Enhancing Cognitive Training: Investigating the Impact of a Suggestion-Offering Robot on Performance and Satisfaction*

Flavio Ruggiero[1], Marco Matarese[2,*], Alessandra Sciutti[2] and Mariacarla Staffa[1]

[1]University of Naples Parthenope, Centro Direzionale of Naples, Isola C4, 80143, Naples, Italy.
[2]Italian Institute of Technology, Via Morego 30, 16163, Genoa, Italy.

Abstract

Advancements in technology have created opportunities to enhance cognitive training through robotic assistance, offering personalized suggestions. This paper investigates how such robots impact performance and satisfaction in cognitive training. Understanding their effects holds significance for psychology, human-computer interaction, and education. Through a review of relevant literature and empirical study, insights into human-robot interaction dynamics and cognitive enhancement are gained. This exploration contributes to optimizing human cognitive abilities in assistive settings. Our results demonstrate that robot tutors' competence and periodic help offerings are insufficient unless human tutees recognize the task's difficulty.

Keywords

Human-Robot Interaction, Robot Tutoring, Cognitive Training, Assistive Robotics

1. Introduction

In the realm of cognitive training, advancements in technology have ushered in new opportunities to enhance human performance and satisfaction. Among these advancements, the integration of robotic assistance stands out as a promising avenue. Robots equipped with artificial intelligence (AI) capabilities not only offer practical assistance but also possess the potential to provide personalized suggestions tailored to individual needs. This intersection of technology and cognitive training prompts an intriguing question: How does the presence of a suggestion-offering robot influence individuals' performance and satisfaction in cognitive training tasks?

This paper delves into the exploration of this question, aiming to shed light on the implications of integrating such robotic systems into cognitive training environments. By examining the effects of a suggestion-offering robot on both performance metrics and subjective satisfaction levels, we seek to uncover valuable insights into the dynamics of human-robot interaction within the context of cognitive enhancement. Understanding the impact of a suggestion-offering robot on cognitive training tasks holds significance for various fields, including psychology, human-computer interaction, and education. Insights gleaned from this investigation could inform the design of future cognitive training programs and the development of more effective human-robot collaboration frameworks.

In this paper, we will first review relevant literature on cognitive training methodologies, human-robot interaction, and the influence of technology on cognitive performance. Subsequently, we will outline the methodology employed in our study, including the experimental design, participant recruitment, and data collection procedures. We will then present and analyze the results obtained, discussing implications and potential avenues for future research. Through this exploration, we aim to contribute to the growing body of knowledge concerning integrating robotics in cognitive enhancement endeavors, ultimately striving towards optimizing human cognitive abilities in assistive settings.

2. Related Works

The utilization of robots for training and assistance has experienced a notable surge, spanning various application areas such as education [1] and cognitive training customized for individuals with specific requirements, including those diagnosed with autism and the elderly [2, 3]. The manifold benefits of these applications are apparent, encompassing capabilities such as enabling concurrent usage at home under remote supervision by instructors or healthcare providers and facilitating personalized and adaptable training programs tailored to meet individual needs.

Nonetheless, a significant challenge arises in finding the ideal equilibrium between the robot's assistance and the individual's active participation. This predicament is widely acknowledged in the field of robotics rehabilitation, where interventions must carefully balance support to prevent inadvertently encouraging passivity (or idleness) [4]. Offering on-demand assistance is a common strategy, but it carries the risk of improper system usage, leading to suboptimal utilization of assistance, including aversion or misuse [5]. Even within the context of robot-assisted gaming, there is considerable variation in individuals' reliance on technology for support, influenced by factors such as their comfort with technology and personality traits [6, 7]. It is imperative to address instances where individuals overly depend on assistance, as it can hinder the learning process. Therefore, it is essential to understand the impact of a suggestion-offering robot on users and to understand if it can be considered a helpful supporter for clinicians and patients in such tasks. This study endeavors to explore the quantitative and qualitative impacts of a suggestion-offering robot in the completion of a cognitive task by aiming to assess how it influences both the participants' learning outcomes and their perception of the robot as a helper, as well as their inclination to use it again in the future.

Robots for Humans Workshop at the International Conference on Advanced Visual Interfaces, June 03–07, 2024, Arenzano, Italy.

* You can use this document as the template for preparing your publication. We recommend using the latest version of the ceurart style.

*Corresponding author.
flavio.ruggiero@studenti.uniparthenope.it (F. Ruggiero); marco.matarese@iit.it (M. Matarese); alessandra.sciutti@iit.it (A. Sciutti); mariacarla.staffa@uniparthenope.it (M. Staffa)
0000-0003-1719-3745 (M. Matarese); 0000-0002-1056-3398 (A. Sciutti); 0000-0001-7656-8370 (M. Staffa)

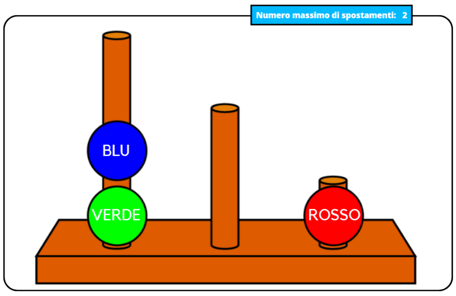

Figure 1: Computerized version of the London Tower test in Android.

3. Materials and Methods

3.1. Tower of London Test (ToL)

The Tower of London (ToL) test is a cognitive test developed in the 1980s by Tim Shallice and Rosaleen A. McCarthy. The test aims to evaluate the capacity for strategic decision-making and effective planning to solve tasks by anticipating and considering the consequences of actions on interconnected elements. The interdependence of elements in complex problems mirrors situations in daily life. The ToL test presents a graded difficulty problem where participants must move perforated balls arranged on a structure to achieve a new configuration, requiring the adoption of appropriate strategies.

Specifically, three operations are vital: (a) formulating a general plan, (b) identifying and organizing sub-goals in a sequence of movements, and (c) storing both sub-goals and the general plan in working memory [8, 9, 10]. Shallice challenges the Supervisory Attentional System (SAS) as a central mechanism in the prefrontal cortex, directing attention towards necessary sub-goals and shifting attention from sub-goals to the general scheme [10].

In the classic version of Shallice the material consists of three pegs of different lengths mounted on a wooden structure and three balls of different colors (red, green, and blue), which are inserted. The test consists of a series of twelve tests of increasing gradual difficulty depending on the number of moves that must be executed to arrive at the solution. In this work, we propose a computerized version of the test consisting of a support with three vertical posts of different heights and three colored balls (green, red, and blue), as shown in Figure 1. The test was displayed on a Pepper robot's tablet as shown in Figure 2.

The objective of the test is to move the balls from an initial configuration to a diverse target configuration, following four rules:

1. Only one ball can be moved at a time.
2. Only one ball can be moved from one post to another to prevent the user from placing the balls on the table or having more than one in his hand at a time.
3. It is possible to insert only one ball on the smaller post, two on the intermediate, and three on the larger one.
4. The maximum number of movements allowed to solve a problem can not be exceeded

To calculate the patient's final score at the end of the test, it is necessary to consider the number of problems solved correctly, the number of moves used to solve them, and the time spent to solve the tasks.

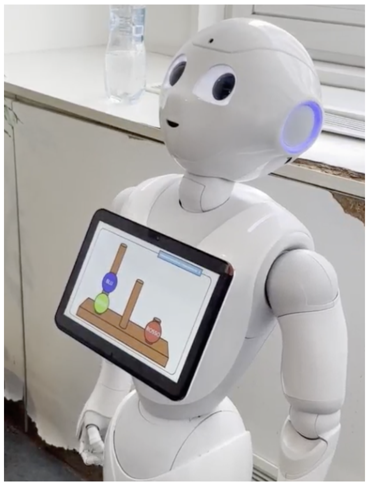

Figure 2: Pepper Robot during the administration of the test.

In each trial, the subject is asked to move the balls across a display to reach the configuration in the upper half of the screen. Participants dragged the ball with their fingers from the free position to one of the other rods to move it. Incorrect movements were recorded on the computer as non-responses. Subjects are asked to solve the task in the minimum possible number of moves.

3.2. Stimuli

The characteristics of the task we used allow for several ways of administration and types of measurements. In this study, we aim to investigate to what extent the adjustable social autonomy of a humanoid social robot employed to administer the London Tower test can impact the users' performance.

Considering this objective, we examined the effect of Suggestion-Offering Robot users. Specifically, to prevent help aversion, the robot proactively asks the users whether they needed a suggestion every 10 seconds. However, we instructed both the participants that they could ask the robot for help in every step of the task.

3.3. Measures

For the evaluation of the test performance, the following metrics will be considered:

- **Time**: time to accomplish the task.
- **Number of Moves (NMOV**: total number of moves performed to complete the problem.
- **Number of Successes (S)**: the number of completed problems.
- **Number of Suggestions requested (NS)**: the number of success w.r.t. the total number of tests.
- **Performance Measure (PM)**: the number of success w.r.t. the total number of tests.

Before and after the interaction with the robot, we submitted questionnaires to participants to collect self-reported measures about the HRI and the participants themselves. In the pre-experiment questionnaire, we submitted participants the Big Five personality traits test [11] and the Sense of Agency scale [12]. Moreover, we collected some items about their perception of the robot, such as the Godspeed questionnaires [13] and the Inclusion of Others in the Self test (IOS) [14]. We also submitted these latter two during the post-experiment questionnaire to measure the change in participants' perception of the robot.

4. Experimental Procedure

4.1. Procedure

The ToL test is extensively used in the context of cognitive training for patients with neurodegenerative disorders, such as Parkinson's disease [15]. Indeed, cognitive training is a promising nonpharmacological treatment option for patients with dementia [16, 17]. Although testing on patients was our goal, we wanted to test the protocol with healthy participants to ensure the feasibility of the study. Thus, we asked 20 computer science students at the University of Naples Parthenope to participate in the study.

In our experimental procedure, the ToL Test is administered by the Pepper robot. After an initial phase in which the robot provides participants with an explanation of the rules, the robot shows a test configuration to be used to verify understanding of the structure of the test by administering it subsequently. At this point, the user views the configuration to be replicated on Pepper's tablet (Figure 2), reproduces the one included, and continues in this way until all the configurations provided by the humanoid are exhausted (12 configurations for each participant). During the ToL test execution, the social robot occasionally offers move suggestions to users.

Namely, participants underwent three experimental phases:

- **Baseline**: a practice phase where requesting suggestions is not allowed, aiming to help the user become familiar with the test.
- **Training**: an intermediate phase in which participants train to complete the test where it is allowed to request suggestions.
- **Assessment**: a final phase where requesting suggestions is not allowed, and the user, with the experience gained from the previous two phases, should demonstrate greater proficiency in completing the test.

4.2. Robot's Suggestions

We needed an expert robot to provide useful and correct on-demand suggestions for our training phase. For this purpose, we designed a graph-based representation of the ToL test. Subsequently, we implemented a solving algorithm that used a breadth-first search (BFS) strategy to find the optimal solution for each ToL test. Starting from the initial ToL configuration representation, the BFS algorithm could find the optimal path to the representation of the final one. Although a BFS-based resolution of the ToL test has a linear time complexity on the number of the graph's edges and nodes, it was completely solvable since we considered a small input space with only three balls and three spots. Hence, when participants press the button to ask for a suggestion, the robot queries the algorithm to retrieve the first action to optimally reach the final game configuration (the goal) from the current one. The robot suggestion involved the description of such an action. There was no limit about the number of suggestions that participants could ask; basically, they could ask for suggestions at each step of the tests.

Measure	Phase	AVG	STD
Time	Assessment	14.35	10.46
	Baseline	17.35	13.19
	Training	**11.38**	9.87
Moves	Assessment	7.24	3.25
	Baseline	7.14	3.58
	Training	**5.49**	3.76
Suggestions	Assessment	0.00	0.00
	Baseline	0.00	0.00
	Training	*0.11*	0.44
Score	Assessment	2.55	0.65
	Baseline	2.48	0.72
	Training	**2.72**	0.57

Table 1
Descriptive Statistics of the Metrics.

4.3. Hypotheses

Our hypothesis regarded both participants' performance and satisfaction. We aim to examine the degree of assistance participants require in completing the task. This holds paramount significance because, while we acknowledge that support might facilitate an immediate understanding of how to solve the test, we are cautious about the potential long-term impact on participants' performance. We want to avoid the risk of participants continually relying on suggestions, which may not lead to sustained improvement and even exacerbate their overall performance.

5. Results and Discussion

We found that the minority of participants asked the robot for help during training (45% on average), with an average number of suggestions asked of 1.53 (with $\sigma = .717$) among those who asked, and a maximum number of 3 suggestions asked per test. Hence, to investigate the difference between the tests in which participants asked the robot for help and those in which they never asked, we introduced in our analysis a binary categorical variable representing whether participants asked for help at least once during each test (the single configuration to solve) of the training.

To measure participants' improvements, we defined the delta between the number of moves needed to solve a ToL test and those performed by participants to solve it as their difference in absolute value. Considering this metric, participants' performance showed an improvement from the Baseline and Training phases ($\mu = 1.89$ with $\sigma = 2.85$ regarding the Baseline phase, and $\mu = 0.988$ with $\sigma = 2.73$ regarding the Training phase). Through a *two-way ANOVA test*, we found an effect of the experimental phase on such measure of learning ($F(2) = 3.57$ with $p = .029$). With a post-hoc Tukey correction, we found that such an improvement was statistically significant ($t = 2.548$ with $p = .03$), while there was no statistical difference between performance in Baseline/Training and Assessment. Hence, we did not find significant improvements in Assessment on average.

Regarding the Training phase, the number of suggestions

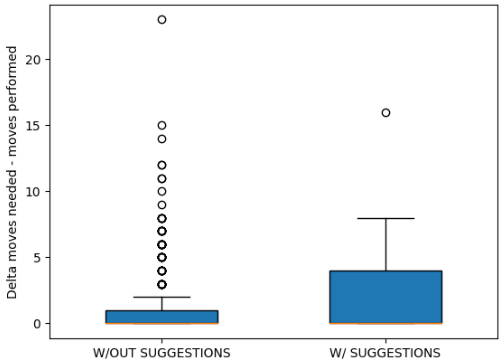

Figure 3: Delta (in absolute values) between participants' actions and those needed to solve the relative ToL configurations during the Training phase, considered as a measure of performance. The closer the delta values are to zero, the higher the participants' performance since it reflects the number of moves they performed to solve the configuration in addition to those required. The x-axis represents the trials in which participants asked for help at least once, or they never asked for help.

asked significantly impacted as a covariate on the number of moves they performed to solve the configurations (*ANCOVA test* $F(1) = 13.9$ with $p < .001$) and completion time (*ANCOVA test* $F(1) = 44.4$ with $p < .001$). We found significantly worse performance in participants who asked the robot for help compared to those who never requested (*Independent Samples t-test,* $t(238) = -2.16$ *with* $p = .032$), as shown in Figure 3. Moreover, we observed that the number of suggestions asked positively correlated with participants completion time (Pearsons' $\rho = .406$ with $p < .001$), number of moves to solve the configurations (Pearsons' $\rho = .226$ with $p < .001$), and the configurations' difficulty (Pearsons' $\rho = .238$ with $p < .001$).

Since the robot suggestions were optimal, these latter correlations explain the worse performance of those who asked for help. We hypothesize that the participants who asked for the robot's suggestions also did not know the game or obtained worse performance. Thus, only bad players needed to ask the robot for help. This hypothesis is supported by the latter correlations, which highlight how the number of suggestions asked grew with the difficulty of the game (and vice versa) and that participants who asked the robot for help more were those who obtained worse performance in solving the game.

As a further measure of performance, we collected participants' completion times, which were defined as the time taken to complete each ToL test. We performed an *Independent Samples t-test* to compare participants' completion times depending on whether they asked the robot for help at least once during the tests or never asked. We found that assisted participants took more time to complete the ToL configurations than the other ($t(398) = -4.79$ with $p < .001$), as shown in Figure 4.

We found similar results regarding the number of moves participants performed to solve the ToL tests. We conducted an *Independent Samples t-test* on the number of moves participants performed to solve the configurations and found a significant effect on whether they asked for help at least once ($t(398) = -2.7$ with $p = .008$) with the assisted participants performing more moves than the others, as shown in Figure 5.

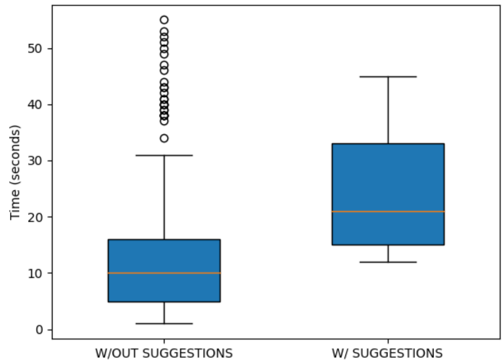

Figure 4: Participants' completion times (in seconds) in both the experimental groups, divided depending on whether they asked the robot for help at least once, or they never asked it for suggestions.

Participants who asked the robot for help are those who perceived its suggestions more useful as the correlation between the perceived usefulness of suggestions and the number of suggestions asked (Pearsons' $\rho = .492$ with $p = .028$). Moreover, by comparing participants' answers to the IOS test, we found that participants felt closer to the robot after having interacted with it, regardless of whether they asked it for help (*Repeated Measures ANOVA*, $F(1) = 5.716$ with $p = .028$). A post-hoc Tukey correction highlighted a significant difference between the pre- and post-experiment test answers ($t = -2.39$ with $p = .028$).

Finally, we found that participants' personality dimensions (Big 5 and Sense of Agency tests) did not affect their behavior during the training and assessment phases. The two robot modalities did not elicit differences in participants perception of it (Godspeed test). We neither found differences between the groups regarding the perceived closeness to the robot (IOS test).

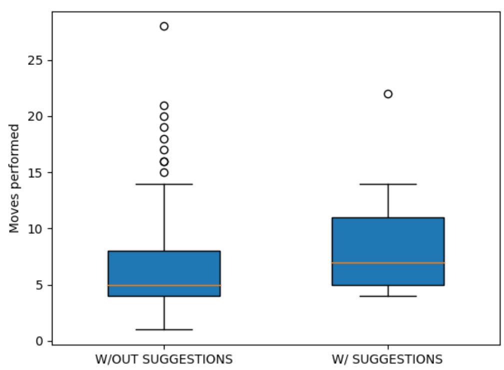

Figure 5: Number of moves participants performed to solve the ToL configurations in both the experimental groups divided depending on whether they asked the robot for help at least once or they never asked it for suggestions.

6. Conclusions

Despite the robot's availability to provide suggestions and its correct and complete knowledge of the task, a large number of participants (55%) never asked for help during the game. However, they could have gained from the robot's support, as their performance during the training was suboptimal. Interestingly, participants who asked for help had more difficulty with the task (as highlighted by an average longer completion time and a larger number of moves). Furthermore, they also recognized the usefulness of the suggestions at the end of the game. Nonetheless, the vast majority did not take advantage of the robot's help but preferred to play autonomously. This was probably due to the perception that the task was sufficiently easy to be addressed and by the desire to challenge themselves. Although this is certainly positive, it also raises the question of how to best provide robot support to maximize its utility. Our results demonstrate that perfect competence and periodic help offerings are insufficient unless the participants recognize their difficulty. Future work should focus on identifying novel strategies to counteract help aversion.

Acknowledgments

The work was supported by the research RESTART project (Robot Enhanced Social abilities based on Theory of mind for Acceptance of Robot in assistive Treatments) (CUP: I53D23003780001), funded by the MIUR with D.D. no.861 under the PNRR and by the European Union - Next Generation EU.

References

[1] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, F. Tanaka, Social robots for education: A review, Science Robotics 3 (2018) eaat5954. doi:10.1126/scirobotics.aat5954.

[2] F. Yuan, E. Klavon, Z. Liu, R. P. Lopez, X. Zhao, A systematic review of robotic rehabilitation for cognitive training, Frontiers in Robotics and AI 8 (2021).

[3] S. Rossi, G. Santangelo, M. Staffa, S. Varrasi, D. Conti, A. Di Nuovo, Psychometric evaluation supported by a social robot: Personality factors and technology acceptance, in: 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2018, pp. 802–807. doi:10.1109/ROMAN.2018.8525838.

[4] D. Piovesan, A computational index to describe slacking during robot therapy, in: J. Laczko, M. L. Latash (Eds.), Progress in Motor Control: Theories and Translations, Advances in Experimental Medicine and Biology, Springer International Publishing, Cham, 2016, pp. 351–365. doi:10.1007/978-3-319-47313-0_19.

[5] A. Ramachandran, C.-M. Huang, B. Scassellati, Toward effective robot–child tutoring: Internal motivation, behavioral intervention, and learning outcomes, ACM Transactions on Interactive Intelligent Systems 9 (2019) 1–23. doi:10.1145/3213768.

[6] A. M. Aroyo, F. Rea, G. Sandini, A. Sciutti, Trust and social engineering in human robot interaction: Will a robot make you disclose sensitive information, conform to its recommendations or gamble?, IEEE Robotics and Automation Letters 3 (2018) 3701–3708. Publisher: IEEE.

[7] S. Rossi, D. Conti, F. Garramone, G. Santangelo, M. Staffa, S. Varrasi, A. Di Nuovo, The role of personality factors and empathy in the acceptance and performance of a social robot for psychometric evaluations, Robotics 9 (2020). doi:10.3390/robotics9020039.

[8] R. G. Morris, A. D. Baddeley, Primary and working memory functioning in alzheimer-type dementia, Journal of Clinical and Experimental Neuropsychology 10 (1988) 279–296. doi:10.1080/01688638808408242, pMID: 3280591.

[9] A. M. Owen, J. J. Downes, B. J. Sahakian, C. E. Polkey, T. W. Robbins, Planning and spatial working memory following frontal lobe lesions in man, Neuropsychologia 28 (1990) 1021–1034. doi:https://doi.org/10.1016/0028-3932(90)90137-D.

[10] T. Shallice, Specific impairments of planning, Philosophical Transactions of the Royal Society of London. B, Biological Sciences 298 (1982) 199–209. doi:http://doi.org/10.1098/rstb.1982.0082.

[11] S. D. Gosling, P. J. Rentfrow, W. B. Swann, A very brief measure of the big-five personality domains, Journal of Research in Personality 37 (2003) 504–528. doi:https://doi.org/10.1016/S0092-6566(03)00046-1.

[12] A. Tapal, E. Oren, R. Dar, B. Eitam, The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment, Frontiers in psychology 8 (2017) 1552. doi:10.3389/fpsyg.2017.01552.

[13] C. Bartneck, D. Kulić, E. Croft, S. Zoghbi, Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, International journal of social robotics 1 (2009) 71–81. doi:10.1007/s12369-008-0001-3.

[14] A. Aron, E. N. Aron, D. Smollan, Inclusion of other in the self scale and the structure of interpersonal closeness., Journal of personality and social psychology 63 (1992) 596. doi:10.1037/0022-3514.63.4.596.

[15] R. Biundo, L. Weis, G. Abbruzzese, G. Calandra-Buonaura, P. Cortelli, M. C. Jori, L. Lopiano, R. Marconi, A. Matinella, F. Morgante, et al., Impulse control disorders in advanced parkinson's disease with dyskinesia: the althea study, Movement Disorders 32 (2017) 1557–1565. doi:10.1002/mds.27181.

[16] J. V. Hindle, A. Petrelli, L. Clare, E. Kalbe, Nonpharmacological enhancement of cognitive function in parkinson's disease: a systematic review, Movement Disorders 28 (2013) 1034–1049. doi:10.1002/mds.25377.

[17] B. Guglietti, D. Hobbs, L. E. Collins-Praino, Optimizing cognitive training for the treatment of cognitive dysfunction in parkinson's disease: current limitations and future directions, Frontiers in Aging Neuroscience 13 (2021) 709484. doi:10.3389/fnagi.2021.709484.