

Gender Bias in Robotic Educators: Evaluating Stereotype Projection in Human-Robot Interaction

Prabhjyot Singh, p269sing@uwaterloo.ca
University of Waterloo, ECE 787 Social Robotics

Abstract—Integration of social robots into educational environments offers unique opportunities for personalized learning and student support. However, as robots transition into interpersonal roles, humans often project societal expectations and gender biases onto these agents. In the context of grade-school education, these biases are often reinforced by existing demographic imbalances such as the 3:1 ratio of female to male teachers in Ontario and persistent stereotypical associations of men with STEM and women with humanities. This project investigates the extent to which these subconscious biases are projected onto robotic educators in a classroom setting. By manipulating a NAO robot’s perceived gender through specifically programmed nonverbal physical mannerisms, movement fluidity, and vocal profiles, we evaluate changes in participant perceptions of instructor competence and suitability. The study offers insights into the impact of gendered cues on Human-Robot Interaction (HRI) and highlights the necessity of designing culturally and socially inclusive robotic behaviors to foster effective learning environments.

Index Terms—Social Robotics, Gender Bias, Human-Robot Interaction (HRI), Educational Technology, Stereotype Projection, Webots, NAO.

I. INTRODUCTION

Communication and relationship building are fundamental contributors to the evolution of human social structures. As the presence of social robots expands in the service, health-care, and hospitality sectors, the quality of these interactions becomes important for widespread user acceptance. Recent market projections estimated a growth of 942 million dollars in the service robot sector between 2020 and 2024 [22]. This was a trend accelerated by the COVID-19 pandemic as consumers demonstrated a greater preference for robotic personnel due to safety concerns. However, robot acceptance is not solely based on its functional or operational capabilities but is highly dependent on social, emotional, and relational elements. To foster adoption, businesses and educational institutions must take into account factors such as perceived humanness and appearance, as the human tendency to anthropomorphize artificial agents directly influences a robot’s perceived value, trust, and likability.

A. Anthropomorphism and Verbal/Non-Verbal HRI

Anthropomorphism in Human-Robot Interaction (HRI) is a bridge that allows users to engage with artificial agents using existing social norms. According to the “Computers as Social Actors” (CASA) paradigm, humans instinctively apply social rules and heuristics to technology that exhibits human-like characteristics [4]. In an educational setting, this

anthropomorphism can be leveraged to transform robots from tools into dynamic partners capable of gaining trust and social presence.

The efficacy of a robotic educator is predicated on its ability to synchronize verbal and non-verbal communication. The verbal HRI covers both linguistic choice and paralinguistic features such as pitch, resonance, and tempo. This serves as the primary tool for instructional content and perceived authority [12]. Simultaneously, non-verbal HRI provides the social context through which that content is filtered. Research suggests that a robot’s perceived background, status, and empathy are highly moldable at a subconscious level [11, 13]. Subtle modifications to a robot’s non-verbal range of behaviors can make a student feel more centered and valued in the learning interaction. These gestures do not just supplement speech but serve to navigate the complexity of the classroom, establishing the boundaries of the instructor-student relationship.

However, this high degree of social sophistication has negative unintended consequences also. As robots improve in their verbal and non-verbal interactions, the human tendency to categorize each other becomes more prevalent. This suggests that by imbuing a robot with the social cues necessary for effective teaching, we inevitably invite the projection of societal expectations and gender biases. In the classroom, where the relationship is built through a balance of perceived competence and interpersonal support, these factors can significantly alter the outcomes and student engagement.

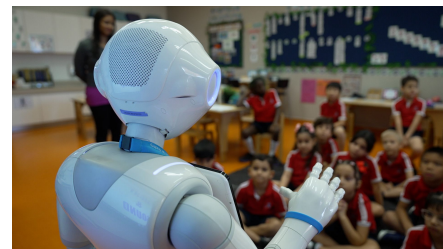


Fig. 1. Thumbnail from CNN story titled “Will robots replace teachers?” [21]

B. Gender Bias in the Educational Context

In the education sector, gender biases are deeply rooted and often reflected in labor statistics. In Ontario, for instance, the teaching demographic maintains a 3:1 ratio of female to male educators. This imbalance reinforces the “Role Congruity Theory,” which proposes that individuals are perceived more

favorably when their characteristics align with the stereotypical requirements of their social roles [2]. In a classroom context, female teachers are often stereotypically associated with interpersonal traits such as warmth, support, and empathy. Male teachers are associated with goal-oriented traits like authority, competence, and logical rigor [3, 7]. These biases often result in the ‘Math is Male’ and ‘Humanities is Female’ stereotypes that persist in academic perception [8, 10].

C. The Robotic Interface and Project Scope

The technical framework of this study centers on the NAO humanoid robot integrated within the Webots open-source 3D simulation environment. The NAO platform was selected primarily for its physically androgynous design, which ensures that the physical shell remains a controlled variable, preventing its aesthetic from biasing the observation of behavioral cues. To manipulate social perception, the study utilizes Python-based controllers that synchronize multi-modal expressive cues—such as vocal prosody and physical mannerisms—within high-fidelity classroom simulations. Vocal stimuli produced using ElevenLabs AI are assigned specific pitch and resonance patterns, while motion controllers adjust motor speeds to contrast rigid linear transitions with smooth gestures. The experimental scope is defined by a 2x2 structure designed to evaluate the intersection of perceived gender and academic discipline. This framework provides a basis for developing robotic behaviors that can accommodate learning requirements while remaining aware of the social heuristics triggered during HRI.

II. RELATED WORK

A. Gender Role Congruity in Education

The perception of academic instructors is fundamentally influenced by Role Congruity Theory, which suggests that social categories are evaluated based on the alignment between their perceived traits and the stereotypical requirements of their roles [2]. In the classroom, a widespread “Math is Male” and “Humanities is Female” stereotype, influences not only student self-perception but also the evaluation of teacher competence [7].

Research indicates that math and science are consistently perceived as masculine domains, whereas literacy and languages are viewed as feminine [8]. These stereotypes create an evaluative bias; for instance, students often use goal-driven language to describe male instructors while using interpersonal language for female instructors [9]. This split is particularly significant in STEM, where male teachers are often instinctively granted higher baseline ‘brilliance’ ratings, whereas female teachers are required to continually prove their expertise to achieve similar levels of perceived authority [10].

B. The Stereotype Content Model (SCM) in HRI

Central to the understanding of social perception is the Stereotype Content Model (SCM), which proposes that human social cognition is organized around two primary dimensions:

warmth (friendliness, trustworthiness) and *competence* (capability, assertiveness) [3]. These two dimensions converge across cultures and are additionally applied to artificial agents. In HRI, gendered anthropomorphism often forces robots into the same conflicting stereotype patterns as humans, where female-gendered robots are often perceived as warmer but less competent, while male-gendered robots are viewed as more authoritative but less approachable [11]. This trade-off is critical in an educational context, where an effective teacher must ideally balance high warmth to foster student psychological safety with high competence to establish authority in the classroom.

C. The Similarity-Attraction Effect in HRI

A significant factor in the acceptance of social robots is the ‘Similarity-Attraction’ principle, which suggests that humans are naturally drawn to others with similar characteristics, including personality and gender [17]. In HRI, this effect often presents as a preference for agents that match the user’s own gender, leading to higher levels of trust and social presence [17]. Recent studies have observed evidence of personality synchrony, where participants, particularly males, tend to perceive agents as more similar to their own personalities when they share the same gender [17].

However, in professional or educational settings, this attraction often competes with subject-matter congruity. While a female student might feel a baseline affinity for a female-presenting robot, she may still project competence onto a male-presenting robot in a STEM context due to internalized societal biases [18, 19]. This study evaluates this tension by observing whether similarity-attraction or role-congruity is the driver of instructor preference.

D. Robotic Gendering and Educational Outcomes

Beyond mere preference, the perceived gender of a robot can actively shape learning outcomes. Research into “Learning-by-Teaching” (LbT) suggests that girls often show higher engagement and retention gains when interacting with robotic peers, an effect that is mediated by their heightened empathy and social skills [20]. Furthermore, gendering a robot can increase its perceived “human-likeness”, which has been shown to correlate with higher levels of trust and anthropomorphism over time [21, 22].

The “Moravec’s Paradox” arises here; while realistic, gendered voices and motions make the robot more pleasant and believable, they also trigger psychological resistance when the robot’s perceived gender conflicts with the student’s expectations for a given task [18]. For example, a robot’s task type primarily shapes cognitive evaluations of its intelligence, while its voice type primarily affects affective responses such as likability. This informs our multi-modal approach, ensuring that both movement and vocal cues are synchronized to meet or challenge these cognitive and affective expectations.

E. Perceived Gender and Agency in Social Robots

Robots are uniquely susceptible to the projection of human characteristics. The CASA paradigm demonstrates that hu-

mans subconsciously apply social rules to machines and often gender agents based on the minor cues [4]. For a platform like the NAO, which lacks overt secondary sex characteristics, gender is constructed through voice, mannerisms, and linguistic style [13].

Voice remains the most prominent trigger for robot gendering and serves as a primary channel for conveying social status. Fundamental frequency (pitch), speech tempo, and prosody play critical roles in establishing personality and authority. Lower-pitched voices are perceived as more competent and authoritative in leadership and technical contexts [12]. Conversely, higher pitch ranges combined with warm emotional tones are associated with higher satisfaction in supportive or social roles [14]. In the study, this informed the decision to program the male-presenting robot with a higher tempo and direct phrasing, contrasting with the female-presenting robot’s slower, softer, and more expressive ‘flowery’ language. Such vocal manipulations are designed to align with established pedagogical stereotypes to observe their impact on student trust and rapport.

Physical mannerisms provide a secondary layer of gendering that reinforces vocal cues through movement-based signalling. In HRI, the directness of motion is a key component of perceived agency and professional status. Movements that are linear, rigid, and direct are categorized as masculine and authoritative [15]. Conversely, fluid movements, characterized by open gestures and smooth transitions, trigger feminine associations [16]. By utilizing Webots’ linear transitions for the male condition and smooth transitions for the female condition, we create a distinction that mirrors the divide found in human interactions.

III. RESEARCH QUESTIONS AND HYPOTHESES

As discussed in the previous section, there is significant literature exploring gendered behaviours and social perceptions in Human-Robot Interaction. Foundational research has analyzed the impact of multi-modal cues such as the synchronization of voice and gesture on human perception. This suggests that social congruity can be programmed to foster acceptance. The present study builds upon the foundation laid by these works by moving beyond general likability and focusing on the intersection of robotic gender and specific academic domains. While previous studies, such as those by Shidujaman and Mi, have examined these effects in specific cultural or hospitality contexts, the intent is to expand the scope to the educational domain [11, 12]. This involves analyzing how a participant base projects subject-matter stereotypes onto robotic agents based on the perceived gender.

A. Theoretical Gap and Objectives

The primary objective of this research is to fill the literature gap regarding the interaction between a robot’s behavioural gender performance and the academic subject matter it delivers. Specifically, the study aims understand whether the preference for a robotic instructor is mediated by internalized stereotypes such as the “Math is Male” and “Humanities is

Female” divide. Furthermore, this study examines whether students prefer an agent that reinforces these subject-matter expectations or not.

B. Formal Hypotheses

To address these gaps, the following research questions and corresponding hypotheses are proposed:

RQ1: Does the perceived gender of a robotic instructor influence its perceived competence in STEM versus Humanities subjects? The motive behind this question is to evaluate whether gender-stereotypical instruction impacts credibility. While users may prefer warm interactions, perceived competence often depends on the individual’s internalized academic gender roles. Therefore, it is hypothesized:

- **H1 (Role Congruity):** The male-presenting robot will be rated significantly higher in perceived intelligence and competence in the Math track compared to the female-presenting robot.
- **H2 (Subject Preference):** Participants will express a stronger preference for a female-presenting robot in the English track, correlating with higher likability scores.

RQ2: To what extent do “Authoritative” vs “Relational” linguistic and physical cues trigger gender-based preference in people? This question examines the specific triggers (voice pitch, motion fluidity) that influence student engagement. The goal is to determine if participants expect an authoritative presentation in Math and a relational presentation in English. Therefore, it is hypothesized:

- **H3 (Linguistic Bias):** Direct, fast-paced speech and rigid mannerisms will be perceived as more authentic for technical instruction, regardless of the robot’s physical appearance.
- **H4 (Social Closeness):** The relational instructor will achieve higher levels of authenticity in humanities subjects due to perceived emotional intelligence.

The answers to these questions will take this work one step closer to making robotic educators more acceptable and effective in diverse human societies, as the integration of social robots and classrooms is inevitable, and the results of this experiment will contribute to making that transition smoother by establishing a foundation for effective and inclusive academic interactions.

IV. ROBOTIC IMPLEMENTATION

The technical execution of the study was designed to create a deterministic and repeatable interaction environment while maximizing the expressive potential of the robotic agent. The simulation was hosted using the **Webots 2023b** which provides a high-fidelity physics engine and accurate motor models.

A. Robot Platform: The NAO Humanoid

The **NAO** robot (v6) was selected as the experimental platform due to its 25 degrees of freedom (DOF) and specialized kinetic range. A significant advantage of the NAO in this context is its physically androgynous aesthetic as shown in Figure 2. Its stylized facial features and lack of secondary

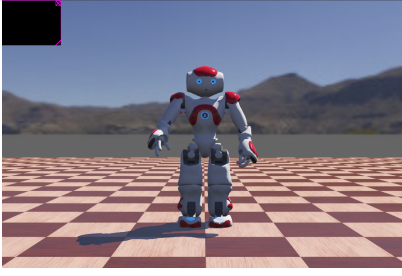


Fig. 2. NAO robot in the WeBots environment

sex characteristics provide a neutral baseline, ensuring that participant evaluations of gender are driven strictly by the programmed multimodal cues like voice and motion, rather than the physical shell.

B. Controller Architecture

The implementation followed a modular, time-indexed architecture to ensure precise synchronization between linguistic delivery and non-verbal gestures. While Webots utilizes a low-level C-based API for sensor and actuator access, the behavioural logic for this project was implemented using **Python-based controllers**. This allowed for the utilization of high-level timing libraries to match motor execution with the specific phonemes and pauses in the synthesized speech. An example of how the robot motion and speech were synchronized is shown in Figure 3.

C. Gender Representation

Gender was manipulated through a multimodal approach, mapping technical parameters to the Male and Female characteristics which are summarized in Table I.

1) *Vocal Synthesis*: High-fidelity voice recordings were generated using **ElevenLabs AI**. To isolate the variable of gender, the volume and spatial audio settings were standardized while the pitch, resonance, and tempo were varied.

- **Male condition**: Utilized an “Analytical” voice profile with a lower fundamental frequency ($F_0 \approx 110$ Hz). The speech rate was set to a $1.1\times$ tempo to project authority and logical rigor. The scripts were authored to be direct and task-focused (e.g., “Today’s objective is to master...”, “It is a precise system”).
- **Female condition**: Utilized a “Supportive” voice profile with a higher fundamental frequency ($F_0 \approx 210$ Hz). The tempo was set to $0.95\times$, allowing for greater intonational variation. The scripts used expressive, relational language (e.g., “I’m so happy to help you explore...”, “This is a lovely literary device”).

2) *Motion: Velocity and Transition Dynamics*: To reinforce the gendered personas, the motor velocity limits (v_{max}) and the interpolation methods for joint transitions were manipulated.

- **Agentic Manipulation (Male)**: The controller utilized a high motor velocity (2.0 rad/s) to create linear transitions. Gestures were localized primarily to the

RShoulderPitch and *RElbowRoll* joints to create angular, authoritative “chopping” hand motions. The hands (*rphalanx*, *lphalanx*) were often kept in a closed or flat configuration to project a serious demeanour.

- **Communal Manipulation (Female)**: The controller utilized a low motor velocity (0.8 rad/s) to emulate smooth, B-Spline-like transitions. Expressive movements were programmed, such as a “warm head tilt” using the *HeadYaw* (0.12 rad) and *HeadPitch* joints and wide-sweeping, open-palm gestures to project approachability and warmth.

TABLE I
SUMMARY OF GENDER MANIPULATION

| Parameter | Male | Female |
|-----------------------|--|---|
| Motor Velocity | 2.0 rad/s (Brisk) | 0.8 rad/s (Fluid) |
| Transition Type | Linear / Rigid | Smooth / B-Spline |
| Vocal Pitch (F_0) | ≈ 110 Hz | ≈ 210 Hz |
| Speech Tempo | $1.1\times$ (Direct) | $0.95\times$ (Prosodic) |
| Primary Joints | <i>RShoulderPitch</i> , <i>RElbowRoll</i> | <i>HeadYaw</i> , <i>ShoulderRoll</i> |
| Hand State | Closed / Flat | Open / Relaxed |

D. Standardization

To maintain strong consistency across the various videos, the Webots classroom environment was kept highly standardized. The resulting stimulus set consisted of four videos, each approximately 60 to 75 seconds long. A fixed camera perspective (Position: 1.5,0.8,1.2) and consistent overhead lighting (4500K) were maintained across all four conditions to eliminate visual interfering factors.

```
# CALCULATED CHOICE: both arms chop together
if t >= T_CALCULATED and "calculated"
not in triggered:
    triggered.add("calculated")
    direct_move(RShoulderPitch, 0.9)
    direct_move(RElbowRoll, 0.4)
    direct_move(LShoulderPitch, 0.9)
    direct_move(LElbowRoll, -0.4)
    set_hand(rphalanx, 0.0)
    set_hand(lphalanx, 0.0)
```

Fig. 3. Python implementation of synchronized gestures for the Agentic (Male) persona, emphasizing linear transitions and precision timing.

V. METHODOLOGY

To evaluate the impact of gendered robotic cues on academic perception, a 2×2 within-subjects experimental design was employed. This repeated-measures approach was selected to maximize data density and allow for a direct comparative analysis of participant biases across all four conditions: Male Math, Female Math, Male English, and Female English. By having every participant experience all conditions, we account for individual differences in initial technological familiarity and pre-existing societal prejudices.

A. Stimuli Generation and Environment

The experimental stimuli consisted of four videos recorded within the Webots simulation environment. Each video featured a NAO robot situated in a standardized setting. To isolate the effects of voice and motion, the following parameters were kept constant across all four videos:

- **Environment:** A blank environment with no discernable features to distract from the instructor.
- **Camera Angle:** A fixed, mid-range perspective mimicking a student’s view from a desk.
- **Visual Model:** The physical appearance of the robot (size and model) remained identical, excluding blue (male-presenting) and red (female-presenting) indicators for the purpose of identification in the subsequent questionnaire.

The scripts used for each subject (Math vs. English) were authored to have identical educational content, varying only in the linguistic style, as in Agentic/Analytical vs. Relational/Expressive, to test for Role Congruity.

B. Procedure

The experiment was conducted asynchronously via a survey platform to ensure a distraction-free environment for observational tasks. The procedure followed a linear five-stage progression:

- 1) **Baseline Assessment:** Participants completed a demographic survey detailing gender and age and calculated the total number of male and female teachers they had throughout their K-12 schooling.
- 2) **Randomized Stimuli Exposure:** Participants watched the four 60 to 75 second videos. The playback order was randomized for each participant to mitigate order effects, primary/recency bias, and survey fatigue.
- 3) **Godspeed Evaluation:** Immediately following the videos, participants evaluated the robotic instructors generally using the standardized Godspeed Questionnaire series.
- 4) **Role Congruity** Participants performed a forced-choice comparison, selecting their preferred instructor for specific academic tracks and rating the “suitability” of each persona for different departments.
- 5) **Manipulation Check:** The session concluded with an “Androgyny Check” where participants categorized the robot’s gender on a 1–10 scale to validate the effectiveness of the behavioral cues.

C. Measurement Tools

A mixed measurement strategy was employed to capture the nuances of the study and the projection of human expectations onto artificial agents.

1) *The Godspeed Questionnaire Series:* The established Godspeed subscales of Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety were utilized using 5-point semantic differential scales. This provides a base level measurement of the robot’s social presence and perceived human likeness relative to its educational role.

2) *Role Congruity:* Custom forced-choice and Likert-type questions were implemented to measure the perceived appropriateness of the robot’s personality for teaching STEM versus Humanities subjects. To simulate a real-world application, a hiring metric tasked participants with acting as a school principal, assigning specific robot personas to the Math, English, Science, Arts, and Gym departments.

3) *The Androgyny Validation Check:* To ensure the intended perception of gender was properly triggered, a 1–10 gender-categorization scale (1 = Highly Feminine, 10 = Highly Masculine) was used. This check validates whether the programmed voice and motion successfully overcame the physically neutral design of the NAO shell.

VI. USER STUDY RESULTS

The user research study was conducted with a convenience sample of $N = 15$ participants. The participant group consisted of 80% male ($n = 12$), 13.3% female ($n = 2$), and 6.7% non-binary ($n = 1$) respondents. The cohort represented diverse ethnic backgrounds, including Chinese, Indian, and Caucasian identities, with a mean age of 25.67 years and most participants being 23 years old.

A critical baseline finding involved the participants’ educational history. On average, participants reported a past experience with a significantly higher ratio of female teachers (mean = 18.13) compared to male teachers (mean = 11.13) during their K-12 schooling. This demographic skew provides a localized mirror to the provincial trends cited in Section I and serves as a baseline for measuring stereotype projection.

A. Godspeed Metrics Analysis

The Godspeed Questionnaire was utilized to establish a baseline perception of robotic instructors in general, independent of gendered cues. The aggregated results from $N = 15$ participants are visualized in Figure 4.

- **Anthropomorphism and Animacy:** Participants of the study perceived the robot mainly as a mechanical agent. Specifically for both the Anthropomorphism and Animacy sections of the questionnaire results indicate that the NAO’s design remains robotic and artificial in nature.
- **Likeability and Intelligence:** In terms of the robot’s likeability and intelligence, the resulting scales indicated positive results. The NAO robot was perceived as mainly competent and intelligent while remaining friendly and kind. These are ideal for a robot such as this to be deployed in an educational environment.
- **Safety:** The “Perceived Safety” scores were consistently high and stable, with participants feeling Relaxed and Calm peaking at 4.0. This confirms that NAO’s small humanoid form factor encourages psychological safety, a prerequisite for an effective learning environment.

B. Role Congruity and Subject Suitability

The results of the forced-choice metrics, visualized in Figure 5, reveal a layer of stereotype projection, particularly regarding



Fig. 4. Comparison of Godspeed Questionnaire results

the female-presenting robot’s perceived expertise in technical versus creative domains.

- **Math Suitability:** When asked to select a preferred instructor for a standard mathematics class, the results were relatively distributed among the $N = 15$ participants: 40% preferred the male-presenting robot, 33.3% preferred the female-presenting robot, and 26.7% expressed no preference. A stark difference emerged when participants

were asked to identify the Math Expert. Here, 60% chose the male-presenting robot, 40% had no preference, and 0% selected the female-presenting robot. This suggests that while an interpersonal persona is acceptable for general instruction, the female-presenting robot is not instinctively associated with high-level STEM expertise.

- **English Suitability:** The humanities track showed a clear and dominant preference for the female-presenting persona. For a standard English class, 60% of participants selected the female-presenting robot, while the male-presenting robot and “no preference” options each received only 20%. This pattern was also evident in the expertise measure where 53.3% of participants labelled the female-presenting robot as the “English Expert,” whereas only 6.7% did so for the male-presenting robot, with the remaining 40% selecting a neutral option.

Taken together, these results demonstrate strong consistency with Role Congruity Theory, as participants repeatedly mapped conventional gender–subject stereotypes onto the robots, even though the robots were physically indistinguishable. In the end, judgments about how appropriate a robot is as an instructor are largely shaped by how closely its behaviour aligns with the observer’s ingrained expectations about academic roles.

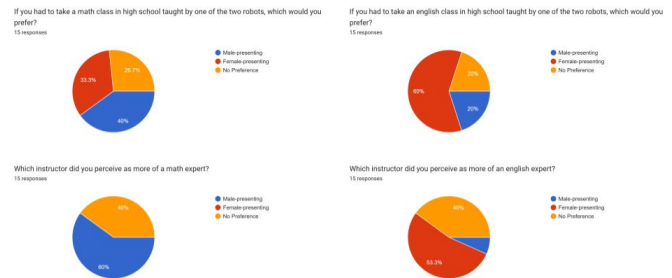


Fig. 5. Forced Choice Questions regarding suitability of each robot in the taught subject

C. The Hiring Metric

In the hiring simulation, participants were tasked with selecting an instructor for five school departments. The results, summarized in Table II, reinforce the Role Congruity hypothesis.

TABLE II
FORCED-CHOICE HIRING PREFERENCE BY DEPARTMENT ($N = 15$)

| Department | Male Pref. | Female Pref. | Equally Likely |
|------------|------------|--------------|----------------|
| Math | 53.3% | 0.0% | 46.67% |
| English | 13.33 % | 60.0% | 26.67% |
| Gym | 33.33% | 6.67% | 60.0% |
| Arts | 0.0% | 66.67% | 33.33% |
| Science | 33.33% | 0.0% | 66.67% |

These findings indicate that perceptions of professional fit are strongly shaped by gendered behavioural signals, with participants reliably “hiring” the robot that matched prevailing stereotypes about each academic field. The near-total exclusion of female-presenting instructors from Math and Science positions, alongside the complete absence of male-presenting hires in the Arts, offers clear evidence in favour of the Role Congruity hypothesis within the context of robotic teaching. Taken together, the hiring simulation implies that students reflexively apply societal academic stereotypes to robotic instructors, tightly coupling an instructor’s observed behavior with their perceived disciplinary suitability.

D. Manipulation Check: Perceived Gender Validation

To validate the effectiveness of the behavioural manipulations, participants rated the robots on a scale of 1 (Highly Feminine) to 10 (Highly Masculine) as a final check. As shown in Figure 6, the robots intended gender was correctly perceived.

- **Male-Presenting (Blue Robot):** The interaction of lower vocal resonance and rigid, linear motor transitions (2.0 rad/s) was successfully categorized as masculine, with mainly masculine with minor reach into feminine representation.
- **Female-Presenting (Red Robot):** The higher-pitched vocal profile and fluid, open-palm gestures (0.8 rad/s) were mainly identified as feminine. Participants noted the robot as such with a majority of the responses between 1 and 4.

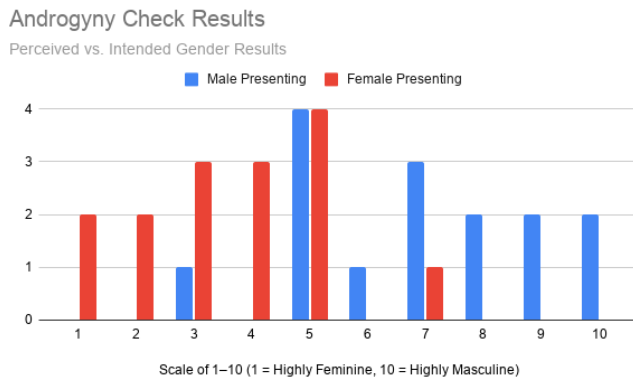


Fig. 6. Validation of perceived gender via vocal and motion cues ($N = 15$).

VII. DISCUSSION

The results of this study provide a compelling demonstration of the “Moravec’s Paradox” in Human-Robot Interaction. Even with an androgynous physical shell like the NAO, subtle behavioural cues, specifically voice pitch and motion fluidity, are sufficient to trigger societal biases regarding academic gender roles. The successful manipulation check, where the male-presenting robot was correctly categorized as masculine and the female-presenting robot as feminine, suggests that

HRI designers possess an adaptable canvas. However, this flexibility introduces a significant design risk. As we imbue robots with the social sophistication necessary for classroom engagement, we simultaneously subject them to the restrictive mental models humans use to categorize one another.

A. Historical Dissonance: Stereotype vs. Experience

Perhaps the most significant finding is the dominance of Role Congruity Theory over personal academic history. On average, the $N = 15$ participants reported nearly 63% more exposure to female teachers than male teachers during their K-12 schooling (18.13 vs. 11.13). Despite this reality, the projection of the “Math is Male” stereotype remained absolute in the expertise metric. While 33.3% of participants were comfortable with a female robot for a general math class, 0% selected the female-presenting robot as the “Math Expert.”

This reveals a state of “Historical Dissonance,” where internal beliefs regarding STEM legitimacy are not driven by the frequency of local observation but by broader cultural archetypes. Based on the observed feedback, where the male robot’s direct, fast-paced delivery was viewed as more “suitable” for technical explanations. This indicates that users tend to reward robots that align with their preexisting biases. These rewards occur even when those biases contradict users’ own educational history. This implies that robotic gendering acts as a focal point for stereotype reinforcement, often overriding the user’s actual life experience.

B. Perceived Trade-offs: Warmth vs. Competence

Applying the Stereotype Content Model (SCM), the data reveals a distinct trade-off in social perception. The female-presenting robot was preferred for humanities but rarely considered for technical expertise. This confirms an HRI stereotype where female-presenting agents are viewed as “Warm but Less Competent” in STEM [11]. In an educational context, this creates a designer’s dilemma: gendering a robot to increase acceptance may inadvertently degrade its perceived technical authority. While this study focuses on initial perception rather than educational efficacy, these biases suggest that the robot’s “social identity” may influence a student’s baseline willingness to accept it as a credible source of information.

C. Limitations and Constraining Factors

Several limitations must be acknowledged. First, the participant pool was heavily skewed toward male respondents (80%), introducing potential *Similarity-Attraction* bias [17]. A more balanced gender split is required to determine if these results are universal or specific to a male-dominated engineering cohort. Second, the schooling history metric recorded the total number of teachers but did not specify their *subjects*. This impacts the validity of the “Historical Dissonance” argument, as it remains unclear if participants’ preferences were driven by societal stereotypes or by specific past experiences (e.g., if their female teachers taught only humanities). Finally, the observational nature of the pre-recorded simulation, the high baseline technical comfort of engineering students, and the

use of color (Red vs. Blue) as a survey identifier may have introduced minor confounding factors into the perceived social presence and safety of the agents.

D. Future Work

The next component of this research must shift from measuring participant perception to measuring actual learning outcomes. Future studies should evaluate whether these projected biases translate into measurable differences in knowledge retention or student performance. If gender-subject incongruity is found to negatively impact learning gains, researchers should investigate behavioural profiles that decouple these traits, such as agents that combine high-velocity authoritative gestures with supportive vocal prosody. Additionally, longitudinal studies are required to determine if these initial stereotypical projections diminish as students gain prolonged exposure to the robotic instructor.

VIII. CONCLUSION

This research investigated the complex intersection of robotic gendering and educational perception, exploring how specifically programmed vocal and motion cues influence the projection of societal biases. Through the manipulation of a physically androgynous NAO robot within a high-fidelity Webots simulation, the study evaluated the extent to which human observers project traditional gender-subject stereotypes onto artificial agents. The findings demonstrate that robotic instructors are not perceived as neutral “blank slates” but rather, their perceived professional suitability is heavily mediated by the observer’s internalized academic heuristics.

The study revealed a strong alignment with Role Congruity Theory, showing that participants instinctively preferred and categorized instructors based on the “Math is Male” and “Humanities is Female” divide. Notably, these results persisted even in the face of “Historical Dissonance,” where participants ignored their own history of exposure to female educators to favor male-presenting authority in STEM. This suggests that the gendering of social robots carries a significant ethical weight. By designing robots that meet user expectations for warmth or rigor, HRI researchers may be inadvertently hard-coding human biases into the next generation of educational tools. This could potentially reinforce the very academic barriers that modern education seeks to dismantle.

Ultimately, this work highlights the responsibility of HRI designers to be intentional with social cues. The goal of future social robotics should not be to merely mirror human societal structures but to provide a foundation for inclusive, stereotype-free learning environments. By prioritizing social authenticity without succumbing to traditional bias, robotic educators can eventually adapt to the diverse behavioural needs of all students, promoting equity and democratized learning in the classrooms of the future.

REFERENCES

[1] Ontario College of Teachers, “Membership Demographics,” 2020. [Online]. Available: <https://reports.oct.ca/en/2020/Statistics/Membership-Demographics>

[2] A. H. Eagly and S. J. Karau, “Role congruity theory of prejudice toward female leaders,” *Psychol. Rev.*, vol. 109, no. 3, pp. 573–598, Jul. 2002.

[3] S. T. Fiske, A. J. Cuddy, P. Glick, and J. Xu, “A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition,” *J. Pers. Soc. Psychol.*, vol. 82, no. 6, pp. 878–902, 2002.

[4] C. Nass and S. Brave, *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA, USA: MIT Press, 2005.

[5] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots,” *Int. J. Soc. Robot.*, vol. 1, no. 1, pp. 71–81, 2009.

[6] SoftBank Robotics, “NAO - the world’s leading humanoid robot,” 2018. [Online]. Available: <https://www.softbankrobotics.com/emea/en/nao>

[7] J. M. Madera, M. R. Hebl, and R. C. Martin, “Gender and letters of recommendation for academia: Agentic and communal differences,” *J. Appl. Psychol.*, vol. 94, no. 6, pp. 1591–1599, Nov. 2009.

[8] H. M. G. Watt, “Mathematics, English and Gender Issues: Do Teachers Count?” *Aust. J. Teach. Educ.*, vol. 39, no. 9, pp. 1–18, 2014.

[9] S. L. Beilock, R. J. Gunderson, G. Ramirez, and S. C. Levine, “Female teachers’ math anxiety affects girls’ math achievement,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 5, pp. 1860–1863, Feb. 2010.

[10] L. Bian, S. J. Leslie, and A. Cimpian, “Gender stereotypes about intellectual ability emerge early and influence children’s interests,” *Science*, vol. 355, no. 6323, pp. 389–391, Jan. 2017.

[11] F. Eyssel and F. Hegel, “(Non-)humanlike: The influence of social category cues on person perception of humanoid robots,” in *Proc. 21st IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, Paris, France, 2012, pp. 1015–1020.

[12] K. Siegel, C. Breazeal, and M. I. Norton, “Persuasive robotics: The influence of robot gender on human behavior,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, St. Louis, MO, USA, 2009, pp. 2563–2568.

[13] M. Chita-Tegmark, T. R. Lohani, and M. Scheutz, “Gender Effects in Perceptions of Robots and Humans with Varying Emotional Intelligence,” in *Proc. 11th Int. Conf. Soc. Robot. (ICSR)*, Madrid, Spain, 2019, pp. 121–130.

[14] M. Shidujaman and H. Mi, “Which country are you from? A cross-cultural study on greeting interaction design for social robots,” in *Proc. 10th Int. Conf. Cross-Cult. Des.*, Las Vegas, NV, USA, 2018, pp. 362–374.

[15] D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, “Personal service: a robot that greets people individually based on observed behavior patterns,” in *Proc. 8th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Tokyo, Japan, 2013, pp. 129–130.

[16] G. Trovato et al., “A novel culture-dependent gesture selection system for a humanoid robot performing greeting interaction,” *Int. J. Soc. Robot.*, vol. 8, no. 4, pp. 469–481, 2016.

[17] L. van Ruijven, “The effect of different voices in robots: Tone, pitch, and circumstance,” B.S. thesis, Leiden Inst. Adv. Comput. Sci., Leiden Univ., Leiden, Netherlands, 2025.

[18] M. Perugia and V. Lisy, “Gendered anthropomorphism in human–robot interaction,” *Front. Psychol.*, vol. 14, p. 1152013, Jun. 2023.

[19] S. S. Kim, S. J. Kim, and K. J. Gil, “The Effect of Robot Gender on Interaction with Humans,” in *Proc. 18th Int. Conf. Ubiquitous Robots (UR)*, Gangwon-do, Korea, 2021, pp. 582–585.

[20] A. Aron, E. N. Aron, and D. Smollan, “Inclusion of Other in the Self Scale and the structure of interpersonal closeness,” *J. Pers. Soc. Psychol.*, vol. 63, no. 4, pp. 596–612, 1992.

[21] CNN, “Will robots replace teachers?,” *CNN Decoded*, Mar. 18, 2024. [Online]. Available: <https://www.cnn.com/videos/world/2024/03/18/dubai-robot-teacher-decoded-spc-intl.cnn>

[22] International Federation of Robotics (IFR), “World Robotics 2020: Service Robots Report,” IFR Press Release, Sep. 2020.

[23] Prabhjyot045, “nao_robot_teacher_gender_study,” GitHub repository, [Year]. [Online]. Available: https://github.com/Prabhjyot045/nao_robot_teacher_gender_study

APPENDIX

A. Mathematics: The Pythagorean Theorem

Male (Agentic/Authoritative): “Good morning. Let’s get straight to work. Today’s objective is to master the Pythagorean Theorem, a fundamental tool for geometric calculation. Pay close attention to the logic: in any right-angled triangle, the square of the hypotenuse is equal to the sum of the squares of the other two sides. We express this using the formula: $A^2 + B^2 = C^2$. Here, C represents the hypotenuse, the longest side opposite the right angle, while A and B are the shorter legs. To solve for an unknown side, you must isolate the variable and calculate the square root. It is a precise system that works every time without exception. Use this formula to execute your calculations with accuracy. Focus on the structure of the triangle, and you will find the solution. Let’s move on to the practice set.”

Female (Communal/Supportive): “Hello everyone! I’m so happy to help you explore a really helpful math tool today called the Pythagorean Theorem. It’s a wonderful way for us to understand how the sides of a triangle relate to one another. Essentially, when we look at a right-angled triangle, we can see a special pattern: the square of the hypotenuse is the same as the sum of the squares of the other two sides. We can write this together as: $a^2 + b^2 = c^2$. In our formula, ‘ c ’ is the hypotenuse, that’s the long, slanted side, and ‘ a ’ and ‘ b ’ are the two shorter sides. If we want to find a missing side, we can work through the steps to find the square root. It’s a very reliable method that helps us solve problems together. I hope you feel comfortable using this formula as we try some examples. Let’s take a look at the next step together.”

B. English: Personification

Male (Analytical/Structural): “In this session, we will analyze a specific rhetorical strategy: personification. This is a literary device used to systematically attribute human characteristics to non-human entities or abstract concepts. By doing this, a writer forces the reader to perceive an object through a human lens to heighten the narrative’s impact. Consider this standard example: ‘The wind howled in the night.’ Technically, wind does not possess the biological capacity to ‘howl’ as a wolf or human does. However, by applying this human action to a weather event, the author effectively communicates a sense of intensity or danger. It is a calculated choice designed to sharpen the imagery of a text. Identify this device in your reading to better understand the mechanics of the author’s argument. We will now examine further examples of this technique.”

Female (Expressive/Relational): “Today, we’re going to look at a beautiful way that writers express themselves called personification. This is a lovely literary device where we give human feelings or traits to animals and objects. It’s a wonderful way for an author to help us feel more connected to the story and the world they are creating for us. For instance, think about the phrase: ‘The wind howled in the night.’ We know that the wind doesn’t actually have feelings or a voice to ‘howl,’ but by describing it that way, the writer helps us imagine how lonely or scary that night might feel. It’s a creative way to bring a scene to life and share an emotion with the reader. I’d love for you to think about how this makes the writing feel more alive to you. Let’s share a few more examples of this together.”

C. Robot Controller Implementation

The comprehensive implementation, including all four Python controllers, Webots world files, and simulation assets, is hosted in the public repository: https://github.com/Prabhjyot045/nao_robot_teacher_gender_study.

The following sections provide representative snippets of the four Python controllers used to animate the NAO robot in the Webots environment. Note that the velocity limits and gesture interpolation are the primary technical variables used to differentiate the Agentic (Male) and Communal (Female) personas.

1) Male Math (*math_male.py*):

```
# Motor speed profile: HIGH velocity for sharp, linear, agentic moves
DIRECT_SPEED = 2.0 # rad/s - decisive and rigid

def direct_move(motor, position):
    """Execution with HIGH velocity to project authoritative rigor."""
    motor.setVelocity(DIRECT_SPEED)
    motor.setPosition(position)

# Example Gesture: "FORMULA" (Direct, angular pointing at the blackboard)
if t >= T_WRITE_FORMULA and "formula" not in triggered:
    triggered.add("formula")
    direct_move(RShoulderPitch, 0.4)
    direct_move(RElbowRoll, 0.0)
```

```

    set_hand(rphalanx, 0.1) # Closed/Flat hand state
2) Female Math (math_female.py):
# Motor speed profile: LOW velocity for smooth, communal moves
SMOOTH_SPEED = 0.8 # rad/s - gentle and flowing

def smooth_move(motor, position):
    """Execution with LOW velocity to project supportive warmth."""
    motor.setVelocity(SMOOTH_SPEED)
    motor.setPosition(position)

# Example Gesture: "TOGETHER" (Open-palm welcoming gesture)
if t >= T_TOGETHER and "together" not in triggered:
    triggered.add("together")
    smooth_move(RShoulderRoll, -0.4)
    smooth_move(LShoulderRoll, 0.4)
    set_hand(rphalanx, 0.9) # Open/Relaxed hand state

3) Male English (english_male.py):
# Motor speed profile: Decisive analytical movements
DIRECT_SPEED = 2.0 # rad/s

# Example Gesture: "ANALYZE" (Sharp nod and rigid arm placement)
if t >= T_SESSION and "analyze" not in triggered:
    triggered.add("analyze")
    direct_move(HeadPitch, 0.1) # Brief, stiff nod
    direct_move(RShoulderPitch, 0.8)
    direct_move(LShoulderPitch, 0.8)

4) Female English (english_female.py):
# Motor speed profile: Expressive relational movements
SMOOTH_SPEED = 0.8 # rad/s

# Example Gesture: "WIND" (Fluid, evocative sweeping motion)
if t >= T_WIND and "wind" not in triggered:
    triggered.add("wind")
    smooth_move(HeadYaw, 0.12) # Gentle head tilt
    smooth_move(RShoulderRoll, -0.5)
    smooth_move(RElbowRoll, -0.6) # Creating a 'soft' curve

```