

Influencing Human Performance: Investigating the Effect of Non-humanoid Robot Feedback on Task Performance

Mohammed Shabaj Ahmed
hi20047@bristol.ac.uk
University of Bristol
Bristol, UK

Manuel Giuliani
manuel.giuliani@hs-kempten.de
Kempten University of Applied
Sciences
Kempten, DE

Paul Bremner
paul2.bremner@uwe.ac.uk
University of the West of England
Bristol, UK

ABSTRACT

This paper reports the findings of an online between subject study that investigated the effectiveness of a non-humanoid socially assistive robot in providing positive reinforcement feedback to aid in improving performance on a cognitively demanding task. Four different feedback conditions were used, including verbal, expressive, neutral, and text-based feedback, to identify which type of feedback could positively influence behaviour. Results showed no significant differences in task performance, perceived workload or robot perception.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in interaction design.*

KEYWORDS

Positive reinforcement feedback, non-humanoid socially assistive robot, persuasion and influence, task engagement and motivation

ACM Reference Format:

Mohammed Shabaj Ahmed, Manuel Giuliani, and Paul Bremner. 2024. Influencing Human Performance: Investigating the Effect of Non-humanoid Robot Feedback on Task Performance. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion)*, March 11–14, 2024, Boulder, CO, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3610978.3640651>

1 INTRODUCTION

The field of human-robot interaction is rapidly advancing with the introduction of socially assistive robots (SAR) in household settings. These robots are designed to be smaller and integrate into our homes, and they have the potential to influence our daily behaviour through subtle persuasive mechanisms. This study investigates the effectiveness of non-humanoid SARs in improving performance in demanding cognitive tasks. Given their significance in fields like education [12], health [16], and rehabilitation [17], we focus on the impact of various feedback mechanisms (verbal, expressive, neutral, and text-based) on user motivation and task performance.

The basis of our research is centred around reinforcement, a core concept in behaviourism, the study of how environmental stimuli can condition individuals to learn new behaviours. Our

study focused on positive reinforcement strategies based on our review of the existing literature. Specifically, we introduced feedback to encourage the desired behaviour. Overall, the mechanisms of behaviour change being researched are social influence and reinforcement feedback.

We conducted an online between-subject study. Participants solved puzzles requiring them to count the number of boxes containing a specified shape (see Figure 1). A combination of objective and subjective measures was employed to evaluate the effects of each feedback type. However, considering that many in-person studies struggle to get enough participants, we aimed to investigate the viability of assessing the impact of robot feedback in an online study, where we have easier access to a large population to sample from.

Our research question was whether social feedback from a non-humanoid robot affects task performance and user perceptions. We hypothesised that positive reinforcement delivered by a non-humanoid social robot would improve the task performance and positively influence user attitudes towards the robot.

The robot chosen to provide feedback for this study was the Digital Dream Labs Vector 2.0¹, a palm size non-humanoid companion robot (see Figure 2). Its small size, affordability, expressive capabilities, and publicly available software development kit made it an ideal candidate for our experimental setup.

Through this study, we aim to contribute to understanding how non-humanoid SARs can be effectively utilised to influence human behaviour, particularly in settings that demand high cognitive engagement. The insights gained could have significant implications for the development of future SARs and their application in various fields, including education, therapy, and workplace productivity.

2 BACKGROUND AND RELATED WORK

Our research extends previous studies on using Socially Assistive Robots (SAR) for persuasive contexts and focuses on how SAR can alter human mental states (attitudes, beliefs, or behaviours) to get people to work harder on a cognitively demanding and repetitive task. Persuasion in this context is defined as the process of changing the mental state of the person being persuaded [11] through communication and interaction.

A robot's influence on humans depends on being perceived as a social actor [7], a concept used to describe actors that engage in interactions and influence others. In the context of SAR, being perceived as a social actor enables the robot to provide effective social feedback, a term used to describe the influence of behaviour or speech on others. When technology presents itself as social, it



This work is licensed under a Creative Commons Attribution International 4.0 License.

¹<https://www.digitaldreamlabs.com/products/vector-robot>

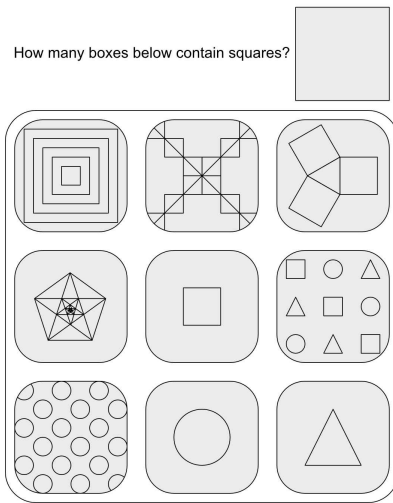


Figure 1: Example of a CAPTCHA puzzle used during the study

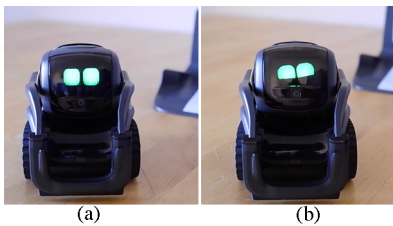


Figure 2: Screenshot of the video of the Vector robot in (a) verbal and neutral feedback condition and (b) example of expressive feedback condition

elicits social responses from humans [13]. This phenomenon has been studied as human susceptibility to flattery from computers just as they are susceptible to flattery from other humans [2].

Previous research by Ham and Midden [8] investigated the effect of robot-generated feedback on energy saving behaviour in a simulated washing machine task. Their study contrasted the impact of positive or negative social feedback against factual feedback displayed on a gauge. They discovered that social feedback, particularly negative feedback, had stronger persuasive effects to save energy than factual feedback. This aligns with our hypothesis about the influence of social feedback on task performance. We have extended their findings by exploring if social feedback from a non-humanoid robot can motivate individuals to exert more effort on a task.

However, contrasting findings by Akalin et al. [1] suggest that participants preferred the robot that provided positive or flattery feedback even if it did not correlate to objective performance measures. This highlights a mismatch between preferred feedback and effective feedback, a gap our research looks to explore further.

Similarly, Chidambaram et al. [5] investigated compliance with robot suggestions in a desert survival task. Participants had to

choose the best tools from a selection to survive in a desert. Their findings suggest that nonverbal immediacies are crucial in persuasive interactions, including bodily cues such as gaze, gesture, and facial expression, and verbal cues such as tone and expressions. We summarised this type of feedback as expressive feedback and will be looking to compare if there are any noticeable differences in performance between expressive and verbal feedback conditions.

We used the framework of feedback dimensions proposed by Pritchard and Montagno [14] to identify the dimensions relevant to our current study. They include positive versus negative feedback, timing of feedback, specificity, absolute versus comparative feedback, and the source’s authority. Our approach incorporates positive reinforcement with positive feedback, associated with increased intrinsic motivation [3]. Additionally, Swift-Spong et al. [15] emphasised the impact of self-comparative, other-comparative, and no comparative feedback, suggesting that the feedback should be biased towards self-comparative feedback.

In summary, our review of existing literature highlights the important role of feedback dynamics and nonverbal immediacies in the context of Socially Assistive Robots (SAR). By examining how various forms of positive feedback influence human behaviour and decision-making, we aim to gain practical insights into the persuasive power of SARs. Our study seeks to build on these findings by exploring the specific impact of non-humanoid robot feedback on motivation and persistence in repetitive tasks.

3 METHODOLOGY

3.1 Study setup

We employed a between-subject online comparative study using the Digital Dream Labs Vector 2.0 with four study conditions. Participants were recruited through an online recruitment portal (Prolific) and were reimbursed for their time.

Participants were asked to solve spatial reasoning puzzles, as shown in Fig. 1, inspired by the Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). We used greyscale puzzles with combinations of triangles, circles, and squares, which were crafted to be universally accessible, avoiding language or mathematical barriers. The repetitive nature of the task aimed to evaluate how participants adapted to the robot’s feedback.

The study was advertised to participants as a test of the effectiveness of a new type of CAPTCHA puzzle. Participants were told the robot would guide them through the survey and was framed as an invigilator. They were not told that their task performance would be correlated to the robot’s feedback.

3.2 Experimental procedure

After consenting to the study terms, tutorial videos of the robot were shown, introducing the study and explaining the rules for solving the puzzles. The introductory section involved practising with two to three trial puzzles, where the robot provided tailored feedback based on the correctness of responses.

In the main phase of the study, participants were assigned to one of four distinct conditions (described in section 3.3). Feedback, when provided, was always positive and encouraging regardless of the correct answer. Additionally, all answers to the CAPTCHAs were presented in multiple-choice format.

The videos were played automatically and could not be paused, rewound, fast-forwarded or replayed. Additionally, there was a time delay before the next page button was displayed to ensure the participant watched most of the videos before going to the next page.

After completing the repetitive task at the end of the study, the participants were asked to complete two questionnaires (Section 3.6). Finally, participants had the chance to provide their opinion of the robot and the study in general in a free-form text field.

3.3 Robot feedback types

Our study was designed with four distinct study conditions to examine the effects of different types of feedback on participant performance using a SAR.

- **Verbal Utterance Condition:** The robot provided positive verbal feedback. This feedback was aimed at reinforcing correct responses or encouraging effort.
- **Expressive Condition:** Feedback was delivered through non-verbal expressions, combining facial expression, gestures and sound effects, because they are a valuable addition to recognising emotion in robots [6].
- **Neutral Condition:** The robot only offered instructional guidance without any feedback on the correctness of the participants' responses.
- **Text Condition:** Feedback similar to that in the verbal utterance condition was presented as text on a screen without the robot's presence.

Feedback videos, lasting between three and six seconds, were centrally displayed against a white background. In the Verbal and Text feedback conditions, we employed three types of positive feedback, supported by the literature on their effectiveness. These included general feedback (e.g., well done), self-comparative (e.g., well done, that was all correct) and other comparative (e.g., you are doing better than average).

We employed neutral and text conditions as control groups; their purpose was to isolate if the presence of the robot alone or feedback alone affects performance.

3.4 Robot interaction model

In each of the robot feedback videos, the robot starts the interaction and demonstrates engagement by looking directly at the camera as if it were looking directly at the participant, in addition to turning to look in the direction it wanted the participant to look.

Initially, a continuous reinforcement schedule was used for the trial puzzles. This schedule has been shown to have the fastest time to learn a new behaviour [10]. This was followed by a fixed interval reinforcement schedule for the main portion of the study. We did this to avoid giving participants too frequent feedback, which could have made them aware of the study's purpose. As a result, participants received feedback after solving three puzzles.

3.5 Population and sampling, participant recruitment

A power analysis based on the assumptions of normality and variance was conducted to ensure statistical reliability for a one-way

ANOVA. To detect a medium effect size (Cohen's $f = 0.25$), with an alpha error of 0.05 and a power of 80%. This calculation led to a required sample size of 180 participants, >45 participants in each study condition.

200 participants, consisting of 100 males and 100 females, with a mean age of 27 ($SD = 7.1$), were recruited through Prolific, with access granted only to those on desktop computers. The study took roughly 15-20 minutes, and participants were compensated £2.5 for their time. Participants were allocated to a condition using pseudo-randomisation. This study was approved by the Ethics Committee of the Faculty of Engineering at the University of the West of England (Approval number FET-2122-103).

3.6 Experimental measures

In our study, we assessed the effect of robot feedback on task performance using a combination of objective and subjective measures. The objective metrics were the number of puzzles solved, the number of correct answers and the time taken to solve them. Subjectively, we evaluated participants' workload using the NASA Task Load Index [9] and their perception of the robot using the Godspeed questionnaire [4] to measure perceived robot anthropomorphism, animacy, likability, and intelligence. Additionally, participants were asked for their open-ended opinions of the robot and the overall study design.

4 RESULTS

4.1 Attention Check

There were 50 participants per study condition. The puzzles were not considered difficult, resulting in participants scoring >75% on average. Those who scored less than 75% were excluded from the study. Additionally, the average time spent on each puzzle across all conditions was 16.7 seconds (lower IQR = 7.71); we excluded any participant who spent less than eight seconds per puzzle for not paying attention. The remaining participants for each study condition were 40 for the verbal, 39 for the expressive, 40 for the neutral, and 42 for the text condition.

4.2 Robot persuasiveness

Before conducting the main analysis, the dataset was tested to assess the assumptions of parametric testing. Levene's test indicated homogeneity of variances across conditions for all measures. However, the Shapiro-Wilk test revealed significant deviations from normality in several conditions for different measures. Due to these normality violations, non-parametric methods were chosen for the subsequent analysis.

The Kruskal-Wallis test was used to evaluate differences across the four feedback conditions (Text, Expressive, Neutral, Verbal) for each measure. This analysis revealed a statistically significant difference in animacy scores ($H = 9.39$, $p = .025$). No significant differences were found for other measures, including task performance ($p = .770$), anthropomorphism ($p = .390$), likeability ($p = .289$), and NASA Task Load Index ($p = .278$).

Following the significant result in the animacy measure, a post-hoc Dunn's test with Bonferroni correction was conducted to explore pairwise group differences. The post-hoc analysis did not indicate any significant pairwise differences among the feedback

conditions, suggesting that the observed variations in animacy scores were not distinct enough to be differentiated between specific pairs of groups.

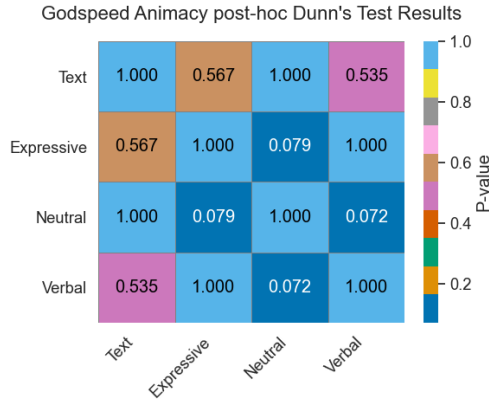


Figure 3: Heatmap illustrating the pairwise post hoc Dunn test for animacy across all study conditions

5 DISCUSSION

This section analyses the key findings from the study, evaluating their implications and exploring potential reasons behind the observed outcomes.

5.1 Task performance

Our analysis did not reveal significant differences in task performance between the experimental conditions. The feedback frequency, the nature of the task, and participant engagement levels are potential influencing factors to consider.

Several participants commented that they felt the robot was trying to encourage them to solve more puzzles:

"I could immediately tell the test was actually about how the robot affected my decisions rather than the captcha itself."

The participants above stopped the study early once they realised the robot was trying to persuade them to solve as many puzzles as possible. It is unclear how many other participants may have figured out the purpose of the study and acted in accordance with the null hypothesis.

The type of puzzles used in our study (visual puzzles) may have influenced participants' performance. Participants who enjoyed solving puzzles may have performed well regardless of the feedback they received:

"I think I enjoyed it more because I like solving puzzles. It was fun."

Visual puzzles were chosen over math or literacy problems to prevent proficiency in the subject from affecting how well someone performed. It is not easy to know if the data recorded is influenced by intrinsic motivation or the feedback from the robot. Future studies may consider using different types of tasks to avoid this potentially confounding variable.

5.2 Perception of the robot

We used the Godspeed questionnaire subscales for anthropomorphism, animacy, likability, and perceived intelligence to assess participants' perceptions of the robot during the study. Kruskal-Wallis analysis revealed a statistically significant difference in animacy between study conditions. However, post hoc analysis showed that the significance was above our significance threshold of 0.05. Although the evidence is insufficient to conclude a difference in animacy, it does suggest that the type of feedback the robot provides has a subtle influence on the perception of a non-humanoid robot's liveliness.

5.3 Perceived workload

Participants perceived workload across each study condition was relatively low, as observed by the group average scores, all falling in the lower quartile of the NASA Task Load Index. Moreover, the average perceived workload was similar across conditions, suggesting that the type of feedback provided did not significantly affect participants' perceived workload.

5.4 Limitations

There may be a combination of factors that contributed to the lack of evidence supporting our hypothesis. The following are some key areas that need to be examined:

- (1) Generalisability: Paid participants were incentivised to take part and may have optimised for receiving payment. Future studies could consider seeking a more diverse participant pool, possibly through in-person recruitment.
- (2) Feedback schedule: The frequency of feedback may have influenced the results, as participants may have identified the purpose of the study. Future studies could implement a variable feedback schedule in addition to having the robot always present on the screen, which may make the process of receiving feedback more natural.
- (3) Task Engagement: The enjoyment derived from solving puzzles may have overshadowed the feedback's impact. Progressively increasing puzzle difficulty, increasing the number of puzzles, and assessing the intrinsic motivations of the participants during the study could help identify and reduce this effect.

6 CONCLUSION

In conclusion, this study was an online study that aimed to evaluate the effectiveness of non-humanoid SAR in providing positive reinforcement feedback to improve task performance on a cognitively dependent task.

No significant differences in task performance or perceived workload were observed across the different feedback conditions. However, the type of feedback provided did slightly impact how participants perceived the robot's level of animacy, although the significance level exceeded our alpha level in a post hoc analysis.

7 ACKNOWLEDGMENTS

This work was supported by the EPSRC Centre for Doctoral Training in Future Autonomous and Robotic Systems, grant no. EP/S021795/1.

REFERENCES

- [1] Neziha Akalin, Annica Kristoffersson, and Amy Loutfi. 2019. The Influence of Feedback Type in Robot-Assisted Training. *Multimodal Technologies and Interaction* 3, 4 (Dec. 2019), 67. <https://doi.org/10.3390/mti3040067> Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
- [2] B. J. Fogg and Clifford Nass. 1996. Silicon Sycophants: The Effects of Computers that Flatter | Elsevier Enhanced Reader. (Oct. 1996). <https://doi.org/10.1006/ijhc.1996.0104>
- [3] Rokhsareh Badami, Mohammad Vaezmousavi, and Mehdi Namazizadeh. 2011. Feedback After Good Versus Poor Trials Affects Intrinsic Motivation. *Research Quarterly for Exercise and Sport* 82 (June 2011), 360–364. <https://doi.org/10.1080/02701367.2011.10599765>
- [4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. <https://doi.org/10.1007/s12369-008-0001-3>
- [5] Vijay Chidambaram, Yueh-Hsuan Chiang, and Bilge Mutlu. 2012. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI '12*. ACM Press, Boston, Massachusetts, USA, 293. <https://doi.org/10.1145/2157689.2157798>
- [6] Sandra Costa, Filomena Soares, and Cristina Santos. 2013. Facial Expressions and Gestures to Convey Emotions with a Humanoid Robot. In *Social Robotics (Lecture Notes in Computer Science)*, Guido Herrmann, Martin J. Pearson, Alexander Lenz, Paul Bremner, Adam Spiers, and Ute Leonards (Eds.). Springer International Publishing, Cham, 542–551. https://doi.org/10.1007/978-3-319-02675-6_54
- [7] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On seeing human: A three-factor theory of anthropomorphism. *Psychological Review* 114, 4 (2007), 864. <https://doi.org/10.1037/0033-295X.114.4.864> Publisher: US: American Psychological Association.
- [8] J. Ham and C. J. H. Midden. 2014. 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 (April 2014), 163–171. <https://doi.org/10.1007/s12369-013-0205-z>
- [9] Sandra G. Hart. 1986. NASA Task Load Index (TLX). <https://ntrs.nasa.gov/citations/20000021487> NTRS Author Affiliations: NASA Ames Research Center NTRS Document ID: 20000021487 NTRS Research Center: Ames Research Center (ARC).
- [10] David Hulac, Nicholas Benson, Matthew C. Nesmith, and Sarah Wollersheim Shervey. 2016. Using Variable Interval Reinforcement Schedules to Support Students in the Classroom: An Introduction With Illustrative Examples. *Journal of Educational Research and Practice* 6, 1 (Nov. 2016). <https://doi.org/10.5590/JERAP.2016.06.1.06>
- [11] Rahul Radhakrishnan Iyer and Katia Sycara. 2019. An Unsupervised Domain-Independent Framework for Automated Detection of Persuasion Tactics in Text. *arXiv:1912.06745 [cs]* (Dec. 2019). <http://arxiv.org/abs/1912.06745> arXiv: 1912.06745.
- [12] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human-Computer Interaction* 19, 1 (June 2004), 61–84. https://doi.org/10.1207/s15327051hci1901&2_4
- [13] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94*. ACM Press, Boston, Massachusetts, United States, 72–78. <https://doi.org/10.1145/191666.191703>
- [14] Robert D. Pritchard and Raymond V. Montagno. 1978. *Effects of Specific Vs. Nonspecific and Absolute Vs. Comparative Feedback on Performance and Satisfaction*. [Department of Defense], Department of the Air Force, [Air Force Systems Command], Air Force Human Resources Laboratory. Google-Books-ID: 71v1r2whbtAC.
- [15] Katelyn Swift-Spong, Elaine Short, Eric Wade, and Maja J Mataric. 2015. Effects of comparative feedback from a Socially Assistive Robot on self-efficacy in post-stroke rehabilitation. In *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. 764–769. <https://doi.org/10.1109/ICORR.2015.7281294> ISSN: 1945-7901.
- [16] Katie Winkle, Severin Lemaignan, Praminda Caleb-Solly, Paul Bremner, Ailie Turton, and Ute Leonards. 2020. In-Situ Learning from a Domain Expert for Real World Socially Assistive Robot Deployment. In *Robotics: Science and Systems XVI*. Robotics: Science and Systems Foundation. <https://doi.org/10.15607/RSS.2020.XVI.059>
- [17] Katie Winkle, Séverin Lemaignan, Praminda Caleb-Solly, Ute Leonards, Ailie Turton, and Paul Bremner. 2019. Effective Persuasion Strategies for Socially Assistive Robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 277–285. <https://doi.org/10.1109/HRI.2019.8673313> ISSN: 2167-2148.