Physical Form Shapes Agency Attribution: A Multi-Robot Study of Embodiment, Perceived Intentionality, and Trust

Ali Fallahi[1], Patrick Holthaus[1], Farshid Amirabdollahian[1], and Gabriella Lakatos[1]

Abstract— **How does a robot's physical form shape the amount of agency that people attribute to it, and does that attribution translate into differences in trust? This paper presents an in-person experiment in which 50 participants interacted with two robots of contrasting embodiment — Pepper, a humanoid social robot, and Fetch, a mobile manipulator — under matched High Agency (HA) and Low Agency (LA) behavioural conditions. Agency was manipulated via a theoretically grounded, multi-channel framework operationalising five factors: Goal Directness, Social Awareness, Accountability, Shared Decision-Making, and Embodiment. Perceived agency was measured with the PA scale; trust and social perception were assessed via the MDMT and RoSAS. We present the experimental design, manipulation framework, and hypotheses, and discuss the expected contributions of this ongoing work for trust calibration design, multi-robot study methodology, and embodiment-sensitive agency manipulation.**

I. Introduction

Trust is central to effective human-robot interaction (HRI), and a key antecedent of trust is the degree of agency people attribute to a robot — the extent to which its actions appear driven by its own intentions rather than external commands [1]. As robots become socially present in homes and workplaces, understanding how perceived agency forms and how it can be reliably manipulated has become essential for researchers wishing to study its effects on trust in rigorous and replicable ways.

Prior work by our group demonstrated that text-based autonomy framing — telling participants a robot was autonomous versus remotely controlled — produced effects only on the Sincerity dimension of the MDMT [2], motivating a richer, behaviourally embedded manipulation. A further gap is that prior agency studies have overwhelmingly used single robot platforms, making it impossible to separate the effect of the manipulation from the baseline agency afforded by the robot's physical form.

Physical presence increases perceived agency even when actual autonomy is held constant [3], and the nature of a robot's possible actions modulate whether intentionality is attributed [4]. A humanoid robot like Pepper affords rich social signalling; a mobile manipulator like Fetch affords physical task handling. These differences may produce fundamentally different agency attributions even when verbal behaviour is matched — a possibility single-platform studies cannot detect.

[1]All authors are with the School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield, United Kingdom. {a.fallahi, p.holthaus, f.amirabdollahian2, g.lakatos}@herts.ac.uk

The experiment presented in this paper addresses both gaps. We present a mixed-design experiment comparing two physically distinct robots under matched HA/LA behavioural conditions, and report on data collection and methodology. The study pursues the following research questions:

- **RQ1** Does a multi-channel behavioural agency manipulation (HA vs. LA) produce reliable differences in perceived agency, trust, and social perception beyond those achievable through autonomy framing alone?
- **RQ2** How does robot embodiment (Pepper vs. Fetch) influence which *dimensions* of agency are attributed, and does embodiment moderate the effect of the behavioural manipulation on trust?
- **RQ3** Do pre-existing attitudes toward robots (NARS) moderate perceived agency or trust, replicating and extending findings from prior autonomy-framing work [2]?

Based on these research questions and prior literature, we propose the following directional hypotheses.

For RQ1: (H1a) Participants in the HA condition will report significantly higher composite PA scores than those in the LA condition, for both robots; (H1b) the HA condition will yield higher MDMT scores, expected most strongly on Sincerity; (H1c) the HA condition will yield higher RoSAS Warmth and Competence, and lower Discomfort.

For RQ2: (H2a) Fetch will be perceived as more goal-directed and responsible than Pepper, reflecting its physical task-handling affordances; (H2b) Pepper will be perceived as more socially aware than Fetch, reflecting its humanoid social signalling affordances; (H2c) embodiment will moderate the HA/LA manipulation such that larger PA differences emerge for Pepper than for Fetch.

For RQ3: (H3) Higher NARS scores will be associated with lower RoSAS Warmth, higher RoSAS Discomfort, and lower MDMT scores, replicating the pattern from prior work [2].

II. Method

A. Participants and Design

Fifty participants (34 male, 16 female; age: $M = 27.18$, $SD = 5.97$, range 19–44) took part in an in-person study at the University of Hertfordshire Robot House, a four-bedroom residential home adapted for HRI research [5]. The study used a mixed design: agency condition (HA vs. LA) was between-subjects (25 per condition), and robot type (Pepper vs. Fetch) was within-subjects, counterbalanced across participants.

B. Robots

Pepper (SoftBank Robotics) is a 1.2 m humanoid social robot with a chest-mounted touchscreen, expressive arms, and an articulated head supporting active gaze. **Fetch** (Fetch Robotics) is a mobile manipulator with a single articulated arm, gripper, RGB-D camera, and mobile base suited to physical object handling and spatial navigation.

C. Agency Manipulation

The manipulation operationalised five agency factors. Four — Goal Directness, Social Awareness, Accountability, and Shared Decision-Making — were differentiated through verbal and non-verbal behaviour between HA and LA conditions. The fifth, Embodiment, was operationalised within-subjects by exposing each participant to both robots. Table I defines each factor with contrasting HA and LA examples; Table II specifies the non-verbal behaviour differences.

The manipulation was delivered across three scenarios. **S1 (Security Check)** targeted Goal Directness, Social Awareness, and Accountability: participants provided a passcode for the robot to verify the house security system, using a mild deception paradigm (participants were fully debriefed). **S2 (Rock-Paper-Scissors)** targeted Shared Decision-Making, Embodiment, and Social Awareness: three rounds were played, with HA robots providing expressive round-result commentary and LA robots giving no feedback. **S3 (Table Cleanup)** targeted all five factors: participants assisted in disposing of a can, a building block, and bread, with Fetch physically handling items in HA while Pepper directed verbally.

D. Measures

Pre-interaction: the Negative Attitude Towards Robots Scale (NARS) [6] (14 items, 5-point) assessed pre-existing robot attitudes across three subscales: Interaction Anxiety, Social Influence Concerns, and Emotional Attitude.

Post-interaction (per robot): the Robot Social Attribute Scale (RoSAS) [7] (18 items, 7-point; Competence, Warmth, Discomfort); a custom five-dimension agency selection task and open-ended agency question; the Perceived Agency (PA) scale [1] (13 items, 5-point) as the primary manipulation check; and the MDMT [8] (16 items, 8-point; Reliability, Sincerity, Capability, Ethical Integrity).

Post-session: five forced-choice comparative questions identifying which robot demonstrated greater goal-directedness, shared decision-making, item-sorting suitability, social awareness, and responsibility; followed by age and gender.

E. Statistical Analysis Plan

Given ordinal data and expected non-normal distributions (confirmed via Shapiro-Wilk), non-parametric methods will be used throughout: Wilcoxon rank-sum tests for HA vs. LA between-condition comparisons; Wilcoxon signed-rank tests for Pepper vs. Fetch within-participant comparisons; two-sided binomial tests against a 50% baseline for forced-choice questions; and Spearman rank correlations for NARS associations with outcome measures.

III. Expected Contributions and Future Work

This study makes two anticipated contributions to the HRI community. First, by using a multi-channel behavioural manipulation grounded in the PA scale item hierarchy and prior agency literature, it provides the most systematic attempt to date to shift perceived agency through embedded robot behaviour — going beyond the text-based framing approach whose limitations were established in prior work [2].

Second, it is the first study to directly compare perceived agency and trust across two physically distinct robot platforms under matched behavioural conditions. Prior single-platform designs cannot determine whether agency attributions are driven by what the robot *does* (the manipulation) or by what the robot *is* (its embodiment). The within-subjects embodiment comparison isolates this contribution for the first time.

We hypothesise that the results will show embodiment to be a constitutive rather than merely a moderating factor: Fetch's physical task-handling affordances (arm, gripper, mobile base) should drive attributions of goal-directedness, capability, and accountability, while Pepper's humanoid social affordances (face, gaze, voice) should drive attributions of warmth and social awareness, independently of the HA/LA condition.

Future work will complement the questionnaire data with behavioural coding of the video recordings and thematic analysis of the open-ended responses collected across both studies, examining whether compliance, hesitation, and non-verbal engagement reveal effects that post-hoc questionnaires may not capture. Longitudinal replication and extension to further robot platforms are also planned.

References

[1] G. Trafton et al., "Perceived agency scale: Development and validation," in *Proc. ACM/IEEE HRI*, 2024.

[2] P. Holthaus, A. Fallahi, F. Förster, C. Menon, L. Wood, and G. Lakatos, "Agency effects on robot trust in different age groups," in *Proc. ACM HAI*, 2024.

[3] C. Roselli et al., "Vicarious sense of agency over a robot's actions," *Frontiers in Psychology*, 2022.

[4] R. B. Jackson and T. Williams, "Language-capable robots may inadvertently harm human social cohesion," in *Proc. ACM/IEEE HRI*, 2021.

[5] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would you trust a (faulty) robot?" in *Proc. ACM/IEEE HRI*, 2015, pp. 141–148.

[6] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "Prediction of human behavior in human-robot interaction using psychological scales for anxiety and negative attitudes toward robots," *IEEE Trans. Robotics*, vol. 24, no. 2, pp. 442–451, 2008.

[7] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (RoSAS): Development and validation," in *Proc. ACM/IEEE HRI*, 2017, pp. 254–262.

[8] D. Ullman and B. F. Malle, "MDMT: Multi-dimensional measure of trust," 2019.

[9] P. A. Hancock et al., "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, vol. 53, no. 5, pp. 517–527, 2011.

[10] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, "Overtrust of robots in emergency evacuation scenarios," in *Proc. ACM/IEEE HRI*, 2016, pp. 101–108.

TABLE I

THE FIVE AGENCY FACTORS OPERATIONALISED IN THIS EXPERIMENT. FOUR FACTORS (GOAL DIRECTNESS, SOCIAL AWARENESS, ACCOUNTABILITY, SHARED DECISION-MAKING) ARE MANIPULATED BETWEEN-SUBJECTS VIA HA/LA BEHAVIOURAL CONTRASTS. EMBODIMENT IS OPERATIONALISED WITHIN-SUBJECTS BY EXPOSING EACH PARTICIPANT TO TWO ROBOTS WITH DISTINCT PHYSICAL AFFORDANCES.

Factor	**Definition**	**High Agency (HA)**	**Low Agency (LA)**
Goal Directness / Intentionality	Actions perceived as purposeful and directed toward self-chosen goals rather than mechanical or reactive.	States purpose before acting; autonomously identifies problems (e.g., “To check all the sensors, I will need the master password”).	Bare requests with no rationale (e.g., “I need a five-digit password”).
Social Awareness	Capacity to perceive, interpret, and respond to social cues adaptively and appropriately.	Expressive, personalised responses (e.g., “That was fun! You won this time!”); adaptive recovery when input is missed.	Functional, neutral language; no emotional responsiveness or round-result feedback.
Accountability / Responsibility	Capacity to justify or take responsibility for actions, promoting transparency and trust.	Explicit ethical reasoning (e.g., “I will terminate the password for safety reasons. It is not ethical to keep it active”).	Action stated without explanation (e.g., “Ok, the password is no longer needed”).
Shared Decision-Making	Collaborative process in which humans and robots jointly participate in decisions.	Invites participant evaluation (e.g., “Do you agree the recycling bin is not ok for the bread?”).	Directive only; no consultation (e.g., “Please throw the bread in a proper bin”).
Embodiment	Physical and perceptual presence of a robot — its body form, movement, and responsiveness — shaping how humans experience its agency.	Operationalised within-subjects: **Fetch** physically receives objects from the participant and places them in the correct bin in S3; **Pepper** visually displays its game choice on its chest screen in S2. The Embodiment factor varies by robot type rather than by HA/LA condition.	

TABLE II

NON-VERBAL BEHAVIOUR DIFFERENCES BETWEEN HA AND LA CONDITIONS. ALL BEHAVIOURS APPLY TO BOTH ROBOTS UNLESS NOTED.

Factor	**High Agency (HA)**	**Low Agency (LA)**
Gaze & Head Orientation	Actively orients toward participant; shifts gaze to relevant objects before initiating dialogue.	Neutral or fixed orientation; no proactive gaze shifts.
Approach & Spatial Behaviour	Moves toward participant at scenario start; leads physical navigation (e.g., moves to kitchen first in S3).	Remains stationary at scenario start; follows participant rather than leading.
Deliberation & Timing	1–2 s pause before key decisions; motion timing varied across executions.	Immediate action after input; deterministic timing; identical motion sequences.
Inter-scenario Behaviour	Performs autonomous purposeful tasks between scenarios (S1.5: temperature check; S2.5: network monitoring).	Remains idle between scenarios; no self-initiated behaviour.
Object Handling & Leading (S3)	Leads participant to kitchen; Fetch places items independently in the correct bin; Pepper directs with specific bin knowledge.	Follows participant to kitchen; Fetch asks participant to show the bin; Pepper asks participant to use the proper bin without specification.