Real-World Implicit Association Task for Studying Mind Perception: Insights for Social Robotics

Tuğçe Nur Pekçetin
Department of Cognitive Science
Graduate School of Informatics
Middle East Technical University
Ankara, Turkey
tugce.bozkurt@metu.edu.tr

Şeyda Evsen
Department of Neuroscience
Bilkent University
Ankara, Turkey
seyda.evsen@bilkent.edu.tr

Serkan Pekçetin
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
serkan.pekcetin@metu.edu.tr

Cengiz Acarturk
Department of Cognitive Science
Jagiellonian University
Krakow, Poland
cengiz.acarturk@uj.edu.pl

Burcu A. Urgen
Department of Psychology
Department of Neuroscience
Bilkent University, UMRAM
Ankara, Turkey
burcu.urgen@bilkent.edu.tr

ABSTRACT

In response to the growing demand for enhanced integration of implicit measurements in Human-Robot Interaction (HRI) research, the need for studies involving physically present robots, and the calls for a transition from lab experiments to more naturalistic investigations, we introduce the Real-World Implicit Association Task (RW-IAT). This report outlines the versatile methodology of the RW-IAT; emphasizing its allowance to present real-life stimuli and capture behavioral data, including response times and mouse tracking metrics in a controlled manner. Sample analyses focusing on communicative and noncommunicative actions between a human actor and the Pepper robot reveal significant effects on the Agency and Experience dimensions of the mind perception. We believe the methodology we proposed will contribute to conducting ecologically valid research in the field of HRI in real-world contexts.

CCS CONCEPTS

• **Human-centered computing** → **Laboratory experiments**.

KEYWORDS

social robotics, human-robot interaction, mind perception; agency; experience, implicit association task; IAT, real-time; naturalistic experiments

1 INTRODUCTION

In Human-Robot Interaction (HRI) research, the widely-used Wizard-of-Oz (WoZ) experimental methodology [23] involves participants interacting with a robot, unaware of human operator control [41]. Beyond its functional inspiration, analyzing characters such as the Scarecrow and Tin Woodman from this iconic novel [2] reveals nuanced aspects of how our observations influence our perception of others - a core aspect of mind perception discussions in HRI. Despite the Scarecrow's "apparent" *lack of a brain*, their consistent displays of wisdom challenge conventional notions, prompting attributions of *mental capacities*, such as rationality. Similarly, the Tin Woodman, with a metallic exterior and *no physical heart*, embodies *emotional intelligence*, leading to attributions, such as empathy. These characters exemplify our human capacity to attribute cognitive and emotional abilities, transcending anatomical structures and emphasizing the multifaceted nature of mind perception [7].

Associating the brain with cognitive functions and the heart with emotional capabilities are recurring themes in literature and research. Gray et al. [16] proposed two main dimensions of mind perception: *Agency*, the ability to do, and *Experience*, the ability to feel. While further studies exploring the possible dimensions of mind perception [29, 48] offer valuable insights, their reliance on self-reports can make their results challenging to reconcile. Acknowledging the limitations of explicit measurements highlights the need for extending the use of implicit measurements [10, 17, 32, 45], grounded in behavioral or neuroimaging data, for in-depth explorations of complex cognitive processes, such as mind perception.

Similarly, attributing mental states to entities, adopting the *intentional stance* [4, 5], is not limited to fiction; mind perception can arise from the need to understand and predict others' behavior [7]. Consistent with causes and consequences outlined by Waytz et al. [47], prior research demonstrated that robots' expressions of social behavior [12, 44], or gestures [42] influence mental state attributions. However, these studies primarily utilized texts, images, or a combination, as discussed in a recent comprehensive review [45]. Notably, humans tend to anthropomorphize a robot more strongly when it is physically present [24, 27, 44].
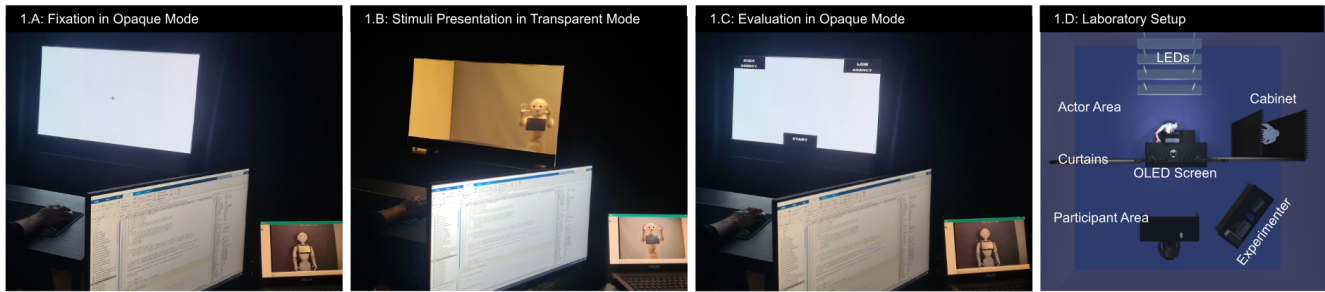
Figure 1: Real-World Implicit Association Task (RW-IAT) setup overview. (A, B, C) The RW-IAT from the experimenter's view. (D) The naturalistic laboratory setup from top-down view.

Among its counterparts, the original Implicit Association Test (IAT) by Greenwald et al. [19], stands out as the most widely used psychological tool assessing implicit attitudes. Participants' reaction times to two conditions of *target stimuli* indicate their implicit attitudes toward the targets based on *attribute stimuli*. For example, in the gender-science IAT [33], participants categorize Male and Female attributes under Science and Liberal Arts. Faster and more accurate categorization in the initial condition (Male = Science) implies stronger associations between science and male attributes. In a recent study, Li et al. [28] introduced the Mind Perception IAT (MP-IAT) to examine mind perception in human-robot interaction. This test measures mental attributions on Agency and Experience dimensions, utilizing attributes from the High and Low ends of these dimensions and using images of humans and humanoid robots as targets. While the MP-IAT is a valuable method with the potential to test concepts, images, or videos as target stimuli, it remains a computerized task, limiting its ability to present real-world stimuli.

The scarcity of studies on how individuals perceive the physically present and active robots, coupled with the necessity for real-world investigations outside controlled labs to enhance ecological validity [21, 45], underscores the need for conducting WoZ studies with actual robots [6]. In this report, we address these gaps as well as the need for implicit measurements by introducing the *Real-World Implicit Association Test (RW-IAT)* to implicitly study mind perception in real-time with physically present robots and humans and live actions, while capturing participants' response times and mouse trajectories in a controlled manner. RW-IAT is a WoZ experiment in a lab environment, yet it uses real-life social stimuli, corresponding to the last level before the fully naturalistic studies [9], serving as a further step towards "HRI in the real world."

2 METHODS AND MATERIALS

2.1 Participants

As part of a broader social robotics project [35], we collected in-person data from 160 participants from four generations (ages 18 to 73). Since this report aims to introduce RW-IAT details, for aligning with prior research [28] ($N = 53$) and facilitating comparisons with studies representing conventional age ranges, we present data from 55 participants (ages 18-35, $M = 25.93$, $SD = 5.59$, 33 females). This study was approved by the Human Research Ethics Committee of Bilkent University. All participants provided informed consent and they received approximately 3 USD in compensation.

2.2 Materials

Target (Action) Stimuli: Following IAT best practices [18], we selected "easy stimuli" systematically via norming. Using Android Studio and Pepper SDK's Animation Editor IDE [8], we created 40 animations based on action clusters [34] and datasets of communicative and noncommunicative actions [30, 50]. We shot standardized videos, consistent in length, aspect ratio, trajectories, and repetitions, with both the Pepper robot and the human actor. 438 participants tested them in two separate online studies [36], in which they identified actions, rated confidence levels, and categorized them as communicative or noncommunicative. We annotated the open-ended answers and calculated the H entropy [43], and the k-means clustering algorithm [20, 46] yielded the four most straightforward communicative (peek-a-boo, saluting, throwing a kiss, hand-waving) and noncommunicative (shooting an arrow, jogging, drinking, driving) actions across human and robot conditions.
Attribute (Lexical) Stimuli: The target stimuli were live actions, the attribute stimuli were verbal concepts: *High Agency*, *Low Agency*, *High Experience*, and *Low Experience*. Following IAT best practices [18], participants underwent training to familiarize themselves with the Agency and Experience dimensions and their High and Low ends. Training concepts were validated in an online norming study with 274 participants, as documented in our previous work [37].

2.3 Procedure

To prevent any familiarity- or interaction-related biases, the participants provided demographic data and consent and underwent additional tests in a separate room. Only after these procedures, they entered the main experiment room where the human and robot actors were positioned behind a curtain and an experimenter was waiting to supervise the session. The *Lexical Training* involved detailed descriptions of Agency and Experience concepts. Participants evaluated twelve concepts for each dimension, receiving feedback until reaching 80% accuracy. Subsequently, participants learned the High and Low ends of Agency and Experience dimensions, evaluating six concepts for each until achieving 80% accuracy. Following Lexical Training, participants moved to *Action Identification*, observing eight actions performed by human and robot actors, respectively. They verbally identified each action, serving as familiarization and manipulation checks. After these steps, participants started the RW-IAT.

2.4 Real-World IAT (RW-IAT)

The RW-IAT utilizes a 55" OLED display, depicted in Figure 1, that functions as both a standard monitor and a transparent screen for presenting real-life stimuli. It transitions between an opaque state for instructions and prompts (Figure 1.A) or evaluations (Figure 1.C) and a transparent state during stimulus presentations (Figure 1.B). The RW-IAT uniquely features real-time, live stimuli, distinguishing it from existing computerized IATs. The use of the OLED screen maintains strict experimental control, setting it apart from fully naturalistic experiments that involve sacrificing control. The use of a single platform for stimuli presentation and evaluations in the RW-IAT aims for a seamless transition between tasks, minimizing potential disruptions in behavioral data crucial for precise response times and mouse trajectories [39].

With inspiration from the Single Category IAT [22], in RW-IAT, we tested the human and robot actor in four separate blocks to address the non-complementary nature of these categories. Block contents included Robot-Agency, Robot-Experience, Human-Agency, and Human-Experience, and their order was counterbalanced across participants while each actor performed the eight actions in a randomized order within the blocks. Participants were instructed to observe each action and actor on a transparent screen carefully and then evaluate what they saw in terms of the required capacity. For instance, in the Robot-Experience Block, participants watched the live "peek-a-boo" action performed by the robot actor. After six seconds —the duration of all actions—, the screen turned opaque, revealing "High Experience" and "Low Experience" choices in the upper left and right corners. Participants clicked on one, deciding whether what they saw required High or Low Experience capacity. Although the participants were encouraged to keep the definitions and examples in the training session in mind during their evaluations, the experimenter also emphasized the need for quick and instinctive responses. Both human and robot actors executed all actions identically to observe potential variations in participants' judgments based on the actor or action or both. After evaluating each of the eight actions in one block, participants waited for the actors to replace, marking the start of the next block.

Figure 1.D illustrates the lab setup with a curtain system dividing participant and actor areas. The participant area includes tables and devices for the experimenter and participants, while the actor area features a cabinet for the actors to wait between blocks and a laptop, positioned on top of a table, displaying the next action to inform the human actor. Ceiling LEDs enhance the transparency or the opaqueness of the screen. Background music is used during actor replacements, and the experimenter observes backstage through a security camera. The task was implemented using Psychtoolbox-3 [3, 25, 40] on MATLAB R2022a. Full details of the laboratory setup, including hardware, software, and materials, are extensively documented in our prior work [38].

The RW-IAT involved two independent variables: Actor type: robot or human and Action type: communicative or noncommunicative. Participants assessed each actor and action combination in both Agency and Experience dimensions. Dependent variables included *Response Time (RT)*, the elapsed time in seconds between the end of an action and the occurrence of a mouse click on one of the response alternatives, *Maximum Deviation* of a Trajectory

(MD), and *Area Under the Curve (AUC)* for mouse trajectories, calculated following the original work [14] and interpreted as indicators of participant hesitations [13, 49]. The fourth dependent variable, *Response*, was categorical, categorized as either High or Low.

3 RESULTS

We processed and visualized data with MATLAB 2023b and performed analyses using RStudio 2023.09.1. After identifying outliers (eight trials), we categorized the dataset (1750 trials) into Agency and Experience blocks and calculated participants' average scores for each dependent variable. Since the Shapiro-Wilk test revealed significant departures from normal distribution ($p < .001$) for both actor and action types in both block dimensions, we used Friedman's ANOVA [15] as a nonparametric alternative [11] and conducted post hoc analyses using the Nemenyi test [31] to identify differences between data groups. We present the results separately for each dependent variable in the Agency and Experience blocks.

3.1 Agency Dimension

3.1.1 Actor Type. The RTs ($\chi^2(1) = 2.20$, $p = .14$), MDs ($\chi^2(1) = 0.16$, $p = .69$), and AUCs values ($\chi^2(1) = 0.16$, $p = .69$) showed no significant changes across actor types. However, the ratio of High responses significantly varied by actor type, $\chi^2(1) = 5.49$, $p = .02$ (see Actor Type in Figure 2.C). Pairwise comparisons indicated that the ratio of High Agency responses for the robot actor was significantly higher than for the human actor, $p = .04$.
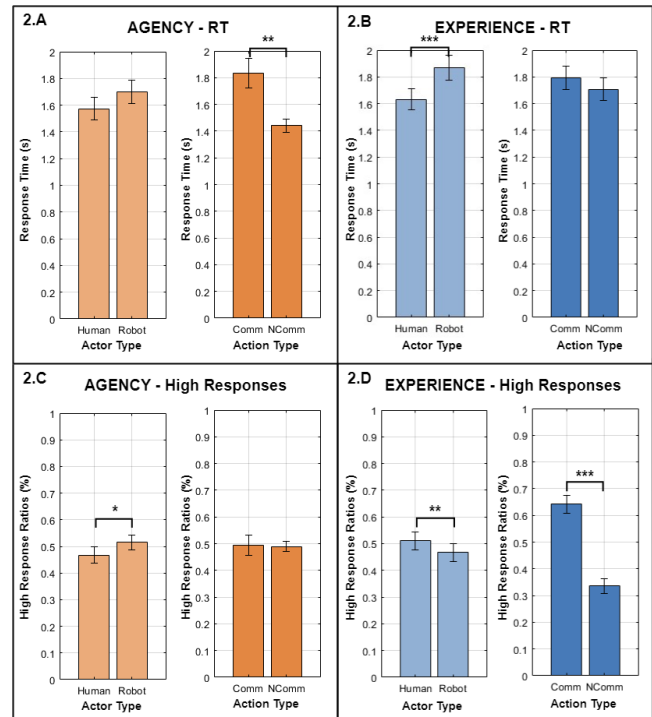
Figure 2: Mean response times (RT) and High response ratios across conditions.
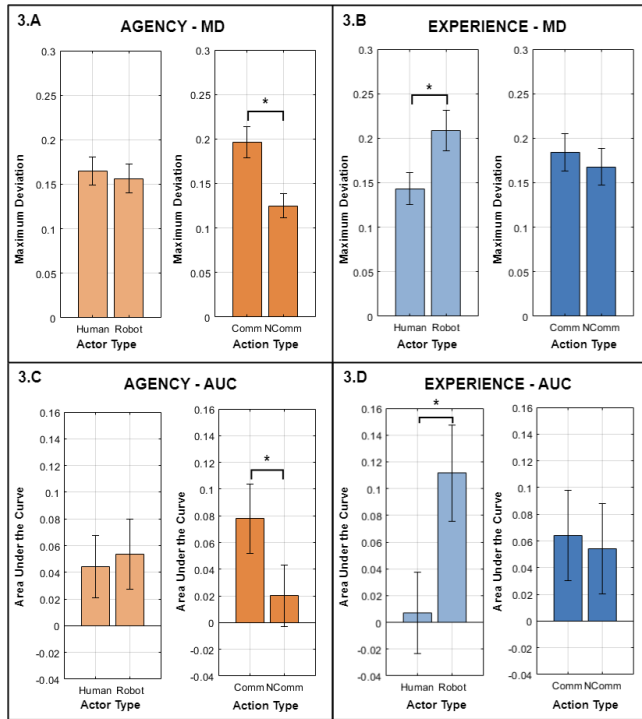
Figure 3: Mean maximum deviation (MD) and area under the curve (AUC) values across conditions.

3.1.2 Action Type. There was a main effect of action type on the RTs, $\chi^2(1)$ = 8.02, p = .005 (see Action Type in 2.A); MDs, $\chi^2(1)$ = 6.56, p = .01 (see Action Type in 3.A); and AUC values, $\chi^2(1)$ = 6.56, p = .01 (see Action Type in 3.C), but not on the ratio of the High responses ($\chi^2(1)$ = .022, p = .88). Pairwise comparisons revealed that the RT, p = .005; MD, p = .01; and AUC values, p = .01 were significantly longer, higher, and larger, respectively, for communicative actions compared to noncommunicative actions.

3.2 Experience Dimension

3.2.1 Actor Type. The type of the actor significantly affected the RTs, $\chi^2(1)$ = 11.36, p < .001 (see Actor Type in Figure 2.B); MDs, $\chi^2(1)$ = 4.09, p = .04 (see Actor Type in Figure 3.B); AUC values, $\chi^2(1)$ = 5.25, p = .02 (see Actor Type in Figure 3.D), and the ratios of the High responses, $\chi^2(1)$ = 7.11, p = .008 (see Actor Type in Figure 2.D). Pairwise comparisons revealed that the robot actor elicited significantly longer RT, p < .001; higher MD, p = .04; and larger AUC values, p = .02, but a significantly lower ratio of High Experience responses compared to the human actor, p = .03.

3.2.2 Action Type. The RTs ($\chi^2(1)$ = 0.45, p = .50), MDs ($\chi^2(1)$ = 0.02, p = .89), and AUC values ($\chi^2(1)$ = 0.45, p = .50) did not significantly change across action types. However, there was a significant effect of the action type on the ratio of High responses of the participants, $\chi^2(1)$ = 15.87, p < .001 (see Action Type in Figure 2.D). Pairwise comparisons revealed that the ratio of High Experience responses for the communicative actions was significantly higher compared to the ratio for the noncommunicative actions, p < .001.

4 DISCUSSION

We introduced the RW-IAT, which we designed by following the best practices of creating an IAT [18], to explore mind perception implicitly, using real-world stimuli performed by both human and robot actors. With the RW-IAT, we investigated how attributions of mental capacity on the Agency and Experience dimensions of mind perception change across different actor and action types. In addition to traditional IAT metrics such as responses and response times, we also tracked participants' mouse trajectories. This additional data enables further investigations into the cognitive processes involved [13], aligning with discussions [49] and recommendations in previous work [45].

In the results, action type notably affected behavioral metrics in the Agency dimension, with communicative actions leading to longer response times and increased hesitations. The robot actor showed a significantly higher response ratio in the Agency dimension, possibly due to participants' lowered expectations of its capabilities, as reflected in their surprised comments while they watched the actions such as "The robot can do everything a human can do!" or "They perform almost the same!", reflecting the influence of the live-action performance of the robot. In contrast, a human performing similar actions might have stayed within or below participants' agency thresholds. Previous studies also suggest a reduced difference in mental state attributions between robots and humans when their behaviors are tested [1, 26].

On the contrary, in the Experience dimension, the effect of the actor type prevails, with the robot actor eliciting longer decision times and more hesitations, yet leading to significantly lower Experience scores compared to the human actor. Notably, action type significantly affected response ratios, with communicative actions yielding higher rates than noncommunicative ones. The observed hesitation differences between human and robot actors may be attributed to the nature of communicative actions. Previous studies have suggested that when robots engage in social behaviors, people's judgments can change [12, 42, 44]. However, robots still received lower attributions, as indicated in most prior work [45].

In conclusion, the RW-IAT shows promise based on observed outcomes in the sample data. While the investigations in applied domains of HRI often extend into real-world scenarios, investigations regarding the dynamics of cognitive processes, such as mind perception, may still require experimental control for behavioral data accuracy. While initially designed for studying mind perception in HRI, RW-IAT's methodology could be adapted to the studies from various domains deploying neuroimaging or eye-tracking methodology and using real-life stimuli with precise timing and pre-defined conditions. Future directions include further validation of this methodology, and we view this naturalistic yet controlled approach as a significant step towards increased ecological validity, embracing investigations into HRI in real-world scenarios.

REFERENCES

[1] Abdulaziz Abubshait and Eva Wiese. 2017. You Look Human, But Act Like a Machine: Agent Appearance and Behavior Modulate Different Aspects of Human–Robot Interaction. *Frontiers in Psychology* 8 (2017). https://doi.org/10.3389/fpsyg.2017.01393

[2] L. Frank Baum. 1900. *The Wonderful Wizard of Oz.* G.M. Hill Company.

[3] David H. Brainard and Spatial Vision. 1997. The Psychophysics Toolbox. *Spatial Vision* 10, 4 (1997), 433–436.

[4] Daniel C. Dennett. 1989. *The Intentional Stance.* MIT Press.

[5] Daniel C. Dennett. 1997. True Believers:The Intentional Strategy and Why It Works. In *Mind Design II: Philosophy, Psychology, and Artificial Intelligence.*

[6] Christian Dondrup, Lynne Baillie, Frank Broz, and Katrin Lohan. 2018. How can we transition from lab to the real world with our HCI and HRI setups. In *Proceedings of the 4th Workshop on Public Space Human-Robot Interaction (PubRob), Barcelona, Spain*, Vol. 3.

[7] Nicholas Epley, Adam Waytz, et al. 2010. Mind Perception. *Handbook of Social Psychology* 1, 5 (2010), 498–541.

[8] SoftBank Robotics Europe. 2019. *Animation Editor.* Retrieved December 6, 2023 from https://qisdk.softbankrobotics.com/sdk/doc/pepper-sdk/ch3_tools/animation_editor.html

[9] S. Fan, O. Dal Monte, and S.W.C. Chang. 2021. Levels of naturalism in social neuroscience research. iScience, 24 (7), 102702.

[10] Russell H. Fazio and Michael A. Olson. 2003. Implicit Measures in Social Cognition Research: Their Meaning and Use. *Annual Review of Psychology* 54, 1 (2003), 297–327. https://doi.org/10.1146/annurev.psych.54.101601.145225 PMID: 12172003.

[11] Andy Field, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R.* Sage Publications.

[12] Marlena R. Fraune, Benjamin C. Oisted, Catherine E. Sembrowski, Kathryn A. Gates, Margaret M. Krupp, and Selma Šabanović. 2020. Effects of robot-human versus robot-robot behavior and entitativity on anthropomorphism and willingness to interact. *Computers in Human Behavior* 105 (2020), 106220. https://doi.org/10.1016/j.chb.2019.106220

[13] Jonathan Freeman, Rick Dale, and Thomas Farmer. 2011. Hand in Motion Reveals Mind in Motion. *Frontiers in Psychology* 2 (2011). https://doi.org/10.3389/fpsyg.2011.00059

[14] Jonathan B. Freeman and Nalini Ambady. 2010. MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods* 42, 1 (2010), 226–241.

[15] Milton Friedman. 1937. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *J. Amer. Statist. Assoc.* 32, 200 (1937), 675–701.

[16] Heather M. Gray, Kurt Gray, and Daniel M. Wegner. 2007. Dimensions of Mind Perception. *Science* 315, 5812 (2007), 619–619. https://doi.org/10.1126/science.1134475 arXiv:https://www.science.org/doi/pdf/10.1126/science.1134475

[17] Anthony G. Greenwald and Mahzarin R. Banaji. 2017. The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist* 72, 9 (2017), 861–871.

[18] Anthony G. Greenwald, Miguel Brendl, Huajian Cai, Dario Cvencek, John F. Dovidio, Malte Friese, Adam Hahn, Eric Hehman, Wilhelm Hofmann, Sean Hughes, et al. 2022. Best research practices for using the Implicit Association Test. *Behavior Research Methods* 54, 3 (2022), 1161–1180.

[19] Anthony G. Greenwald, Debbie E. McGhee, and Jordan L.K. Schwartz. 1998. Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74, 6 (1998), 1464.

[20] J. A. Hartigan and M. A. Wong. 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 1 (1979), 100–108. http://www.jstor.org/stable/2346830

[21] Malte Jung and Pamela Hinds. 2018. Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction. , 5 pages.

[22] Andrew Karpinski and Ross B. Steinman. 2006. The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology* 91, 1 (2006), 16–32.

[23] John F. Kelley. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems (TOIS)* 2, 1 (1984), 26–41.

[24] Sara Kiesler, Aaron Powers, Susan R. Fussell, and Cristen Torrey. 2008. Anthropomorphic Interactions with a Robot and Robot–like Agent. *Social Cognition* 26, 2 (2008), 169–181. https://doi.org/10.1521/soco.2008.26.2.169

[25] Mario Kleiner, David Brainard, and Denis Pelli. 2007. What's new in Psychtoolbox-3? (2007).

[26] Daniel T. Levin, Stephen S. Killingsworth, Megan M. Saylor, Stephen M. Gordon, and Kazuhiko Kawamura. 2013. Tests of Concepts About Different Kinds of Minds: Predictions About the Behavior of Computers, Robots, and People. *Human–Computer Interaction* 28, 2 (2013), 161–191. https://doi.org/10.1080/07370024.2012.697007

[27] Jamy Li. 2015. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies* 77 (2015), 23–37.

[28] Zhenni Li, Leonie Terfurth, Joshua Pepe Woller, and Eva Wiese. 2022. Mind the Machines: Applying Implicit Measures of Mind Perception to Social Robotics. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* 236–245. https://doi.org/10.1109/HRI53351.2022.9889356

[29] Bertram F. Malle. 2019. How Many Dimensions of Mind Perception Really Are There?. In *CogSci.* 2268–2274.

[30] Valeria Manera, Ben Schouten, Cristina Becchio, Bruno G. Bara, and Karl Verfaillie. 2010. Inferring intentions from biological motion: a stimulus set of point-light communicative interactions. *Behavior Research Methods* 42, 1 (2010), 168–178.

[31] Peter Bjorn Nemenyi. 1963. *Distribution-free multiple comparisons.* Princeton University.

[32] Brian A. Nosek, Carlee Beth Hawkins, and Rebecca S. Frazier. 2011. Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences* 15, 4 (2011), 152–159.

[33] Brian A. Nosek, Frederick L. Smyth, Natarajan Sriram, Nicole M. Lindner, Thierry Devos, Alfonso Ayala, Yoav Bar-Anan, Robin Bergh, Huajian Cai, Karen Gonsalkorale, et al. 2009. National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences* 106, 26 (2009), 10593–10597.

[34] G.A. Orban, M. Lanzilotto, and L. Bonini. 2021. From Observed Action Identity to Social Affordances. *Trends in Cognitive Sciences* 25, 6 (2021), 493–505.

[35] Tuğçe Nur Pekçetin, Cengiz Acarturk, and Burcu A Urgen. 2024. Investigating Mind Perception in HRI through Real-Time Implicit and Explicit Measurements. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction.* https://doi.org/10.1145/3610978.3638366

[36] Tuğçe Nur Pekçetin, Gaye Aşkın, Şeyda Evsen, Tuvana Dilan Karaduman, Asli Eroglu, Badel Barinal, Jana Tunç, and Burcu A Urgen. 2023. A Set of Communicative and Noncommunicative Action Video Stimuli for Human-Robot Interaction Research. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 45.

[37] Tuğçe Nur Pekçetin, Badel Barinal, Jana Tunç, Cengiz Acarturk, and Burcu A Urgen. 2023. Studying Mind Perception in Social Robotics Implicitly: The Need for Validation and Norming. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction.* 202–210.

[38] Tuğçe Nur Pekçetin, Şeyda Evsen, Serkan Pekçetin, Cengiz Acarturk, and Burcu A Urgen. 2023. A Naturalistic Setup for Presenting Real People and Live Actions in Experimental Psychology and Cognitive Neuroscience Studies. *JoVE (Journal of Visualized Experiments)* 198 (2023), e65436.

[39] Tuğçe Nur Pekçetin, Şeyda Evsen, Serkan Pekçetin, Tuvana Dilan Karaduman, Cengiz Acarturk, and Burcu A Urgen. 2024. A Naturalistic Laboratory Setup for Real-World HRI Studies. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction.* https://doi.org/10.1145/3610978.3641113

[40] Denis G. Pelli and Spatial Vision. 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 10 (1997), 437–442.

[41] Laurel D. Riek. 2012. Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1, 1 (2012), 119–136.

[42] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2011. Effects of Gesture on the Perception of Psychological Anthropomorphism: A Case Study with a Humanoid Robot. In *Social Robotics: Third International Conference, ICSR 2011, Amsterdam, The Netherlands, November 24-25, 2011. Proceedings 3.* Springer, 31–41.

[43] C. E. Shannon. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27, 3 (1948), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[44] Ilona Straub. 2016. 'It looks like a human!' The interrelation of social presence, interaction and agency ascription: a case study about the effects of an android robot on social agency ascription. *AI & Society* 31 (2016), 553–571.

[45] Sam Thellman, Maartje de Graaf, and Tom Ziemke. 2022. Mental State Attribution to Robots: A Systematic Review of Conceptions, Methods, and Findings. *J. Hum.-Robot Interact.* 11, 4, Article 41 (sep 2022), 51 pages. https://doi.org/10.1145/3526112

[46] Robert L. Thorndike. 1953. Who belongs in the family. *Psychometrika* (1953), 267–276. https://doi.org/10.1007/BF02289263

[47] Adam Waytz, Kurt Gray, Nicholas Epley, and Daniel M. Wegner. 2010. Causes and consequences of mind perception. *Trends in Cognitive Sciences* 14, 8 (2010), 383–388. https://doi.org/10.1016/j.tics.2010.05.006

[48] Kara Weisman, Carol S. Dweck, and Ellen M. Markman. 2017. Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences* 114, 43 (2017), 11374–11379. https://doi.org/10.1073/pnas.1704347114

[49] Zonghuo Yu, Fei Wang, Dengfeng Wang, and Mike Bastin. 2012. Beyond Reaction Times: Incorporating Mouse-Tracking Measures into the Implicit Association Test to Examine its Underlying Process. *Social Cognition* 30, 3 (2012), 289–306. https://doi.org/10.1521/soco.2012.30.3.289

[50] Hazlin Zaini, Jonathan M. Fawcett, Nicole C. White, and Aaron J Newman. 2013. Communicative and noncommunicative point-light actions featuring high-resolution representation of the hands and fingers. *Behavior Research Methods* 45, 2 (2013), 319–328.