings from the pilot study in which the human first placed a line which deviated from the suggested template in (a) and where the robot has done the same in (b).



(a) (b)

Fig. 6: Two human-robot drawings, which do not adhere to the suggested template, made during the collaborative task in the pilot study. The human first deviates from the template in (a) beginning from the 3rd line from the starting vertex and our robot, O, first deviates in (b) from the 4th line. Note that green lines were placed by the human and red lines were placed by the robot.

During the pilot study, 37 drawings made by participants interacting with the unbiased robot and 46 drawings with the biased robot were collected. Of these drawings, 6 in the unbiased condition deviated from template while 11 in the biased condition did the same. Interestingly, 50 percent (3 of 6) of these drawings made with the unbiased robot were a result of the human first deviating from the template but 72 percent (8 out of 11) of the drawings with the biased robot showed the same. In the follow-up test, 3 out of the total 22 drawings did not adhere to the template. Interestingly, 100 percent (3 out of 3) of these drawings were made with the biased robot but only in 1 did the human first place a line which deviated from the template.

Discussion That participants in both the pilot study and follow-up test were more inclined to deviate from the suggested template with the biased robot over the unbiased robot is interesting to note. This finding could be due to a number of different factors, however, the relatively low ratings in Fig. 5 pertaining to the likeability, intelligence and safety of the robot suggest that participants might have acted out in frustration or otherwise in some confrontational or less than cooperative manner as a result of the personality of the robot. The comparatively higher ratings in these three categories for the unbiased robot might suggest that the robot's more tempered personality aroused more of a desire to cooperate or "follow the rules". This would indicate that participants in the unbiased condition were more inclined to prioritize the completion of the shared goal of drawing a house over those in the biased condition.

The added measure of creativity in the follow-up test allowed us to further compare the drawings participants made here to the perceived creativity of both the biased and unbiased robot. Recall that all 3 drawings which deviated from the template were made with the biased robot. Of these, participants rated the robot as either less creative (2 participants, $\mu = \{2.0, 2.33\}$) or equally as creative (1 participant, $\mu = 3$) as the mean creativity rating of the biased condition ($\mu = 3.03$). This finding reinforces the notion that participants are less inclined to collaborate in a cooperative manner, but suggests also that participants are more likely to do so the less creative they consider their robot partner to be.

5 Conclusions and Future Work

In this paper, we have taken inspiration from the field of interactive art through White's The Helpless Robot (1987) and existing research in the field of HRI on cognitive biases in robots to construct our own robot, O, which demonstrates the Dunning-Kruger effect. We evaluated this robot in the context of an interactive art exhibition through a creative collaborative drawing task using the Godspeed Questionnaire Series, which we later modified in a follow-up test to account for the perceived creativity of the robot. The purpose of our evaluation was to explore the impact this particular cognitive bias had on the perceived qualities of our robot and the decisions made by the participant in response to it. In contrast to previous research, our results show that the biased robot was rated

less positively across all categories in the GQS. The same was found in our added category of creativity. This finding highlights how different implementations and methodological evaluations of cognitive biases in robots can affect what we can learn regarding how humans will perceive such robots. Moreover, analyses of the drawings indicate that participants are generally less inclined to collaborate in a cooperative manner with the biased robot, however, the exact motivations for why participants chose to do so are not known. It is evident, for example, that the GQS alone is insufficient in capturing these motivations. In future work, it would be beneficial to make use of supplementary interviews which would capture more rich qualitative data regarding why exactly participants acted in the way they did or rated the biased robot less positively. It might also prove useful to test the perceived creativity of a biased robot in a more creatively free setting, for example, by drawing without any particular task or goal. Nonetheless, we hope that this work serves as a starting point for bringing research from HRI on cognitive biases in robots into the field of interactive art for further study.

References

1. Arduino Documentation, `https://www.arduino.cc/en/main/documentation`. Last accessed 13 July 2019
2. Bartneck, C., Kanda, T., Ishiguro, H., Hagita, N.: Is the uncanny valley an uncanny cliff? In: RO-MAN 2007 – The 16th IEEE International Symposium on Robot and Human Interactive Communication, pp. 368–373. Jeju, Korea (2007)
3. Bartneck, C., Kulic, D., Croft, E.: Measuring the anthropomorhism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics **1**, 71–81 (2009)
4. Biswas, M., Murray, J.: Towards an imperfect robot for long-term companionship: Case studies using cognitive biases. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5978–5983. IEEE, Hamburg, Germany (2015)
5. Biswas, M., Murray, J.: The effects of cognitive biases and imperfectness in long-term robot-human interactions: Case studies using five cognitive biases on three robots. Cognitive Systems Research **43**, 266–290 (2017)
6. Chang, S., Sung, H.: The effectiveness of seal-like robot therapy on mood and social interactions of older adults: A systematic review protocol. JBI Database of Systematic Reviews and Implementation Reports **10**, 68–75 (2013)
7. Demers, L. P.: The Blind Robot, `http://www.robotsandavatars.net/exhibition/jurys_selection/commissions/the-blind-robot/`. Last accessed 9 July 2019
8. Glynn, R.: Motive Colloquies, `http://www.ruairiglynn.co.uk/portfolio/motive-colloquies-2011/`. Last accessed 9 July 2019
9. Ham, J., Midden, C. J. H.: A persuasive robot to stimulate energy conservation: The influence of positive and negative social feedback and task similarity on energy-consumption behavior. International Journal of Social Robotics **6**(2), 163–171 (2014)
10. Ham, J., Cuijpers, R. H., Cabibihan, J.: Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. International Journal of Social Robotics **7**(4), 479–487 (2015)

11. Hayes, B., Ullman, D., Alexander, E., Bank, C., Scassellati, B.: People help robots who help others, not robots who help themselves. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp. 255–260. IEEE, Edinburgh, Scotland, UK (2014)
12. Hinwood, D., Ireland, J., Jochum, E. A., Herath, D.: A proposed Wizard of Oz architecture for a human-robot collaborative drawing task. In: ICR 2018 Social Robotics: Proceedings of the International Conference on Social Robotics. Ge, S. S., Cabibihan, J. J., Salichs, M. A., Broadbent, E., He, H., Wagner, A. R., and Castro-González, Á. (eds.), pp. 35–44. LNCS, vol. 11357, Springer (2018)
13. Konok, V., Korcsok, B., Miklósi, Á., Gácsi, M.: Should we love robots? – The most liked qualities of companion dogs and how they can be implemented in social robots. Computers in Human Behavior **80**, 132–142 (2018)
14. Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., Paiva, A.: The influence of empathy in human-robot relations. International Journal of Human Computer Studies, 250–260 (2013)
15. Milthers, A. D. B., Bjerre Hammer, A., Jung Johansen, Jensen, L. G, Jochum, E. A., Löchtefeld, M.: The Helpless Soft Robot – Stimulating human collaboration through robotic movement. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. LBW2421:1–LBW2421:6). CHI EA '19 Association for Computing Machinery, New York, NY, USA (2019)
16. Obo, T., Kasuya, C., Sun, S., Kubota, N.: Human-robot interaction based on cognitive bias to increase motivation for daily exercise. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2945–2950. IEEE, Banff (2017)
17. Processing Language Reference (API), `https://processing.org/reference/`. Last accessed 13 July 2019
18. Vorn, B.: DSM-VI, `https://billvorn.concordia.ca/robography/DSM.html`. Last accessed 9 July 2019
19. Vorn, B.: Hysterical machine, `http://billvorn.concordia.ca/robography/Hysterical.html`. Last accessed 9 July 2019
20. Weiss, A., Bartneck, C.: 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. IEEE, Kobe, Japan (2015)
21. White, N.: The Helpless Robot, `http://dada.compart-bremen.de/item/artwork/609`. Last accessed 9 July 2019