

Imitation of joint attention in human-robot interaction (HRI) during two-matchstick problem solving

Anastasia N. Stolyarova, postgraduate student, research engineer

Russian Presidential Academy of National Economy and Public Administration, Moscow (Russia)

National Research Center Kurchatov Institute, Moscow (Russia)

E-mail: anastasiyas050298@gmail.com

ORCID: <https://orcid.org/0009-0005-7155-0426>

Received 06.02.2026

Revised 19.03.2026

Accepted 19.03.2026

Abstract: Problem. In the field of human-robot interaction (HRI), despite the development of anthropomorphic robots capable of initiating joint attention, it remains unclear how different types of cues (gesture+gaze vs. gaze only) and their accuracy influence automatic human following and subjective evaluation of the interaction. **Aim.** To compare the effectiveness of different types of cues from an anthropomorphic robot (pointing gestures combined with gaze and gaze-only cues) in a task requiring joint attention, as well as to assess the influence of cue accuracy on participants' behaviour. **Methods.** The study involved 43 students from RANEPA (Russian Presidential Academy of National Economy and Public Administration): 38 females and 5 males aged 19 to 27 years ($M=20.51$; $SD=1.82$). The number of participant movements following the robot's cues and coinciding with the cue direction was evaluated to assess the effectiveness of the robot's cues in each condition. To study participants' reactions to the cues after each task, a questionnaire based on Danek's metacognitive scales was used. **Results.** The results demonstrated the robot's ability to imitate the process of joint attention during problem-solving with the participant. The hypothesis regarding the greater effectiveness of robot cues using pointing gestures combined with gaze and head movement compared to gaze-only cues was confirmed. The hypothesis regarding the greater effectiveness of correct cues compared to incorrect robot cues was confirmed. **Conclusions.** The robot's ability to imitate the joint attention process during problem-solving with the participant was demonstrated; participants paid attention to the robot's cues and attempted to follow them in both correct and incorrect cue conditions. However, in the condition with correct cues, the percentage of response attempts coinciding with the cue direction was significantly higher than in the condition with incorrect cues.

Keywords: Human-Robot Interaction; HRI; anthropomorphic robots; joint attention; gaze and gesture cues; modified Knoblich tasks; Danek's metacognitive scales

Acknowledgements: This work was carried out within the state assignment to the National Research Center Kurchatov Institute. The development of the methodology, including the experimental conditions, gestures, and communication strategies of the anthropomorphic interface, was carried out as part of the state assignment of the Kurchatov Institute National Research Center. The statistical analysis of the experimental results, including comparisons of the experimental conditions and the analysis of the participants' subjective and behavioral assessments, was carried out as part of the Russian Science Foundation grant No. 25-78-10154, <https://rscf.ru/project/25-78-10154/>.

For citation: Stolyarova A.N. Imitation of joint attention in human-robot interaction (HRI) during two-matchstick problem solving. *Evidence-based education studies*, 2026, no. 1, pp. 27–36. DOI: <https://doi.org/10.18323/3034-2996-2026-1-64-3>.

INTRODUCTION

Contemporary research in the field of human-robot interaction (HRI) is increasingly shifting focus from purely technical aspects to considering the psychological characteristics of the human as a full participant in joint activity. The importance of this approach is confirmed by a systematic review by Morandini et al. (2025) [1], which showed that adapting collaborative robots to human cognitive and emotional states (attention, workload, stress) is becoming a key direction for improving the efficiency and safety of joint work. The present study develops this line of research, focusing on automatic mechanisms of joint attention in interaction with a robot.

Human development is characterised by a tendency towards joint attention, defined as the ability to share and direct attention to an object or event of common interest. Joint attention is attention shared between two individuals. It can be visual, auditory, or gestural. The development of joint attention is facilitated by mechanisms such as imitation, turn-taking behavior, and social referencing [2]. Integrating the fundamental mechanisms of joint attention into human-robot interaction represents an important task for robotics, especially for social robots: developing the robot's capacity for joint abilities with humans through visual attention and self-assessment mechanisms, and the robot's response to the emotional context of joint attention events [2].

© Stolyarova A.N., 2026

However, most approaches to joint attention in robotics require additional evaluation with real users. Human-robot interaction (HRI) can be classified according to several criteria: by task type (depending on the task at hand), by robot morphology (depending on its appearance and application area), by human attitude towards the robot (positive or negative), by the level of joint interaction between teams (depending on team composition), and by the roles of interaction between human and robot (supervisor, operator, team member, programmer, observer) [3]. To study joint attention processes in HRI, methods of non-verbal communication are used: gestures and gaze directions [4; 5]. Unconscious cognitive states (load, attention) can be measured and used to optimize collaborative work [6].

Considering approaches to using humanoid robot agents in joint attention research, two main approaches to the robots used can be distinguished. In more classical studies, robot faces are presented on a screen or software-based technical devices are used; the use of such stimuli allows addressing the question of the role of human likeness and human-nature interaction in triggering joint attention mechanisms [7]. That is, with the help of artificial humanoid agents, we can test whether similarity to humans is a decisive factor for engaging in joint attention. In more interactive protocols with embodied humanoids, the advantage of their use lies in overcoming the problems of previous interactive protocols by offering superior experimental control on one hand, while providing increased ecological validity and social presence on the other [8; 9].

According to data from [10], participants interacting with a humanoid robot achieved better recognition memory performance in the initiating joint attention condition than in the responding condition. At the same time, their study showed that human-human and human-robot interactions did not lead to quantifiable differences in recognition memory. Sometimes participants show significantly better task performance with a robot compared to a human [11]. Recent studies have shown that embodied robots can effectively attract attention, although the observed effects do not depend on whether the robot is controlled by a human or artificial intelligence [12]. Joint activity with a robot depends on the influence of individual variability (strategy) and task variability (complexity) on the quality of the outcome (task success) and on human workload [13].

In most existing studies of joint attention, either gaze direction [8; 9; 14] or various gestures [7; 15] are used to attract participants' attention. At the same time, it remains unclear which type of interaction is more effective for performing collaborative activity with a human. The literature contains almost no empirical studies aimed at comparing both methods of attracting human attention by a robot. To investigate this problem, we studied the ability of the robot's non-verbal cues (pointing and visual) to influence the solution of matchstick algebra problems, facilitating or complicating the participants' problem-solving process through correct and incorrect cues.

The research question was whether the effectiveness of the robot's cues differs depending on their type (gestures + gaze vs. gaze only) or the correctness of the presented cues.

The aim of the study was to compare the effectiveness of different types of cues from an anthropomorphic robot (pointing gestures combined with gaze and gaze-only cues) in a task requiring joint attention, as well as to assess the influence of cue accuracy on participants' behavior.

METHODS

Sample

The study involved 43 students from the Russian Presidential Academy of National Economy and Public Administration (RANEPA): 38 females and 5 males aged 19 to 27 years ($M=20.51$; $SD=1.82$). This sample is typical for humanities students in higher education institutions. The sample consisted of volunteers – students of RANEPA. All participants signed written informed consent to participate in the study, including consent for video recording of their behavior. Participants were informed that they could withdraw from further participation in the experiment at any time. The participants were not familiar with the tasks presented. Participants received credit points in psychology disciplines for participation in the experiment. Individuals with vision not corrected to normal, motor impairments, traumatic brain injuries, or cognitive function disorders were excluded from the experimental sample. The study was approved by the Minutes of the Meeting of the Commission on Intra-University Surveys and Ethical Evaluation of Empirical Research Projects No. 1/2026.

Equipment and materials

The study was conducted using the anthropomorphic robot F-2 (Fig. 1) created by the staff of the National Research Center “Kurchatov Institute”, Moscow, Russia. This robot can be used to create and experimentally test strategies of communication with humans through speech, facial expressions, and gestures. The F-2 robot resembles cartoon characters that do not look like humans but are emotional and likable due to their gestures and facial expressions. The robot demonstrates facial expressions, can turn its head, and move its arms. The robot's design is made as simple as possible. The robot's interface (face) is a 13 cm (5-inch) monitor, which displays moving images of eyes and mouth. The robot's manipulators and neck are assembled from Dynamixel AX-12A actuators. Each manipulator can rotate and rise. The robot can rotate its head sideways, raise and lower its head.

The robot's operating voltage is 12 V. An external power supply unit is used to connect to a 220 V network. The robot is controlled from a personal computer or laptop, connecting via USB and HDMI cables, and is plugged into a 220 V network using a power supply unit. The robot was controlled by the experimenter using a laptop. The experiment was recorded on camera.

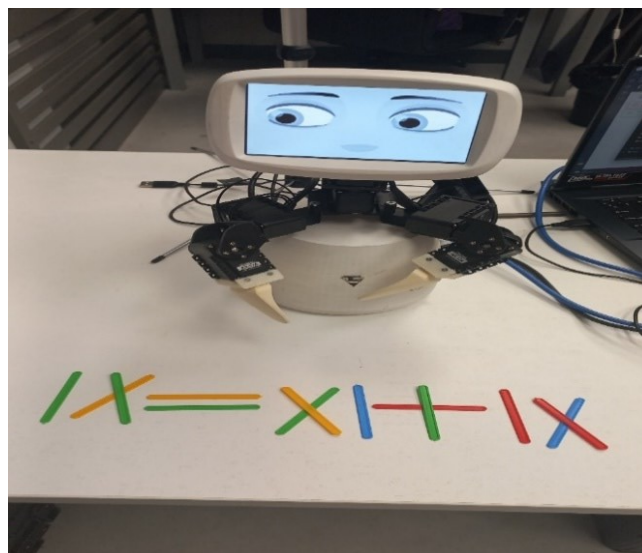


Fig. 1. Example of a matchstick algebra problem laid out in front of the robot
Рис. 1. Образец примера спичечной алгебры, выложенный перед роботом

Stimulus materials

The study employed modified Knoblich matchstick algebra problems [16]. According to his classification, there are four types of problems: *A*, *B*, *C*, and *D*. Type *A* problem is solved by relaxing the numerical constraint – an arithmetic operation (e.g., $IV=III+III$). Type *B* problem requires relaxing the arithmetic sign constraint (the arithmetic operation sign cannot be changed: $III=V+III$). Type *C* problem is solved by relaxing the tautology constraint (the form of the equation is unchanged: $III=III+III$). Type *D* problem involves relaxing the numerical constraint and chunk decomposition (e.g., $XI=III+III$). Knoblich’s experiments confirmed that type *A* problems are easier to solve than type *B* problems, which in turn are easier than type *C* problems; type *D* problems are more difficult to solve than type *A* problems [16; 17]. The study used modified types of matchstick algebra problems for two-matchstick puzzles [18]; these are more complex tasks that allow assessing the effect of joint attention imitation during the prolonged process of solving each problem.

Each participant was presented with four matchstick algebra problems [18]: $IX=VIII-III$ (value-operator), $VI=IX+IV$ (hybrid), $III=XI-IV$ (value), $IX=XI+IX$ (tautology). Solving each problem required moving two sticks. Each of the presented problems could be solved in different ways: the value-operator problem required changing both the digit value and the arithmetic sign; solving the tautology problem required creating a double equality, etc. The problems were laid out using physical counting sticks that the participant could manipulate.

Instrumentation

A specially designed brief questionnaire was used after each problem to assess the perception of interaction with the robot. It consisted of four items, each representing a statement rated on a 7-point Likert scale (1 – “strongly disagree”, 7 – “strongly agree”). All ques-

tionnaire items were presented to participants as a list with complete formulations.

Specific items and measured constructs:

1. Robot effectiveness: Statement: “The robot’s cues were helpful in solving the problem.” A higher score reflected higher perceived usefulness.

2. Distraction from the robot: Statement: “The robot’s actions or presence distracted me and interfered with my concentration.” A higher score corresponded to a stronger feeling of distraction.

3. Subjective problem complexity: Question: “How would you rate the difficulty of the problem you just solved?” (1 – “very easy”, 7 – “very difficult”).

4. Effort expended: Question: “How much mental effort did you need to solve this problem?” (1 – “very little”, 7 – “very much”).

The scales for assessing complexity and effort used bipolar descriptors, which is standard practice for measuring metacognitive assessments. The questionnaire was presented on paper immediately after completing each of the four tasks. These items were formulated by the authors for this study to directly assess key aspects of interaction identified in the pilot study.

Procedure and design

The two-stick problems were presented to participants in four randomized conditions: correct gaze, incorrect gaze, correct gesture + gaze, incorrect gesture + gaze). The types of cues were alternated across task types randomly. Each type of cue was used throughout the solution of one of the four problems. Through cues, the robot indicated the correct or incorrect side of the example from which two counting sticks needed to be moved, depending on the current condition.

Before the experiment began, participants were given instructions for solving tasks with counting sticks. After greeting the participant, the robot delivered the instructions:

"You will solve matchstick moving tasks together with me. You need to move two sticks so that the equality becomes correct. Roman numerals must be laid out strictly according to the template. (The Roman numeral template was placed next to the participant). You have ten minutes to solve each task. I will give you hints in different ways during the solution."

The robot provided cues in different ways: every 30 seconds, the robot gave a cue depending on the current condition: it provided correct gaze cues, incorrect gaze cues, correct pointing gestures combined with gaze, and incorrect gestures combined with gaze.

After the instructions were delivered, the participant had the opportunity to ask any remaining questions to the experimenter.

A countdown began, and the participant started solving the problem. In the event of a correct solution to the problem, the participant was informed that they had solved the problem correctly; in the absence of a correct solution, the participant continued solving until the time elapsed.

Correct cues were the robot indicating the part of the equation from which the counting stick needed to be taken for its correct movement. Incorrect cues were the robot indicating the part of the equation opposite to the one from which the counting stick needed to be taken for correct movement.

After each task, participants were asked to complete a questionnaire of four questions to study subjective evaluations of the robot's cues.

Data analysis

Video data processing was carried out by two independent experts (the author of the paper and an external specialist in non-verbal communication). Participant gestures, head turns, and gaze direction that coincided with or contradicted the direction of the robot's cue were recorded.

The consistency of expert assessments was evaluated using Cohen's kappa coefficient ($\kappa=0.87$). A two-way analysis of variance (ANOVA) with assessment of between-group effects in SPSS was used to compare participants' non-verbal responses depending on the cue type (gesture or gaze) and its correctness. Assumptions of normality (Shapiro-Wilk test) and homogeneity of variances (Levene's test) were checked in advance; no significant violations were found. Observations in the experiment are independent of each other, and the dependent variable was measured on an interval scale. The significance level was set at $p<0.05$. Effect size was assessed using η^2 .

Questionnaire data were processed as follows: mean values and standard deviations were calculated for each item. Internal consistency of the scale was checked using Cronbach's α coefficient ($\alpha=0.82$). The same two-way analysis of variance was used to compare scores under different conditions.

RESULTS

Behavioural data

Analysis of video materials tracking participant movements in response to different robot cue types revealed that participant movements in the direction of the robot's cues

under the condition with correct pointing gestures combined with robot gaze ($M=5.84$, $SD=3.40$) coincided most often (90.15 % of all movements in this condition). Participants followed cues slightly less often under the correct robot's gaze condition ($M=4.53$, $SD=3.15$) – 80.05 %. In contrast, under conditions with incorrect robot's pointing gestures combined with gaze ($M=4.33$, $SD=3.47$) and with incorrect robot's gaze ($M=3.18$, $SD=2.91$), participants followed cues in significantly fewer cases, 59.51 % and 44.32 % respectively.

Participant movements contradicting the robot's cues were most frequent in the conditions with incorrect gaze ($M=2.33$, $SD=2.19$) and with incorrect pointing gestures combined with gaze ($M=1.91$, $SD=2.64$) – 45.60 % and 33.54 % of all movements in each of these conditions. Meanwhile, in the conditions with correct robot's gaze ($M=0.53$, $SD=0.83$) and correct pointing gestures combined with robot's gaze ($M=0.44$, $SD=0.70$), participant movements contradicted the cues in significantly fewer cases, 11.82 % and 7.20 %. These data are consistent with the hypotheses regarding the greater effectiveness of robot cues using pointing gestures compared to cues using gaze alone, and the greater effectiveness of correct cues compared to incorrect ones.

Comparing the factors of cue type and cue correctness using two-way analysis of variance (ANOVA), it was found that the mean number of participant movements in the direction of the cue differed significantly between cue types (gaze only ($M=3.86$, $SD=3.09$) and pointing gestures combined with gaze ($M=5.08$, $SD=3.49$): $F(3,171)=6.103$, $p<0.01$, $\eta^2p=0.035$, and between correct ($M=5.19$, $SD=3.32$) and incorrect cues ($M=3.76$, $SD=3.24$): $F(3,171)=8.375$, $p<0.01$, $\eta^2p=0.047$). Thus, participants followed correct cues more often and preferred robot cues combining gaze direction and pointing gestures (Fig. 2).

Subjective ratings

Indicators of subjective evaluation of robot effectiveness differed significantly depending on the cue type ($F(3,171)=9.745$, $p<0.01$, $\eta^2p=0.05$). According to the questionnaire data, problems in the condition with combined cues (gesture+gaze) were rated as more effective ($M=4.36$, $SD=2.02$) than in the gaze-only condition ($M=3.38$, $SD=2.13$) (Fig. 3).

Notably, participants rated combined cues as more effective even in the condition with incorrect cues ($M=4.81$, $SD=2.00$).

Indicators of subjective evaluation of distraction caused by the robot were not significant according to the questionnaire data, both in the conditions with combined cues (gesture + gaze) ($M=2.74$, $SD=1.96$) and visual cues ($M=2.55$, $SD=1.84$) and in the conditions with correct cues ($M=2.78$, $SD=1.82$) and incorrect cues ($M=2.51$, $SD=1.97$) (Fig. 4).

Indicators of subjective problem complexity differed significantly depending on the cue type ($F(3,171)=12.889$, $p<0.01$, $\eta^2p=0.07$). According to the questionnaire data, problems in the condition with combined cues (gesture+gaze) were rated as more complex ($M=4.55$, $SD=1.50$) than in the gaze-only condition ($M=3.74$, $SD=1.46$) (Fig. 5).

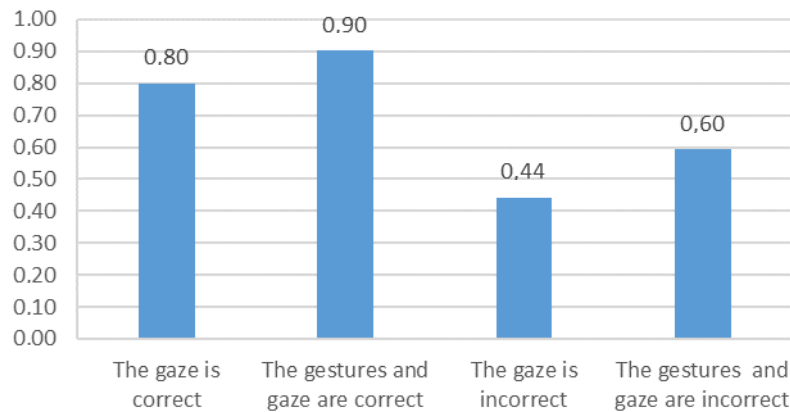


Fig. 2. Proportion of participants' movements in the direction of the robot's cue (%) depending on cue type (gaze / gesture + gaze) and cue correctness. Bars represent mean values; errors indicate standard deviations
Рис. 2. Доля движений испытуемых в направлении подсказки робота (в %) в зависимости от типа подсказки (взгляд / жест + взгляд) и её правильности. Столбцы – средние значения; ошибки – стандартные отклонения

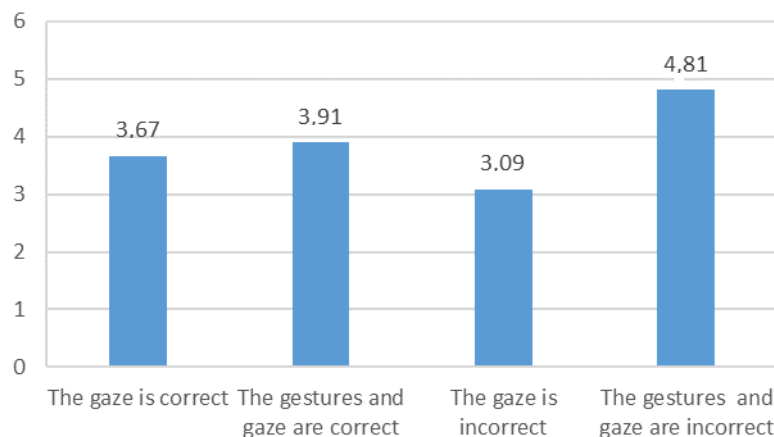


Fig. 3. Mean ratings of robot effectiveness according to the questionnaire (in points) depending on cue type (gaze / gesture + gaze) and cue correctness. Bars represent mean values; errors indicate standard deviations
Рис. 3. Средние показатели эффективности робота по опроснику в баллах в зависимости от типа подсказки (взгляд / жест + взгляд) и её правильности. Столбцы – средние значения; ошибки – стандартные отклонения

Indicators of subjective evaluation of effort expended on solving problems differed significantly depending on the cue type ($F(3,171)=17.467, p<0.01, \eta^2p=0.09$). According to the questionnaire data, problems in the condition with combined cues (gesture+gaze) were rated as requiring more effort ($M=5.02, SD=1.49$) than in the gaze-only condition ($M=4.02, SD=1.67$) (Fig. 6).

DISCUSSION

The experimental results demonstrate that the robot's ability to attract joint attention depends both on the fact of the cue itself and on its modality and correctness. The highest proportion of coinciding movements was observed under conditions with combined cues (gesture + gaze), with this effect being particularly pronounced for correct cues.

This indicates that multimodal signals are perceived as more reliable guides.

The obtained data confirm that combined cues (gesture+gaze) are significantly more effective for initiating joint attention than gaze-only cues (Fig. 2). This aligns with studies showing that multimodal signals are perceived as more reliable [10]. Likely, the pointing gesture acts as a spatial anchor that reinforces and clarifies the direction indicated by the gaze. Interestingly, even incorrect combined cues elicited more coinciding movements than correct gaze-only cues (59.51 % vs. 80.05 %). This may indicate a priority of the motor signal over the visual signal under conditions of uncertainty. Subjective ratings of robot effectiveness (Fig. 3) confirm that participants not only objectively followed combined cues better but also perceived the interaction as more successful. Participants rated the robot's

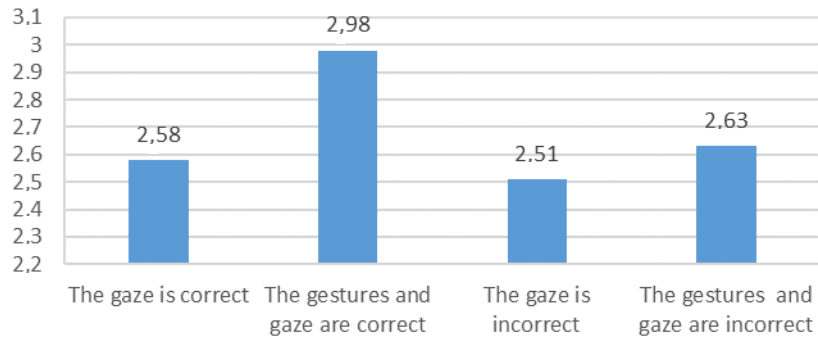


Fig. 4. Mean ratings of robot as a distraction according to the questionnaire (in points) depending on cue type (gaze / gesture + gaze) and cue correctness. Bars represent mean values; errors indicate standard deviations

Рис. 4. Средние показатели робота как помехи по опроснику в баллах в зависимости от типа подсказки (взгляд / жест + взгляд) и её правильности. Столбцы – средние значения; ошибки – стандартные отклонения

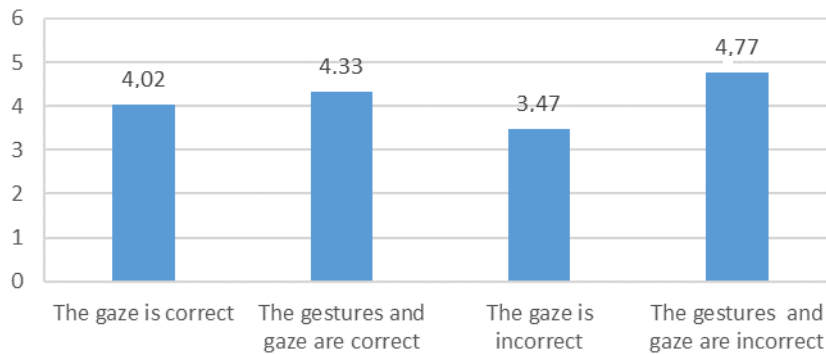


Fig. 5. Mean ratings of problem complexity according to the questionnaire (in points) depending on cue type (gaze / gesture + gaze) and cue correctness. Bars represent mean values; errors indicate standard deviations

Рис. 5. Средние показатели сложности решения задачи по опроснику в баллах в зависимости от типа подсказки (взгляд / жест + взгляд) и её правильности. Столбцы – средние значения; ошибки – стандартные отклонения

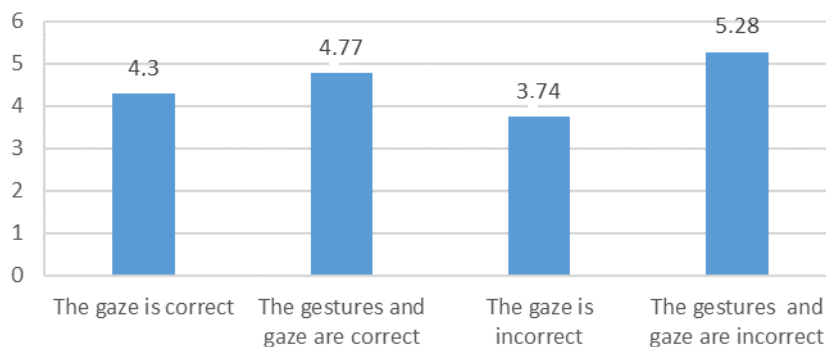


Fig. 6. Mean ratings of effort expended by participants according to the questionnaire (in points) depending on cue type (gaze / gesture + gaze) and cue correctness. Bars represent mean values; errors indicate standard deviations

Рис. 6. Средние показатели количества усилий, затраченных испытуемыми, по опроснику в баллах в зависимости от типа подсказки (взгляд / жест + взгляд) и её правильности. Столбцы – средние значения; ошибки – стандартные отклонения

effectiveness significantly higher under conditions where pointing gestures were combined with gaze compared to gaze-only conditions (Fig. 3). The higher ratings of robot effectiveness under conditions with gestures are likely related to the fact that multimodal signals are perceived as more natural and reliable, which is also supported by the objective indicators of following the cues. Participants’ subjective ratings approximated actual success rates under the gaze conditions (both correct and incorrect); under the combined cue conditions, subjective ratings were opposite to the objective data, as participants rated incorrect gesture cues even higher than correct ones. Thus, for further development of human-robot interaction design, it is necessary to improve the robot’s manipulators, equipping them with fingers for greater precision and localization of pointing gestures, as well as to enhance the expressiveness of the robot’s eyes and the clarity of gaze focusing on specific objects.

It can be hypothesised that the enhanced clarity of gesture cues facilitated their perception and thereby attracted more participant attention to the robot’s actions.

Despite the robot being a weak agent of joint attention, capable of acting only within limited cue parameters, it succeeded in imitating joint attention during interaction with participants. Respondents rarely consciously distinguished between correct and incorrect cues, rating the robot as a helpful assistant.

Our data show an interesting dissociation: although participants objectively followed incorrect cues almost as often as correct ones, they subjectively rated the robot as a highly effective assistant. This aligns with the idea that orientation towards social signals, such as gaze and gesture, is largely an automatic, “implicit” mechanism [2; 12], which may not be accompanied by conscious analysis of their correctness. Likely, participants did not realise the moment of error at the action level, and their subsequent explicit evaluation was formed based on the overall positive experience of interaction, rather than on a step-by-step analysis of the success of each cue. As noted in [12], the use of anthropomorphic robots often triggers precisely this pattern: an automatic response to social signals is combined with a persistent attribution of social qualities to the robot, including usefulness, even under conditions of imperfect performance.

Our findings that participants followed the robot’s cues without recognising their erroneous nature are consistent with recent research on automatic mechanisms of joint attention. For instance, [19] demonstrated that after even short-term collaboration, a robot’s hand begins to attract automatically human attention, even without conscious control. This automaticity likely underlies the following of cues observed in our study. Moreover, the connection identified by the authors between the perception of robot competence and the degree of attention shift explains why our participants, despite the robot’s errors, rated it as a helpful assistant: the very fact of successful joint activity forms a positive image that influences subjective evaluation more strongly than the analysis of individual failures.

Our results are consistent with studies [12; 20] that also demonstrated the advantage of using embodied robots and multimodal signals in HRI. The fact that participants rated the robot as a helpful assistant despite its objective errors

may be related to the process of successful coordination and adaptation of the robot (even imperfect) triggers positive social mechanisms. As shown in the work by Ehrlich et al. [21], even implicit neural signals of approval/disapproval can be used to train robots, contributing to the formation of an image of the robot as an “understanding” and, consequently, useful partner.

The high rating of robot usefulness, despite its “errors”, may be related to the interactive process itself (the robot tries to help, looks, and points) eliciting a socially positive attitude and attribution of “assistant” qualities [12]. However, we extended these findings by showing that the effect persists not only in scenarios where the robot reacts to the human but also when it acts as the initiator of joint action. Conversely, our finding of no differences between correct and incorrect cues in combined conditions regarding participants’ subjective ratings of robot effectiveness, distraction, subjective problem complexity, and effort expended (Figs. 3–6) does not align with data from [11; 14]. Participants in the study [11] performed better in solving card-matching tasks in the condition with robot assistance compared to the condition without robot assistance. According to [14], changes in participant behavior across different phases of joint attention are episodic. These discrepancies may be related to the fact that respondents rarely consciously distinguished between correct and incorrect cues and attempted to follow all of the robot’s cues.

Indicators of subjective evaluation of effort expended on solving problems in the combined gesture conditions, especially incorrect ones, were higher compared to the gaze conditions (Fig. 6). This may be associated with the noisy movement of the robot’s joints, which could periodically create a distracting effect, which is consistent with data from [12].

Study limitations

1. Sample: The uneven gender distribution (38 women, 5 men) combined with the small sample size ($N=43$) does not allow the results to be considered representative of the general population. The academic motivation for participation (course credit) may have influenced behavior. It should be noted that participation in the experiment was incentivized with course credits, which could have created additional external motivation and potentially affected participant behavior. Future studies should consider using alternative forms of incentive or compare results with groups participating voluntarily.

2. Interaction design: Remote control of the robot by an operator (Wizard of Oz paradigm) did not reflect the operation of a fully autonomous system and could have introduced artificial delays.

3. Measurement methodology: Manual analysis of video recordings for coding movements may contain subjectivity. The absence of eye tracking does not allow precise conclusions about participants’ visual attention.

4. Ecological validity: The use of a laboratory setting and specific tasks (“matchstick puzzles”) may limit the generalization of findings to real-world collaborative scenarios.

5. Technical artifact: The noise from the robot’s manipulators, as noted in the discussion, was an uncontrolled variable that may have influenced participants’ subjective ratings.

Addressing these limitations in future research will allow enhancing the validity and generalizability of the results.

Further research aimed at enhancing the imitation of joint attention with a robot using simpler tasks of a different type, such as non-insight puzzle-solving tasks like chess, tangram, or logic puzzles, will help to better understand the features of joint attention in the context of human-robot interaction, assess the extent to which anthropomorphic robots can imitate joint attention, and contribute to the creation of more human-centered embodied robots.

CONCLUSIONS

1. The study has experimentally demonstrated the possibility of imitating the joint attention process by an anthropomorphic robot acting as an active initiator of interaction during spatial problem-solving. This result directly confirms the achievement of the study’s aim.

2. The effectiveness of imitation significantly increases when using multimodal cues (pointing gestures combined with gaze and head direction) compared to unimodal cues (gaze only), which is manifested both in objective behavioral indicators and in participants’ subjective ratings.

3. Imitation of joint attention in HRI is meaningful for humans: the effectiveness of following cues and their evaluation directly depend on their semantic correctness, indicating the emergence of genuine interaction rather than a simple reaction to movement.

4. The identified technical limitation (manipulator noise) highlights the need to consider both the social and the ergonomic component when designing such interactions.

REFERENCES

- Morandini S., Curro F., Parlangei O., Pietrantoni L. Collaborative Robots Adapting Their Behavior Based on Workers’ Psychological States: A Systematic Scoping Review. *Human Behavior and Emerging Technologies*, 2025, vol. 2025, no. 1, article number 6361777. DOI: [10.1155/hbe2/6361777](https://doi.org/10.1155/hbe2/6361777).
- Garcia-Martinez J., Gamboa-Montero J.J., Castillo J.C., Castro-Gonzalez A. Analyzing the Impact of Responding to Joint Attention on the User Perception of the Robot in Human-Robot Interaction. *Biomimetics*, 2024, vol. 9, no. 12, article number 769. DOI: [10.3390/biomimetics9120769](https://doi.org/10.3390/biomimetics9120769).
- Galín R.R., Serebrennyy V.V., Tevyashov G.K., Shirokiy A.A. Human-robot interaction in collaborative robotic systems. *Proceedings of Southwest State University*, 2020, vol. 24, no. 4, pp. 180–199. DOI: [10.21869/2223-1560-2020-24-4-180-199](https://doi.org/10.21869/2223-1560-2020-24-4-180-199).
- Cochet H., Guidetti M. Contribution of Developmental Psychology to the Study of Social Interactions: Some Factors in Play, Joint Attention and Joint Action and Implications for Robotics. *Frontiers in Psychology*, 2018, vol. 9, article number 1992. DOI: [10.3389/fpsyg.2018.01992](https://doi.org/10.3389/fpsyg.2018.01992).
- Urakami J., Seaborn K. Nonverbal Cues in Human-Robot Interaction: A Communication Studies Perspective. *ACM Transactions on Human-Robot Interaction*, 2023, vol. 12, no. 2, pp. 1–21. DOI: [10.1145/3570169](https://doi.org/10.1145/3570169).
- Jing Yang, Barragan J.A., Farrow J.M., Sundaram C.P., Wachs J.P., Denny Yu. An Adaptive Human-Robotic Interaction Architecture for Augmenting Surgery Performance Using Real-Time Workload Sensing – Demonstration of a SemiAutonomous Suction Tool. *Human Factors* 66, 2024, vol. 4, pp. 1081–1102. DOI: [10.1177/00187208221129940](https://doi.org/10.1177/00187208221129940).
- Salminen-Saari J.F.A., Moreno-Esteva E.G., Haataja E., Toivanen M., Hannula M.S., Laine A. Phases of collaborative mathematical problem solving and joint attention: a case study utilizing mobile gaze tracking. *ZDM Mathematics Education*, 2021, vol. 53, pp. 771–784. DOI: [10.1007/s11858-021-01280-z](https://doi.org/10.1007/s11858-021-01280-z).
- Anzalone S.M., Xavier J., Boucenna S., Billeci L., Narzisi A., Muratori F., Cohen D., Chetouaniet M. Quantifying patterns of joint attention during human-robot interactions: An application for autism spectrum disorder assessment. *Pattern Recognition Letters*, 2018, vol. 118, pp. 42–50. DOI: [10.1016/j.patrec.2018.03.007](https://doi.org/10.1016/j.patrec.2018.03.007).
- Kumazaki H., Yoshikawa Y., Yoshimura Y., Ikeda T., Hasegawa C., Saito D.N., Tomiyama S., Kyung-min An, Shimaya J., Ishiguro H., Matsumoto Y., Minabe Y., Kikuchi M. The impact of robotic intervention on joint attention in children with autism spectrum disorders. *Molecular Autism*, 2018, vol. 9, article number 46. DOI: [10.1186/s13229-018-0230-8](https://doi.org/10.1186/s13229-018-0230-8).
- Kim M., Kwon T., Kim K. Can Human–Robot Interaction Promote the Same Depth of Social Information Processing as Human–Human Interaction? *International Journal of Social Robotics*, 2017, vol. 10, pp. 33–42. DOI: [10.1007/s12369-017-0428-5](https://doi.org/10.1007/s12369-017-0428-5).
- Mwangi E., Barakova E.I., Díaz-Boladeras M., Mallofré A.C., Rauterberg M. Directing Attention Through Gaze Hints Improves Task Solving in Human–Humanoid Interaction. *International Journal of Social Robotics*, 2018, vol. 10, pp. 343–355. DOI: [10.1007/s12369-018-0473-8](https://doi.org/10.1007/s12369-018-0473-8).
- Chevalier P., Kompatsiari K., Ciardo F., Wykowska A. Examining joint attention with the use of humanoid robots—A new approach to study fundamental mechanisms of social cognition. *Psychonomic Bulletin and Review*, 2020, vol. 27, no. 2, pp. 217–236. DOI: [10.3758/s13423-019-01689-4](https://doi.org/10.3758/s13423-019-01689-4).
- Fournier É., Jeoffrion C., Hmedan B., Pellier D., Fiorino H., Landry A. Human-Cobot Collaboration’s Impact on Success, Time Completion, Errors, Workload, Gestures and Acceptability During an Assembly Task. *Applied Ergonomics*, 2024, vol. 119, article number 104306. DOI: [10.1016/j.apergo.2024.104306](https://doi.org/10.1016/j.apergo.2024.104306).
- Hoang-Long Cao, Simut R.E., Krepel N., Vanderborght B., Vanderfaellie J. Could NAO Robot Function as Model Demonstrating Joint Attention Skills for Children with Autism Spectrum Disorder? An Exploratory Study. *International Journal of Humanoid Robotics*, 2022, vol. 19, no. 4, article number 2240006. DOI: [10.1142/S0219843622400060](https://doi.org/10.1142/S0219843622400060).
- Okafuji Y., Baba J., Nakanishi J., Kuramoto I., Ogawa K., Yoshikawa Y., Ishiguro H. Can a humanoid robot continue to draw attention in an office environment? *Advanced Robotics*, 2020, vol. 34, no. 14, pp. 931–946. DOI: [10.1080/01691864.2020.1769724](https://doi.org/10.1080/01691864.2020.1769724).
- Knoblich G., Ohlsson S., Haider H., Rhenius D. Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1999, vol. 25, no. 6, pp. 1534–1555. DOI: [10.1037/0278-7393.25.6.1534](https://doi.org/10.1037/0278-7393.25.6.1534).

17. Spiridonov V.F., Erofeeva M.A., Klovayt N.O., Ardislamov V.V., Morozov M.I., Zdilars S. Interactive problem solving revisited: replicating the effects of interactivity using matchstick algebra problems. *Psychological studies*, 2021, vol. 14, no. 79, pp. 1–46. DOI: [10.54359/ps.v14i79.119](https://doi.org/10.54359/ps.v14i79.119).
18. Öllinger H.M., Jones G., Knoblich G. Heuristics and representational change in two-move matchstick tasks. *Advances in Cognitive Psychology*, 2006, vol. 2, no. 4, pp. 239–253. DOI: [10.2478/v10053-008-0059-3](https://doi.org/10.2478/v10053-008-0059-3).
19. Scorza Azzarà G., Zonca J., Rea F., Joo-Hyun Song, Sciutti A. Collaborating with a robot biases human spatial attention. *iScience*, 2025, vol. 28, no. 7, article number 112791. DOI: [10.1016/j.isci.2025.112791](https://doi.org/10.1016/j.isci.2025.112791).
20. Pöysä-Tarhonen J., Shupin Li, Hautala J., Awwal N., Häkkinen P. Making the Invisible Visible: Exploring Joint Attention Behaviour in Remote Collaborative Problem-Solving. *Journal of Computer Assisted Learning*, 2025, vol. 41, no. 4, article number e70068. DOI: [10.1111/jcal.70068](https://doi.org/10.1111/jcal.70068).
21. Ehrlich S.K., Dean-Leon E., Tacca N., Armleder S., Dimova-Edeleva V., Gordon Cheng. Human-robot collaborative task planning using anticipatory brain responses. *PLoS ONE*, 2023, vol. 8, no. 7, article number e0287958. DOI: [10.1371/journal.pone.0287958](https://doi.org/10.1371/journal.pone.0287958).
22. Spiridonov V.F., Erofeeva M.A., Klovayt N.O., Ardislamov V.V., Morozov M.I., Zdilars S. Interactive problem solving revisited: replicating the effects of interactivity using matchstick algebra problems // ZDM Mathematics Education. 2021. Vol. 53. P. 771–784. DOI: [10.1007/s11858-021-01280-z](https://doi.org/10.1007/s11858-021-01280-z).
23. Anzalone S.M., Xavier J., Boucenna S., Billeci L., Narzisi A., Muratori F., Cohen D., Chetouaniet M. Quantifying patterns of joint attention during human-robot interactions: An application for autism spectrum disorder assessment // Pattern Recognition Letters. 2018. Vol. 118. P. 42–50. DOI: [10.1016/j.patrec.2018.03.007](https://doi.org/10.1016/j.patrec.2018.03.007).
24. Kumazaki H., Yoshikawa Y., Yoshimura Y., Ikeda T., Hasegawa C., Saito D.N., Tomiyama S., Kyung-min An, Shimaya J., Ishiguro H., Matsumoto Y., Minabe Y., Kikuchi M. The impact of robotic intervention on joint attention in children with autism spectrum disorders // Molecular Autism. 2018. Vol. 9. Article number 46. DOI: [10.1186/s13229-018-0230-8](https://doi.org/10.1186/s13229-018-0230-8).
25. Kim M., Kwon T., Kim K. Can Human-Robot Interaction Promote the Same Depth of Social Information Processing as Human-Human Interaction? // International Journal of Social Robotics. 2017. Vol. 10. P. 33–42. DOI: [10.1007/s12369-017-0428-5](https://doi.org/10.1007/s12369-017-0428-5).
26. Mwangi E., Barakova E.I., Díaz-Boladeras M., Mallofré A.C., Rauterberg M. Directing Attention Through Gaze Hints Improves Task Solving in Human-Humanoid Interaction // International Journal of Social Robotics. 2018. Vol. 10. P. 343–355. DOI: [10.1007/s12369-018-0473-8](https://doi.org/10.1007/s12369-018-0473-8).
27. Chevalier P., Kompatsiari K., Ciardo F., Wykowska A. Examining joint attention with the use of humanoid robots-A new approach to study fundamental mechanisms of social cognition // Psychonomic Bulletin and Review. 2020. Vol. 27. № 2. P. 217–236. DOI: [10.3758/s13423-019-01689-4](https://doi.org/10.3758/s13423-019-01689-4).
28. Fournier É., Jeoffrion C., Hmedan B., Pellier D., Fiorino H., Landry A. Human-Cobot Collaboration's Impact on Success, Time Completion, Errors, Workload, Gestures and Acceptability During an Assembly Task // Applied Ergonomics. 2024. Vol. 119. Article number 104306. DOI: [10.1016/j.apergo.2024.104306](https://doi.org/10.1016/j.apergo.2024.104306).
29. Hoang-Long Cao, Simut R.E., Krepel N., Vanderborght B., Vanderfaillie J. Could NAO Robot Function as Model Demonstrating Joint Attention Skills for Children with Autism Spectrum Disorder? An Exploratory Study // International Journal of Humanoid Robotics. 2022. Vol. 19. № 4. Article number 2240006. DOI: [10.1142/S0219843622400060](https://doi.org/10.1142/S0219843622400060).
30. Okafuji Y., Baba J., Nakanishi J., Kuramoto I., Ogawa K., Yoshikawa Y., Ishiguro H. Can a humanoid robot continue to draw attention in an office environment? // Advanced Robotics. 2020. Vol. 34. № 14. P. 931–946. DOI: [10.1080/01691864.2020.1769724](https://doi.org/10.1080/01691864.2020.1769724).
31. Knoblich G., Ohlsson S., Haider H., Rhenius D. Constraint relaxation and chunk decomposition in insight problem solving // Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999. Vol. 25. № 6. P. 1534–1555. DOI: [10.1037/0278-7393.25.6.1534](https://doi.org/10.1037/0278-7393.25.6.1534).
32. Спиридонов В.Ф., Ерофеева М.А., Кловайт Н.О., Ардисламов В.В., Морозов М.И., Здилар С. Репликация эффектов интерактивного решения задач спичечной алгебры // Психологические исследования. 2021. Т. 14. № 79. С. 1–46. DOI: [10.54359/ps.v14i79.119](https://doi.org/10.54359/ps.v14i79.119).
33. Öllinger H.M., Jones G., Knoblich G. Heuristics and representational change in two-move matchstick tasks //

СПИСОК ЛИТЕРАТУРЫ

1. Morandini S., Curro F., Parlangei O., Pietrantonio L. Collaborative Robots Adapting Their Behavior Based on Workers' Psychological States: A Systematic Scoping Review // Human Behavior and Emerging Technologies. 2025. Vol. 2025. № 1. Article number 6361777. DOI: [10.1155/hbe2/6361777](https://doi.org/10.1155/hbe2/6361777).
2. Garcia-Martinez J., Gamboa-Montero J.J., Castillo J.C., Castro-Gonzalez A. Analyzing the Impact of Responding to Joint Attention on the User Perception of the Robot in Human-Robot Interaction // Biomimetics. 2024. Vol. 9. № 12. Article number 769. DOI: [10.3390/biomimetics9120769](https://doi.org/10.3390/biomimetics9120769).
3. Галин Р.Р., Серебряный В.В., Тевяшов Г.К., Широкий А.А. Взаимодействие человека и робота в коллаборативных робототехнических системах // Известия Юго-Западного государственного университета. 2020. Т. 24. № 4. С. 180–199. DOI: [10.21869/2223-1560-2020-24-4-180-199](https://doi.org/10.21869/2223-1560-2020-24-4-180-199).
4. Cochet H., Guidetti M. Contribution of Developmental Psychology to the Study of Social Interactions: Some Factors in Play, Joint Attention and Joint Action and Implications for Robotics // Frontiers in Psychology. 2018. Vol. 9. Article number 1992. DOI: [10.3389/fpsyg.2018.01992](https://doi.org/10.3389/fpsyg.2018.01992).
5. Urakami J., Seaborn K. Nonverbal Cues in Human-Robot Interaction: A Communication Studies Perspective // ACM Transactions on Human-Robot Interaction. 2023. Vol. 12. № 2. P. 1–21. DOI: [10.1145/3570169](https://doi.org/10.1145/3570169).
6. Jing Yang, Barragan J.A., Farrow J.M., Sundaram C.P., Wachs J.P., Denny Yu. An Adaptive Human-Robotic Interaction Architecture for Augmenting Surgery Performance Using Real-Time Workload Sensing – Demonstration of a SemiAutonomous Suction Tool // Human Factors. 2024. Vol. 4. P. 1081–1102. DOI: [10.1177/00187208221129940](https://doi.org/10.1177/00187208221129940).
7. Salminen-Saari J.F.A., Moreno-Esteve E.G., Haataja E., Toivanen M., Hannula M.S., Laine A. Phases of collaborative mathematical problem solving and joint atten-

- Advances in Cognitive Psychology. 2006. Vol. 2. № 4. P. 239–253. DOI: [10.2478/v10053-008-0059-3](https://doi.org/10.2478/v10053-008-0059-3).
19. Scorza Azzarà G., Zonca J., Rea F., Joo-Hyun Song, Sciutti A. Collaborating with a robot biases human spatial attention // iScience. 2025. Vol. 28. № 7. Article number 112791. DOI: [10.1016/j.isci.2025.112791](https://doi.org/10.1016/j.isci.2025.112791).
20. Pöysä-Tarhonen J., Shupin Li, Hautala J., Awwal N., Häkkinen P. Making the Invisible Visible: Exploring Joint Attention Behaviour in Remote Collaborative Problem- Solving // Journal of Computer Assisted Learning. 2025. Vol. 41. № 4. Article number e70068. DOI: [10.1111/jcal.70068](https://doi.org/10.1111/jcal.70068).
21. Ehrlich S.K., Dean-Leon E., Tacca N., Armleder S., Dimova-Edeleva V., Gordon Cheng. Human-robot collaborative task planning using anticipatory brain responses // PLoS ONE. 2023. Vol. 8. № 7. Article number e0287958. DOI: [10.1371/journal.pone.0287958](https://doi.org/10.1371/journal.pone.0287958).

УДК 159.9.072

doi: <https://doi.org/10.18323/3034-2996-2026-1-64-3>

Имитация совместного внимания при взаимодействии робота и человека (HRI) при решении двухспичечных задач

Столярова Анастасия Николаевна, аспирант, инженер-исследователь

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Москва (Россия)

Национальный исследовательский центр «Курчатовский институт», Москва (Россия)

E-mail: anastasiyas050298@gmail.com

ORCID: <https://orcid.org/0009-0005-7155-0426>

Поступила в редакцию 06.02.2026

Пересмотрена 19.03.2026

Принята к публикации 19.03.2026

Аннотация: Проблема. В области человеко-роботного взаимодействия (Human-Robot Interaction, HRI) несмотря на развитие антропоморфных роботов, способных инициировать совместное внимание, остаётся неясным, как различные типы подсказок (жест+взгляд vs. только взгляд) и их правильность влияют на автоматическое следование человека и его субъективную оценку взаимодействия. **Цель.** Сравнить эффективность различных типов подсказок антропоморфного робота (указательные жесты, совмещённые со взглядом, и подсказки только взглядом) в задаче, требующей совместного внимания, а также оценить влияние правильности подсказок на поведение испытуемых. **Методы.** В исследовании приняли участие 43 студента РАНХИГС: 38 девушек и 5 юношей в возрасте от 19 до 27 лет ($M=20,51$; $SD=1,82$). Количество движений испытуемых, следующих за подсказками робота и совпадающих с направлением подсказки, оценивались для оценки эффективности подсказок робота в каждом из условий. Для изучения реакции испытуемых на подсказки после каждой задачи применялся опросник, созданный на основе метакогнитивных шкал Данек. **Результаты.** Результаты показали, что способность робота имитировать процесс совместного внимания при решении задач с испытуемым была доказана. Гипотеза о большей эффективности подсказок робота с помощью указательных жестов, объединённых с перемещением взгляда и головы робота, по сравнению с подсказками подтвердилась. Гипотеза о большей эффективности правильных подсказок по сравнению с неправильными подсказками робота подтвердилась. **Выводы.** Способность робота имитировать процесс совместного внимания при решении задач с испытуемым была доказана, испытуемые обращали внимание на подсказки робота и старались следовать им как в условиях правильных подсказок, так и неправильных. Однако в условиях с правильными подсказками процент совпадающих по направлению с подсказками попыток решения был значительно выше, чем в условиях с неправильными подсказками.

Ключевые слова: взаимодействие человека и робота; Human-Robot Interaction; HRI; антропоморфные роботы; совместное внимание; подсказки взглядом и жестами; модифицированные задачи Кноблиха; метакогнитивные шкалы Данек

Благодарности: Работа выполнена в рамках государственного задания НИЦ «Курчатовский институт». Разработка методики, в том числе экспериментальные условия, жесты и стратегии общения антропоморфного интерфейса было выполнено в рамках государственного задания НИЦ «Курчатовский институт». Статистический анализ результатов эксперимента, включающий сравнения экспериментальных условий, анализ субъективных и поведенческих оценок участников выполнялось в рамках гранта Российского научного фонда № 25-78-10154, <https://rscf.ru/project/25-78-10154/>.

Для цитирования: Столярова А.Н. Имитация совместного внимания при взаимодействии робота и человека (HRI) при решении двухспичечных задач // Доказательная педагогика, психология. 2026. № 1. С. 27–36. DOI: <https://doi.org/10.18323/3034-2996-2026-1-64-3>.