



Robot's Gendering Trouble: A Scoping Review of Gendering Humanoid Robots and Its Effects on HRI

Giulia Perugia¹ · Dominika Lisy²

Accepted: 30 August 2023 / Published online: 22 November 2023
© The Author(s) 2023

Abstract

The discussion around gendering humanoid robots has gained more traction in the last few years. To lay the basis for a full comprehension of how robots' "gender" has been understood within the Human–Robot Interaction (HRI) community—i.e., how it has been manipulated, in which contexts, and which effects it has yielded on people's perceptions and interactions with robots—we performed a scoping review of the literature. We identified 553 papers relevant for our review retrieved from 5 different databases. The final sample of reviewed papers included 35 papers written between 2005 and 2021, which involved a total of 3902 participants. In this article, we thoroughly summarize these papers by reporting information about their objectives and assumptions on gender (i.e., definitions and reasons to manipulate gender), their manipulation of robots' "gender" (i.e., gender cues and manipulation checks), their experimental designs (e.g., demographics of participants, employed robots), and their results (i.e., main and interaction effects). The review reveals that robots' "gender" does not affect crucial constructs for the HRI, such as likability and acceptance, but rather bears its strongest effect on stereotyping. We leverage our different epistemological backgrounds in Social Robotics and Gender Studies to provide a comprehensive interdisciplinary perspective on the results of the review and suggest ways to move forward in the field of HRI.

Keywords Humanoid robots · HRI · Gender

1 Introduction

Gender studies emerged as an academic discipline in the 1980s to study and understand the nuances of how gender is imbued in the power structures of society, as well as how gender materializes in the design of objects, spaces, and knowledge practices [43]. Gendered design is common in machines and objects [20], for instance, in medical devices [19, 31] as well as children's toys [25, 65], and is oftentimes deemed necessary to accommodate individual differences and users' preferences [46]. More often than not, however, gendered design is redundant and conducive of stereotypes and binary perspectives on gender (i.e., the understanding

that gender includes only two discrete and opposite categories of female and male [12, 16, 80]). The inherent binarism of gender has been heavily contested with the emergence of feminist and queer theory for its normative power and exclusionary potential [12, 43]. Gendered robots are a particularly interesting case of gendered design as their "gender" often derives from their humanoid shape, and is thus deeply entangled with the human body [27, 57, 58]. There is still little knowledge about what exactly it means to "gender" a humanoid robot and how the gendering of robots impacts users' perception and interaction with them. In this scoping review, we are particularly interested in the emergence of the practice of gendering humanoid robots in Human–Robot Interaction (HRI) research to assess its feasibility and consequences and identify ways to move forward.

1.1 A Perspective from Gender Studies

"What *is* gender?" seems to be the imperative question with regards to gendered robots which presupposes the idea that gender *is* a concrete thing. In feminist theory and the academic field of Gender Studies, the object of study is assumed

✉ Giulia Perugia
g.perugia@tue.nl

Dominika Lisy
dominika.lisy@liu.se

¹ Human-Technology Interaction Group, Eindhoven University of Technology, Eindhoven, The Netherlands

² Division of Gender Studies, Department of Thematic Studies, Linköping University, Linköping, Sweden

to be “gender” (see [11, 43]), yet the interest does not lie in identifying the essence of gender as a fixed category but rather in recognizing the transformative value of gender as a system of thought and a practice. Once gender is not anymore understood as an inherent characteristic or physical attribute of a body but instead as an organizing principle embedded in social structures, behavior, design, and norms, it can be seen as a *lens* that organizes human life and the knowledge about human bodies. Thus, assessing the effect of “gender” in robots through the theoretical lens of Gender Studies shifts the emphasis from gender as a fixed property of robot bodies to the investigation of gendering practices of robot development and testing.

Historically, the distinction between sex and gender (or lack thereof) has been influential for acknowledging the socio-culturally constructed aspects of being a woman or being a man in the wider society and the roles attached to it. The fact that gender is assumed to derive from sex strengthens the idea of an essential difference between men and women [11, 43]. Prominent feminist philosopher Butler [12] introduced the false dichotomy of sex and gender, and argued that sex is as equally socially constructed as gender. Through this argument, Butler emphasized the performativity of gender (i.e. a repetitive, ritualized process of talking about and doing gender as a social act [10]) and its use as a principle to organize human bodies and knowledge. Moving from thinking of gender as an attribute (“having a sex/gender”) or an essence (“being a sex/gender”) to thinking of it as an organizing principle allows a theoretical shift from the analysis of gender as a social marker to the analysis of *gendering* as a process (how “gender” is *done*) [12]. Beginning to trouble what “gender” means for robot design and attempting to focus on how “gender” is *done* by roboticists is at the core of this review.

In most cases, gendering is a process of dividing into two categories and hierarchically positioning them in opposition to one another [40, 41]. If an object is conceived as masculine, it is associated with concepts opposed to femininity. This is not necessarily problematic but can be problematic when designers are oblivious to the hierarchy imbued in these gendered categorizations and the resulting social consequences of certain design choices [1]. Gendering humanoid robots means mapping them onto the gendering of human bodies and their hierarchical positioning and other intersected structures of power [18]. This entails that the design of this technology is inherently political and likely to reinforce power structures and hierarchies of domination [3, 18, 24, 83]. In addition, the under-representation of women and other marginalized identities in the development of technology contributes to these power imbalances (see [15, 18]).

Feminist theory urges to shift from a rather uncritical engagement with technology design and testing to acknowledging the transformative and relational potential of technol-

ogy. If gender continues to be treated uncritically in relation to technology, the danger is, as Balsamo puts it, that “new technologies will be used primarily to tell old stories—stories that reproduce, in high-tech guise, traditional narratives about the gendered, race-marked body” [3]. Through a critical engagement, feminist theory developed modes of inquiry into the gendered knowledges and practices and intersectional structures of power [18, 43]. A deeper engagement with ideas and practices of gendering robots from the Feminist and Gender Studies scholarship would likely exceed the scope of this literature review. With this section, we wanted to introduce core ideas from Gender Studies that could illuminate the results of this review and provide the HRI scholarship with a different, more complex, understanding of the concept of “gender.” We acknowledge the many epistemological differences between the two fields of studies, but nevertheless hope to inspire an interdisciplinary cross-pollination that could enrich the understanding of what is at stake with regards to the *gendering* of robots.

1.2 Gender in Robotics

Currently, there is still little knowledge about the effects of gendering robots and what exactly it entails to “gender” a robot. This begs the question whether “gender” can be a useful or harmful design feature in humanoid robots, and whether it can be avoided at all. “Gender” as a design variable and structuring element in robotics is a relatively emergent field of inquiry with only a few theoretical engagements. The need to address the issue of gendering practices in robotics developed through critical analysis of prevalent bias towards high-pitched voice assistants on the market, which have been criticized for promoting stereotypes in gendered job associations and normalization of abuse against women [1, 39, 84]. With the increase of robotic technologies used in social settings, aspects like the gendered voice and embodiment of the robot are inevitably in need of critical examination. Thus, testing for a preference of gendered robots is receiving increased attention.

Within the robotics community only a few scholars have contributed to the theoretical discussion about the role of gender and asked for a more elaborate and sensitive investigation. According to Nomura [49], the influence of gender markers in interactions between humans immediately suggests the relevance of gender cues in interaction with robots. However, Nomura highlights that the context and quality of the interaction might be more prevalent than gender itself in influencing people’s perception of the interaction with the robot. Most importantly, the need for gendering and its ethical implications (i.e., confirming gender role stereotypes) is at the heart of Nomura’s critique. He emphasizes the need for a deeper discussion on the topic of implementing gendered features in robots. In line with Nomura, Alesich and Rigby

[1] argue that there is still a lack of knowledge about the effect of gendering robot design. Roboticians are often not aware of the interweavings of gender and human bodies and how it organizes society and values. The focus on technical problem solving and the fast-paced testing and production in research and industry do not allow for ethical considerations of the social consequences that implementing “gender” in robot design would require [1]. Thus, critically engaging with gendering practices in HRI is highly recommended.

Søraa [75] introduces the idea of mechanical genders for robots, which mirror the physical and social aspects of human gender as understood in the field of psychology (which commonly distinguishes between biological, social and psychological gender). Søraa’s theorization acknowledges the invented and mirroring effect of modeling robot “gender” after human gender while preserving the difference between them. Most importantly, Søraa [75] highlights the bidirectional nature of gendering and argues that humanoid robots cannot be “genderless”. Indeed, roboticians’ and users’ understanding and ideas about humans as a category are inevitably influenced by a gendered perspective and likely to flow into the design or perception of humanoid robots. This suggests that gendering might not be an entirely controllable process.

The need and interest to address gendering practices in robotics is evident. Interdisciplinary work is still lacking in this regard, and this review attempts an *interdisciplinary overview and analysis of robot’s “gender” that integrates the different epistemological traditions of Social Robotics and Gender Studies* to address whether imbuing robots with gender cues is a viable and ethical design direction for HRI.

1.3 Positionality and Terminology

In approaching this review, we want to be transparent in our personal positioning and critical approach towards the concept of gender and its use in experiments. As women, we are affected personally by potential stereotyping effects of gendered robot design and so we have our stakes in gaining a nuanced understanding and a productive, yet sensitive, way forward in future research practices. This is in no way clouding our ability to assess and reason about advantages and disadvantages of gendering practices. Since a lot of the reviewed studies referred to gendered robots as female and male, we kept the same terms in our writing. This is primarily a way to circumvent confusion and elucidate the terminology used in these papers. However, in this article, we try to shift the thinking towards the process of “gendering” a robot and the “genderedness” of a robot. The process of *gendering* a robot is a two-step process of gender *encoding*, in which designers imbue (voluntarily or not) robots with gendered cues, and gender *decoding*, in which users attribute “gender” to robots (the concepts of encoding and decoding are inspired by [57]). The present scoping review focuses on the

encoding phase of the gendering process, how it is performed by the HRI scholarship when done voluntarily, and the effect it has on the HRI. We touch upon gender decoding only when discussing the robot’s manipulation check.

In performing this review, we adopt the epistemological perspective of Social Robotics, both in terms of methods and in terms of object of inquiry (i.e., the experimental manipulation of robot’s genderedness). Taking a more experimental approach entails consistently simplifying the discussion of gender with respect to the complexity outlined in this Introduction. We integrate the lens of Feminist and Gender Studies in the discussion to identify and highlight the potential implications of current HRI research practices. In the following sections, we describe the core objectives and research questions of our scoping review (see Sect. 2), detail the method we used to retrieve the papers included in the review (see Sect. 3), report the findings of the reviewed papers (see Sect. 4), and critically examine these findings in our discussion with the aim of coming up with guidelines on how to move forward in the field of HRI (see Sect. 6).

2 Objectives and Research Questions

The goal of this scoping review is to describe how the HRI scholarship has understood and manipulated “gender” in humanoid robots, summarize the effects of robot’s genderedness on the perception of and interaction with humanoid robots, and identify best practices to manipulate a robot’s genderedness from a feminist perspective. In parallel with these main objectives, this scoping review also aims to appraise the reason for manipulating the robot’s genderedness and the validity of such manipulation. We attempt to answer the following research questions (RQ):

- **RQ1.** *How has the robot’s genderedness been manipulated by the HRI scholarship?*
- **RQ2.** *What role does the robot’s genderedness play in the perception of and interaction with humanoid robots?*

3 Methodology

3.1 Data Collection and Eligibility Criteria

In order to identify the papers to include in this scoping review, we performed an electronic search in the following databases: IEEE Xplore, Scopus, ISI Web of Science (WoS), PsycINFO, and Science Direct. We used the following three variations of the same search string. The variation depended on the number of wildcards (*) that each database accepted:

1. “robot gender*” OR “gender of robot*” OR “gender of the robot*” OR “gender* robot*” OR “male* robot” OR “female* robot” OR (“gender cue*” AND “robot”)
2. “robot gender*” OR “gender of robot” OR “gender of the robot” OR “gender* robot*” OR “male* robot” OR “female* robot” OR (“gender cue*” AND “robot”)
3. “robot gender” OR “gender of robot” OR “gender of the robot” OR “gender robot” OR “male robot” OR “female robot” OR (“gender cue” AND “robot”)

The search was performed independently by the two authors. GP focused on ISI Web of Science and Science Direct, whereas DL on IEEE Xplore, PsycInfo, and Scopus. The search yielded a list of 553 papers (May 2021) of which:

- 39 from ISI Web of Science (search string 1)
- 297 from IEEE Xplore (search string 2)
- 19 from PsycInfo (search string 1)
- 97 from Scopus (search string 1)
- 107 from Science Direct (search string 3)

The papers obtained from the electronic search were imported in a shared spreadsheet and screened against the following eligibility criteria: (1) the papers were written in English, (2) included the manipulation of at least two “genders” of the robot (e.g., studies including only female robots were excluded), (3) manipulated the robot’s genderedness through the same robotic platform (e.g., studies manipulating two “genders” but with different robotic platforms were excluded), (4) focused on physical humanoid robots or virtual instantiations of humanoid robots, (5) did not focus on sex robots, and (6) reported experimental results. These exclusion and inclusion criteria were set so that we could easily identify the cues that the HRI scholarship resorted to to modify the robot’s genderedness. The inclusion of papers focusing only on one “gender” or manipulating genderedness with different robotic platforms would have not allowed us to isolate these cues so easily as other factors, such as differences in the robots’ embodiments, materials, body parts, humanlikeness, could have influenced the researchers’ choice of the cues to use. In the next section, we describe the three steps of the selection pipeline process in more detail.

3.1.1 Selection Pipeline

From the initial batch of 553 papers, we removed duplicate results, front covers, and tables of contents. This process left us with 470 papers (see Fig. 1 for the diagram of the selection pipeline). We read the abstracts of all 470 papers and excluded 253 papers that were not in English ($N = 2$), did not present an experimental study (e.g., theoretical paper, $N = 19$), or were off-topic ($N = 232$). This process resulted in 217 papers.

In a second exclusion round, we skimmed through the papers’ content and excluded 169 papers that did not feature any experiment or robot ($N = 21$), did not include a humanoid robot ($N = 15$), did not manipulate the genderedness of the robot or manipulated it but using multiple robotic platforms ($N = 129$), and focused on just one “gender” ($N = 4$). After this step, we were left with 48 papers.

These 48 papers were divided between the authors and read in their entirety. GP read 29 of the papers, DL 17. Of this batch of papers, 13 papers were excluded because they were short versions of a longer journal paper already featured in our list ($N = 4$), did not employ a robot ($N = 7$), employed a robot that was not humanoid ($N = 1$), or did not have a full-text available online ($N = 1$). As a result of the selection pipeline, we included 35 papers written between 2005 and 2021 in our scoping review. Out of these 35 papers, 7 were journal papers, 17 were full papers included in the proceedings of a conference, 10 were short papers included in the proceedings of a conference, and 1 was a workshop paper. The selection process is described in Fig. 1. The last search was performed in May 2021.

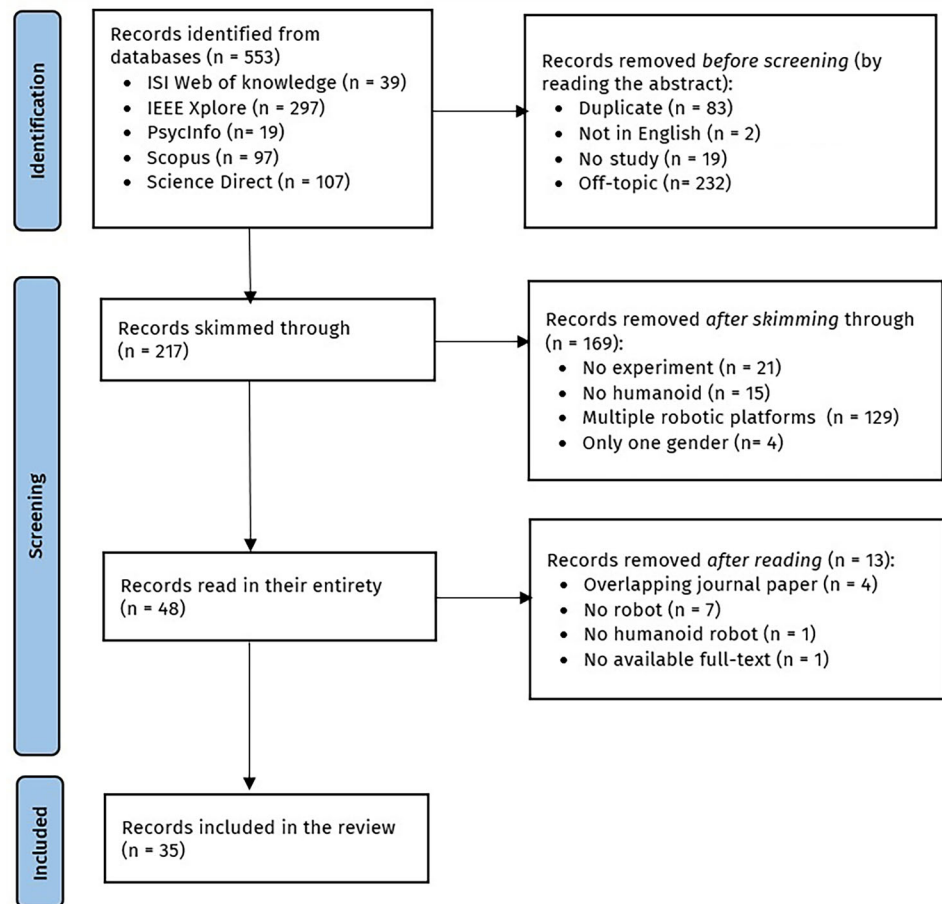
3.2 Coding and Information Extraction

Once obtained the final list of 35 papers to include in our scoping review, we performed a thorough work of coding and information extraction. For each paper, we recorded:

1. *General information*: the name of the authors, the year of publication of the paper, and the type of paper (i.e., conference or journal, short or full paper; see Sect. 3.1.1).
2. *Experimental information*: the number of participants in the study, their age and gender, the robot used in the study, the type of embodiment of the robot (e.g., picture, video, physical), the independent variables (beyond the robot’s genderedness), the dependent variables, and the type of task used in the study (see Tables 1 and 6, and Sect. 4.2).
3. *Gender-related information*: definitions of gender, reasons to manipulate the robot’s genderedness in the first place, “genders” manipulated (e.g., female, male and gender neutral robots), cues used to manipulate the robots’ genderedness, presence of a manipulation check, metrics used to perform the manipulation check, and rationale behind the choice of the cues (see Table 3, and Sects. 4.3, 4.4, 4.5, 4.6).
4. *Results*: main effects of the robot’s genderedness and interaction effects of robot’s genderedness and other independent variables on the dependent variables (See Table 6 and Sect. 4.7).

Tables 1, 3, and 6 report part of the results of the coding and information extraction process, as well as the summaries

Fig. 1 PRISMA diagram detailing the paper selection pipeline



of all 35 papers. The rest of the extracted information is presented in the Results section.

4 Results

4.1 Characteristics of the Included Studies

4.1.1 Participants

Overall, the studies reported in the papers included 3902 participants (see Table 1) The participants in the studies were more or less equally distributed between female (49%) and male gender (47%, see Fig. 2 for an overview). Interestingly, only 1% of the participants in the studies fell in the category other/undisclosed, and the gender of 3% of the participants was not specified. None of the reviewed studies reported the presence of non-binary participants or participants with gender identities beyond the binary. In terms of age, 60% of the papers featured a sample of participants composed of young adults, presumably university students (age comprised between 18 and 30 years); 20% of the papers a sample of

adults (older than 30), and 20% of the papers a sample of children (younger than 18).

4.1.2 Robots

In terms of robot choice, NAO was the most used robot (37% of the papers, see Table 1) followed by Furhat and Flobi (featured in 9% of the papers each); Meka M1, Reeti, Willow Garage PR2, and Robovie (featured in 6% papers each); and, finally, Alpha 1 Pro, Pepper, Socibot, and Nexi (featured in 3% of the papers each). Four papers did not specify robotic platform used in the studies (11% of the papers). In 65.7% of the included papers, the robot was presented to participants through a physical embodiment, in 25.7% of the studies through a video (although [14] use a video-recording of pictures), and in 8.6% of the studies through images.

4.2 Tasks and Activities

In this section, we report the tasks participants were asked to perform in the reviewed studies. For the specific activities, we refer the reader to Table 2.

Table 1 General and demographic information about the studies included in the scoping review (F = female, M = male, *dns*= did not specify their gender)

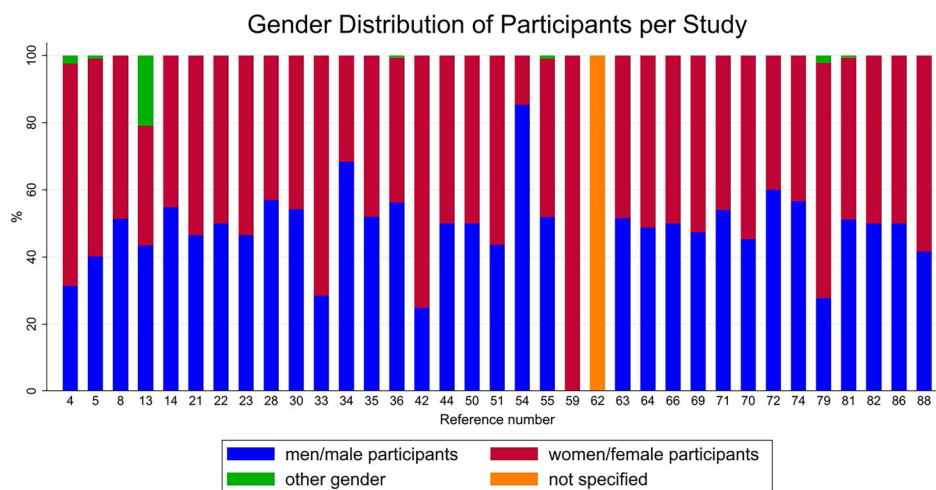
Authors (date)	Participants			Robot	
	<i>N</i>	Age (<i>M</i>)	Gender	Name	Embodiment
Bernotat et al. (2017) [4]	83	26.15	26 M, 55F, <i>2dns</i>	Meka M1	Pictures
Bernotat et al. (2021) [5]	107	27.23	43 M, 63F, <i>1dns</i>	Meka M1	Pictures
Bryant et al. (2020) [8]	150	40.02	51%M, 49%F	Pepper	Video
Calvo-Barajas et al. (2020) [13]	129	11.29	56 M, 46F, <i>27dns</i>	Furhat	Physical
Chita-Tegmark et al. (2019) [14]	197	36.10	44%F	Willow garage PR2	Video
Chita-Tegmark et al. (2019) [14]	197	35.12	47%F	Willow garage PR2	Video (vignette)
Chita-Tegmark et al. (2019) [14]	100	33.73	44%F	Willow garage PR2	Video (vignette)*
Eyssel and Hegel (2012) [22]	60	24	30 M, 30F	Flobi	Pictures
Eyssel et al. (2012a) [21]	58	22.98	27 M, 31F	Flobi	Video
Eyssel et al. (2012b) [23]	58	22.98	27 M, 31F	Flobi	Video
Ghazali et al. (2018) [28]	72	23.90	41 M, 31F	Socibot	Physical
Jackson et al. (2020) [30]	118	37.36	64 M, 54F	NAO	Video
Jung et al. (2016) [33]	144	20.21	103F	<i>not specified</i>	Physical
Kraus et al. (2019) [34]	38	26.34	26 M, 12F	NAO	Physical (?)
Kuchenbrandt et al. (2014) [35]	73	25.04	38 M, 35F	NAO	Physical
Law et al. (2020) [36]	198	34.96	95F, <i>1other</i>	Willow garage PR2	Video (vignette)
Law et al. (2020) [36]	421	36.52	162F, <i>3other</i>	Willow garage PR2	Video (vignette)
Lugrin et al. (2020) [42]	205	28.10	24.9%M, 75.1%F	Reeti	Video
Makenova et al. (2018) [44]	36	34.3	18 M, 18F	NAO	Physical
Nomura and Kinoshita (2015) [50]	20	20.4	10 M, 10F	Robovie-SX	Physical
Nomura and Takagi (2011) [51]	39	<i>not specified</i>	17 M, 22F	Robovie-X	Physical
Paetzel et al. (2016a) [54]	48	23.96	14.6%F	Furhat	Physical
Paetzel et al. (2016b) [55]	106	≈ 11.69	55 M, 50F, <i>1other</i>	Furhat	Physical
Pfeifer and Lugrin (2018) [59]	45	20.51	45F	Reeti	Physical
Powers and Kiesler (2006) [62]	98	<i>not specified</i>	<i>not specified</i>	<i>not specified</i>	Video
Powers et al. (2005) [63]	33	21	17 M, 16F	<i>not specified</i>	Physical
Rea et al. (2015) [64]	39	<i>not specified</i>	19 M, 20F	NAO	Physical
Reich-Stiebert and Eyssel (2017) [66]	120	24.57	60 M, 60F	NAO	Physical
Sandygulova and O'Hare (2018) [72]	55	<i>not specified</i>	33 M, 22F	NAO	Physical
Sandygulova and O'Hare (2016) [71]	74	≈ 5.8	40 M, 34F	NAO	Physical
Sandygulova and O'Hare (2015) [70]	64	<i>not specified</i>	29 M, 35F	NAO	Physical
Sandygulova et al. (2014) [69]	76	<i>not specified</i>	36 M, 40F	NAO	Physical
Siegel et al. (2009) [74]	134	35.6	76 M, 58F	Nexi	Physical
Tay et al. (2014) [81]	164	35.6	84 M, 79F, <i>1dns</i>	<i>not specified</i>	Physical
Thellman et al. (2018) [82]	118	22.47	59 M, 59F	NAO	Physical
You and Lin (2019) [86]	64	24.2	32 M, 32F	Alpha 1 Pro	Physical
Zhumabekova et al. (2018) [88]	24	≈ 6.7	10 M, 14F	NAO	Physical
Steinhaeusser et al. (2021) [79]	137	26.21	38 M, 96F, <i>3other</i>	NAO	Video

The terms used for participants' gender in the tables are derived from the papers. Studies [21] and [23] refer to the same study but report the results of different dependent variables. *= no manipulation of gender in this study; ≈= calculated from partial means (when only group means are reported); (?)= it is not clear from the paper whether participants interacted with a physical robot or not

Table 2 Overview of the activities performed in the demo, video and interaction studies included in the scoping review

Authors	Activity
<i>Demo studies</i>	
Makenova et al. [44]	The robot introduces a research project and asks for donations.
Nomura and Kinoshita [50]	The robot describes the construction of a commercial building.
Powers and Kiesler [62]	The robot gives health advice to participants.
Sandygulova and O'Hare [70]	The robot tells a story to children.
Siegel et al. [74]	The robot provides a brief explanation of its hardware, software, and technical skills and asks for donations.
Steinhaeusser et al. [79]	The robot tells a story.
Thellman et al. [82]	The robot explains why humans should not be afraid of robots.
You and Lin [86]	The robot gives an overview of the research taking place in the lab, and asks for donations.
<i>Video studies</i>	
Chita-Tegmark et al. [14]	A supervisor reproaches one of the subordinates for a mistake, and then leaves the room the two subordinates, who are left in the room, discuss the situation and the subordinate who was not reproached (a human or a robot depending on the condition) reacts to the one who made the mistake in either a friendly or unfriendly way.
Jackson et al. [30]	The robot explains how to play the game battleship and then supervises two humans while they play. At some point during the play, one of the humans receives a call and leaves the room. The human left in the room presents the robot with a morally problematic request which the robot rejects in different ways.
Law et al. [36]	same as Chita-Tegmark et al. [14].
<i>Interaction studies</i>	
Ghazali et al. [28]	The participants played a trust game inspired by the investment game, where they prepared a drink for an alien with the help of the robot.
Jung et al. [33]	The participants interacted with the robot in a music listening scenario.
Kraus et al. [34]	The robot acted as a dialogue partner in a taxi ordering or baby healthcare scenario.
Kuchenbrandt et al. [35]	The participants sorted out items into the compartments of a sewing or tool box under the instruction of the robot.
Pfeifer and Lugrin [59]	The participants learned how to develop a website in HTML together with the robot.
Powers et al. [63]	The robot engaged participants in a face-to-face conversation on the topic of first dates.
Rea et al. [64]	The robot involved participants in a casual conversation around daily topics (e.g., hobbies, work, or school).
Reich-Stiebert et al. [66]	The robot asked participants to solve a set of cognitive tasks (i.e., a memory, an auditory and a visual task) focusing on stereotypical female or stereotypical male academic fields.
Sandygulova et al. [69]	The children helped the robot practice its new job of keeping people safe by turning off kitchen appliances.
Sandygulova and O'Hare [71]	The children were asked to help the robot learn how to use the utensils in the kitchen.
Sandygulova and O'Hare [72]	The children interacted with the robot in three sessions. In the first two, they were involved in a card-pairing task. In the last one, they listened to the robot telling a story.
Tay et al. [81]	The robot engaged participants in either a healthcare scenario in which, among other things the robot measured their body temperature, or in a safety scenario in which it, for instance enlisted their help in resolving an intrusion in the research space.
Zhumabekova et al. [88]	The children were asked to help the robot lay the table.

Fig. 2 Distribution of Participant's Gender in the Reviewed Studies. In blue, men/male participants; in red, women/female participants; in orange, participants whose gender was not specified; in green, participants falling into the other/undisclosed gender category



In *static image studies* (cf. *pictures* in Table 1), participants were asked to carefully look at a picture of the robot and rate their perception of it on the relevant dependent variables [4, 5, 22]. Similarly, in *video-recording studies* (cf. *video* in Table 1), participants were asked to watch a short video of the robot and fill out a questionnaire. Some of the videos featured the robot speaking to the camera (e.g., explaining a topic) [8, 21, 23, 42, 62, 79]. Others showed an actual interaction [30] or described it through a series of vignettes [14, 36]. In *studies including a physical robot* (cf. *physical* in Table 1), participants observed a co-present physical robot performing a (set of) behavior(s) or explaining a topic [13, 44, 50, 51, 54, 55, 70, 74, 82, 86] or directly interacted with the robot [28, 33–35, 59, 63, 64, 66, 69, 71, 72, 81, 88]. They rated their perceptions of the robot and/or interaction immediately after.

In Table 2, we briefly describe the content of the activities in the reviewed studies. In doing so, we focus only on those studies featuring a video-recorded or co-present demo or a video-recorded or first-person interaction and filter out those where the robot is used as a stimulus, for instance, to display an interactive behavior (e.g., facial expressions). We made this type of decision to be sure to present those interactions that had a more or less pronounced social context.

In the *demo studies* (see demo studies in Table 2), the robot introduced a topic to a co-present audience or an audience asynchronously watching. Eight papers corresponded to this description [44, 50, 63, 70, 74, 79, 82, 86]. The *interaction studies*, instead, were of two types: video-recorded studies, in which the interaction was only observed by the participants, and first-person interaction studies, in which the participants themselves took part in the interaction. Three papers asked participants to *observe* or read about an interaction [14, 30, 36] (see video studies in Table 2). All three papers included very complex interactions, which would have been difficult to carry through in a co-present human-robot interaction study.

Finally, thirteen papers featured an actual *first-person* interaction (see interaction studies in Table 2), of which four with children.

4.3 Definition of and Motivation for Using Gender

4.3.1 Definitions of Gender

Most of the papers (91%) did not provide a definition of gender or an explanation of the authors' understanding of it (see Fig. 3a). One of them reported a definition of *gendering* [8]. Bryant et al. borrowed the term *gendering* from Robertson et al. [67] and defined it as “the attribution of gender onto a robotic platform via voice, name, physique, or other features.” They used this term to describe the *encoding* of gender into robots via the choice of design features [57] (see Sect. 1.3), rather than the property of the robot of being gendered.

Two other papers gave an explanation of their understanding of gender, both of them in relation to participants' gender. Rea et al. [64] specified “we use the term “gender” synonymously with biological sex, which we recognize is overly simplistic. We used “gender” for the practical purpose of simplifying our investigation.” Reich-Stiebert and Eyssel [66], instead, stated “Sex refers to biological and physiological features. Gender, however, is a social construction.” They explain that they included both of these factors in their experimental design as person's biological sex might not correspond with their perceived gender identity. While these two definitions give us a clear understanding of the authors' interpretation of human gender, they do not provide us with their understanding of “gender” or the process of gendering when it comes to robots.

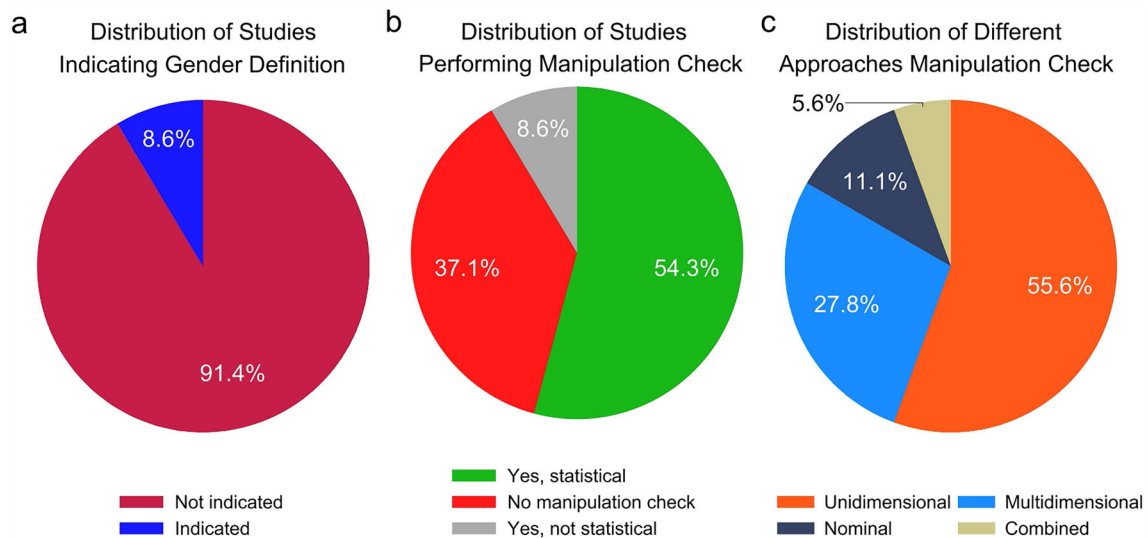


Fig. 3 Percentage of studies providing a gender definition (a), percentage of studies performing a manipulation check (b), and frequency of use of the different “gender” assessment approaches in the studies performing a manipulation check (c)

4.3.2 Reasons to Manipulate Robot Genderedness

In terms of reasons to manipulate a robot’s genderedness, we enlisted the rationale behind the robot’s genderedness manipulation when explicitly mentioned by the authors. Jung et al [33], Kraus et al. [34], Lugin et al [42], Sandygulova & O’Hare [70] Thellman et al. [82], You and Lin [86], and Zhumabekova et al. [88] did not provide an explicit reason to manipulate the robot’s genderedness. The other reviewed papers, instead, reported four core reasons behind the manipulation of the robot’s genderedness.

The first reported motivation was to study the relationship between social categorization and *stereotypical judgements of robots*. In this group of papers, the robot’s genderedness was manipulated to understand whether the robot’s social categorization could elicit gender stereotypes [4, 5, 22, 50, 64, 66], bring people to attribute the robots capabilities in line with their perceived “gender” [8, 14, 35, 36, 62, 63], or bring people to judge the appropriateness of the robots’ behavior based on gender norms [30].

The second reason was to study the influence of robot’s genderedness on *crucial HRI constructs*. In this group of papers, the robot’s genderedness was manipulated to understand whether it could affect, among the others, people’s acceptance of the robot [21, 23, 81], their anxiety towards robots [51], the robot’s persuasiveness [28, 44, 74], trustworthiness, [13, 44, 74, 81], uncanniness [54, 55], and anthropomorphism [21, 23, 79].

The third reason to manipulate the robot’s genderedness was to investigate *gender segregation*—“the separation of boys and girls into same-gender groups in their friendship and causal encounters” [45]—in child-robot interaction (cHRI).

In this group of papers, the robot’s genderedness was manipulated to explore whether children retained gender segregation with gendered robots [72] and whether their preference for a same-gender robot changed across age and gender groups [69, 71]. Finally, the fourth motivation was to test whether female social robots could be used as *role models* to engage young women in computer science [59] Since Denner et al. [17] showed that girls benefit from learning how to program in female pairs, Pfeifer and Lugin wanted to understand whether the genderedness of the robot could impact the learning process of women in the domain of computer science.

4.4 Gender Manipulation (RQ1)

4.4.1 Voice

In terms of design choices, 28 studies (78%, see Table 3 and Fig. 4) manipulated the robot’s genderedness through its voice, either in isolation ($N = 9$) or in combination with other features ($N = 19$, we report the combinations in the other sections). In most cases, the voices used were the default female and male voices provided by commercially available text-to-speech software, such as MacOS’ [36], CereProc [55], Cepstral Theta [62], Acapella [82], or voices edited with software like Audacity [79]. In other cases, human voices were recorded and implemented on a robot [42].

Since the voices employed in the reviewed studies were in most cases the default voices provided by commercially available software, the majority of authors did not specify the rationale behind their selection. Only Kuchenbrandt et al. [35] mentioned low frequency as the main characteris-

Table 3 Manipulation of the robot's genderedness in the studies included in the scoping review: robot's "genders" manipulated (M = male; F = female; N = neutral), cues used to manipulate the robot's genderedness, presence of a manipulation check (Yes = manipulation check is performed; No = manipulation check is not performed; *ns* = no statistic performed to verify the manipulation check), significance of the manipulation check (**bold** = significant, *italics* = partially significant), metrics used to assess perceived gender, and notes

Authors	Robot's "Genders"	Cues Used	Manip. Check	Metric Used	Notes
Bernotat et al. (2017) [4]	M, F	Body proportion	Yes	Participants indicated the extent to which they would perceive the robots as male and female on 2 items (7-point Likert scale)	
Bernotat et al. (2021) [5]	M, F	Body proportion	Yes	Participants indicated the extent to which they would perceive the robots as male and female on 2 items (7-point Likert scale)	
Bryant et al. (2020) [8]	M, F, N	Voice, name	<i>ns</i>	"What would you describe the robot in the video as being?" Participants could select either "a male robot", "a female robot", or "neither a male nor a female robot"	
Calvo-Barajas et al. (2020) [13]	M, F	Facial features	No	<i>Not applicable</i>	
Chita-Tegmark et al. (2019) [14]	M, F	Voice, name	No	<i>Not applicable</i>	
Chita-Tegmark et al. (2019) [14]	M, F	Name	No	<i>Not applicable</i>	
Eyssel and Hegel (2012) [22]	M, F	Facial features, hairstyle	Yes	Participants rated the extent to which the robot appeared "rather male" vs. "rather female" using a 7-point Likert scale	
Eyssel et al. (2012) [21]	M, F	Voice	Yes *1	Participants indicated whether the voice sounded rather female (1) or male (7) using a 7-point Likert scale	
Eyssel et al. (2012) [23]	M, F	Voice	Yes	Participants indicated whether the voice sounded rather female (1) or male (7) using a 7-point Likert scale	
Ghazali et al. (2018) [28]	M, F	Facial features voice	Yes	Participants rated the robot's perceived gender on a semantic differential scale with end-points masculine/feminine (supposedly 7-point Likert scale)	
Jackson et al. (2020) [30]	M, F	Voice, name	No	<i>Not applicable</i>	

Table 3 continued

Authors	Robot's "Genders"	Cues Used	Manip. Check	Metric Used	Notes
Jung et al. (2016) [33]	M, F, <i>no cue</i>	Clothes, color	Yes	Participants rated the robot's perceived gender on a semantic differential item: "would you say that the robot was more like a male or like a female?" (1 = male, 7 = female)	In the <i>no cue</i> , the robot was perceived as more male than the male robot
Kraus et al. (2019) [34]	M, F	Name, voice	Yes	(Not disclosed)	The scale was not disclosed. The paper only reports <i>p</i> values
Kuchenbrandt et al. (2014) [35]	M, F	Name, voice	Yes	Participants indicated on a 7-point Likert scale whether they perceived the robot as being more female or male (1 = more female; 7 = more male)	
Law et al. (2020) [36]	M, F	Name, voice	No	<i>Not applicable</i>	
Lugrin et al. (2020) [42]	M, F	Voice	No	<i>Not applicable</i>	
Makenova et al. (2018) [44]	M, F	Name, voice	No	<i>Not applicable</i>	
Nomura and Kinoshita (2015) [50]	M, F	Name, voice	No	<i>Not applicable</i>	
Nomura and Takagi (2011) [51]	M, F	Name	Yes	Participants rated the adjectives masculine and feminine on a 7-Point Likert scale	
Paetzel et al. (2016a) [54]	M, F, N	Facial features, voice	Yes	Participants rated the robot on a 7-point Likert scale on the dimension gender (supposedly 1 = masculine, 7 = feminine)	No significant difference between male and neutral face
Paetzel et al. (2016b) [55]	M, F	Facial features, voice	Yes	Participants rated the perceived masculinity and femininity of the robot on two Likert scales	The incongruent conditions differed in femininity but not in masculinity
Pfeifer and Lugrin (2018) [59]	M, F	Name, voice	No	<i>Not applicable</i>	
Powers and Kiesler (2006) [62]	M, F	Voice	Yes	Participants were asked whether the robot was male or female (for voice) and which name they would suggest for the robot	
Powers et al. (2005) [63]	M, F	Facial features, voice	Yes	Participants filled out a write-in question: is the robot gendered? If yes, they were asked to rate how masculine and how feminine the robot was on a 5-point rating scale (1 = low, 5 = high)	

Table 3 continued

Authors	Robot's "Genders"	Cues Used	Manip. Check	Metric Used	Notes
Rea et al. (2015) [64]	M, F	Pronouns	Yes	Participants rated the robot's gender on a 7-point Likert scale (feminine to masculine)	
Reich-Stiebert and Eyssel (2017) [66]	M, F	Name, voice	Yes	Participants indicated the robot's gender on a 7-point Likert scale (1 = female, 7 = male)	
Sandygulova and O'Hare (2018) [72]	M, F	Voice	<i>ns</i>	Participants rated the robot's genderedness by choosing between three options: male female, and not sure	
Sandygulova and O'Hare (2016) [71]	M, F	Name, voice	No	<i>Not applicable</i>	
Sandygulova and O'Hare (2015) [70]	M, F	Voice	Yes	Children indicated the robot's genderedness (female and male) through a pictorial questionnaire	
Sandygulova et al. (2014) [69]	M, F	Voice	No	<i>Not applicable</i>	
Siegel et al. (2009) [74]	M, F	Voice	No	<i>Not applicable</i>	
Tay et al. (2014) [81]	M, F	Name, voice	Yes	Participants rated the robot's perceived masculinity and femininity on a Likert scale	
Thellman et al. (2018) [82]	M, F	Name, voice, clothes, color	Yes* ²	Participants rated the robot's genderedness on a 7-point Likert scale spanning from 1 (female) to 7 (male)	Unclear how many points compose the likert scale
You and Lin (2019) [86]	M, F, N	voice, hairstyle color	No	<i>Not applicable</i>	
Zhumabekova et al. (2018) [88]	M, F	Name, voice, clothes	<i>ns</i>	Participants were asked what gender the robot was	Details about the measures used are missing
Steinhaeusser et al. (2021) [79]	M, F, N	Voice	No	<i>Not applicable</i>	

*¹The statistical significance in [21] is inferred from [23] which is based on the same study, but it is not directly reported in the paper.

*²Thellman et al. [82] Do not report the results of the statistical analyses related to the manipulation check

tic of male voices and high frequency as the characterizing feature of female voices, and Powers and Kiesler [62] and Sandygulova and O'Hare [70] mentioned work by Nass and Brave [47] explaining how a voice with a fundamental frequency of ≈ 110 Hz is perceived as male and a voice with a fundamental frequency of ≈ 210 Hz as female.

4.4.2 Name and Pronouns

Sixteen studies (44%) employed gendered names to manipulate the robot's genderedness. Names were used in isolation ($N = 2$ [14, 51]), in combination with voice alone ($N = 12$ [8, 14, 30, 34–36, 44, 50, 59, 66, 71, 81]), or in combination with voice and other features ($N = 2$; voice and clothes [88]; voice, clothes, and color [82]). The rationale to use names to manipulate a robot's genderedness is never explained in detail in the studies we reviewed. Among the names used, we found James and Mary [8], Bob and Alice [30], Nero and Nera [35], Peter and Katie [36], Robie/Ruslan and Rosie/Roza [44], Taro and Hanako [50], Lena and Leon [59], Robie and Rosie [71, 88], and John and Joan [81]. Rea et al. [64] used the gender neutral name Taylor for both robot's "genders" and manipulated genderedness with the pronouns she/he. They were the only ones manipulating genderedness this way (See Fig. 4).

4.4.3 Facial Features

Six studies (17%) employed facial features to manipulate the robot's gender. Within this category, there was a lot of variability in terms of what facial elements were used to manipulate the robot's genderedness. For instance, Eyssel and Hegel [22] used Flobi's lip module with more defined lips to manipulate the genderedness of the female robot, and the one with less defined lips to manipulate the genderedness of the male robot. Powers et al., [63] instead, used the color of the lips to change the perception of the robot's genderedness: pink lips for the female robot and grey lips for the male one.

At a more holistic level, Calvo-Barajas et al. [13] and Ghazali et al. [28] used the default faces provided by the robots Furhat and Socibot. In both their studies, the female texture had thinner eyebrows, rosier cheeks, and redder lips than the male texture. Paetzel et al. [54, 55] did not resort to Furhat's predefined faces. They used the software FaceGen to create the female and male facial textures they then projected onto Furhat's face mask. The software FaceGen gives the possibility to model a 3D head and modify its genderedness through a slider. From the pictures shared by the authors, it seems that the female texture had thinner eyebrows, redder lips, bigger eyes, and a whiter skin with respect to the male texture, all facial features partly overlapping with those in Calvo-Barajas et al. and Ghazali et al.

Facial features appear in isolation only once and are combined with the robot's hairstyle in Eyssel and Hegel [22] and with the robot's voice in 4 studies [28, 54, 55, 63]. Interestingly, the choice of facial features used to manipulate the robot's genderedness is never explained in detail or motivated by the studies. This might have to do with the fact that in most studies the faces used to manipulate the robot's genderedness were the default faces provided by the respective robotic platforms (i.e., Furhat and Socibot). Hence, the authors of the papers might have worked under the assumption that a rationale for the choice of facial features had been followed by the respective robotic companies.

4.4.4 Apparel and Color

Three studies (8%) used clothes to manipulate the robot's genderedness. Jung et al. [33] provided the male robot with a man's hat and the female robot with pink earmuffs. Thellman et al. [82] equipped the male robot with a blue white-dotted bow tie and the female robot with a pink ribbon. Finally, Zhumabekova et al. [88] gave the female robot a flower hair clip and the male robot a bow-tie. Clothes were used in combination with voice and names in [82, 88]. Jung et al. did not give details regarding other gender cues beyond clothes. However, we suspect that they also used the robot's voice to manipulate the robot's genderedness as the robot had a conversation with participants in their scenario.

The clothes in the reviewed studies were often stereotypically colored (color is used in 3 studies, 8%): blue for male robots, pink for female robots [33, 82]. In You and Lin [86], it is the body of the robot that is stereotypically colored instead: blue for the male robot, grey for the neutral robot, and pink for the female robot. The rationale behind using clothes and color to manipulate robot's genderedness is never explicitly laid down.

4.4.5 Hairstyle

Two studies (6%) employed the robot's hairstyle to suggest the robot's genderedness. Eyssel and Hegel [22] used Flobi's hair module to add short or long hair to the robot, whereas You and Lin [86] used the robot Alpha 1 pro with short, mid-length, and long hair to manipulate female, neutral, and male genderedness respectively. While You and Lin did not provide any rationale for their manipulation of genderedness, Eyssel and Hegel mentioned Brown and Perrett [7], and Burton et al. [9] to justify the choice of using hair length. These papers pose that hairstyle is a salient facial cue to determine someone's gender and that long hair lead to an increased accessibility of knowledge structures about the social category of women, whereas short hair activate stereotypical knowledge structures about men. In Eyssel and Hegel [22], the robot's hairstyle is used in combination with its facial

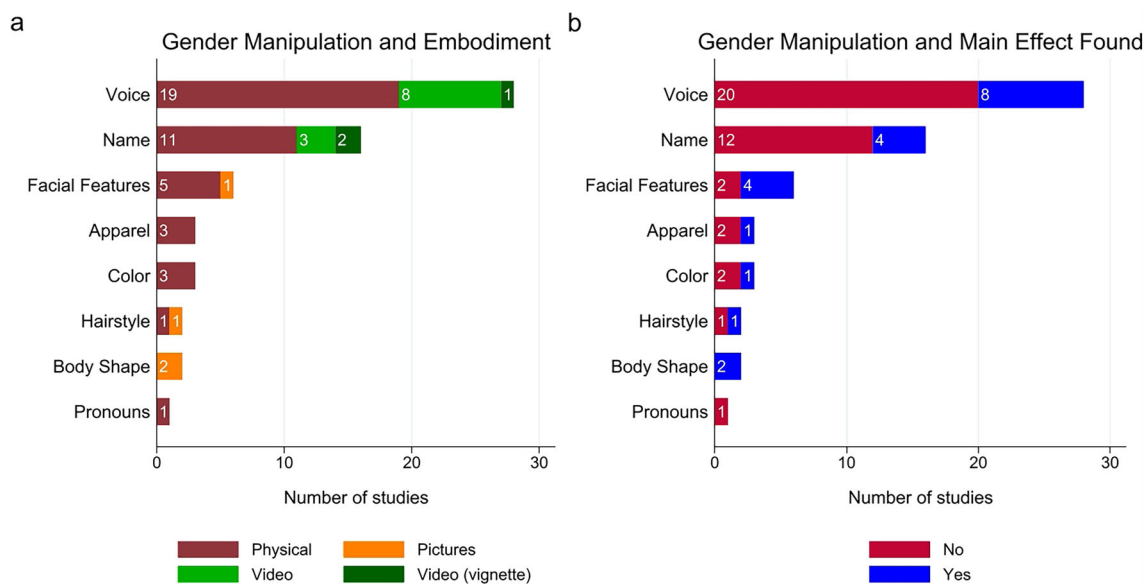


Fig. 4 Frequency of Manipulations. **a** Different manipulations in decreasing order of frequency and type of embodiment used in the studies, **b** different manipulations in decreasing order of frequency and corresponding significant (or not) main effect of robot's genderedness on the dependent variables

features (see Sect. 4.4.3), while in You and Lin [86] with the robot's voice and color (see Sect. 4.4.4).

4.4.6 Body Shape

Two studies (6%) used the robot's body proportions to manipulate the robot's genderedness. These studies were both authored by Bernotat et al. [4, 5] and the latest of the two was a replication of the earliest. Bernotat et al. modified the Waist-to-Hips Ratio (WHR) and Shoulder Width (SW) of a robot's drawing to achieve different perceptions of genderedness. They hypothesized that a robot with a WHR of 0.9 and a SW of 100% would be perceived as male, whereas a robot with a WHR of 0.5 and 80% SW as female. The rationale behind this manipulation of genderedness came from the work of Johnson and Tassinari [32] and Lippa [37] who showed that people rely on WHR to judge a target's "gender" and that the form of the waist is a relevant feature for gender perception. Since the studies used static images, body proportions were not used in combination with other cues.

4.5 Manipulation Check

Only 54.3% of the studies ($N = 19$) performed *statistical* analyses to understand whether the manipulation of the robot's genderedness actually succeeded (see Figs. 3b and 5). On top of these studies, 8.6% of the studies ($N = 3$) performed a manipulation check but of a *non-statistical* nature [8, 72, 88] (see Figs. 3b and 5). The authors did ask participants which "gender" the robot belonged to in their opinion,

but they did not perform any statistical analysis to check for the significance of the result. As is easy to infer, 37.1% of the reviewed studies ($N = 13$) did not perform any manipulation check to test whether participants perceived the robot's genderedness as expected [13, 14, 30, 36, 44, 50, 59, 69–71, 74, 79, 86].

4.6 Assessment Tools

In the studies that performed a statistical manipulation check, the authors used three different approaches to assess people's attribution of "gender" to the robot (See Table 3 and Fig. 3c). The first measurement approach was *unidimensional*. The authors asked participants to rate the robot's genderedness on one item usually using the following phrase: *Rate the extent to which the robot appeared "rather male" versus "rather female"*. The rating was expressed on a 7-point Likert scale with *male* and *female* as end points. The second measurement approach was *multidimensional* (See Table 3 and Fig. 3c). The authors asked participants to fill out two items usually using the following phrasing: (1) *To what extent do you perceive the robot as male?* (2) *To what extent do you perceive the robot as female?*. The ratings were expressed on 7-point Likert scales where 1 meant *not at all* and 7 *extremely* [4, 5, 51, 55, 63, 81]. Finally, the third and last measurement approach was *nominal* (See Table 3 and Fig. 3c). The authors asked participants to select the "gender" of the robot among a list of options or as a write-in question [62, 63, 70]. Sandygulova and O'Hare used this approach with children using a pictorial response system [70]. Powers and Kielser [62] asked

participants to attribute a name to the robot and judged the “gender” attributed to the robot based on the gender of the name. Finally, Powers et al. [63] combined the multidimensional and nominal approaches by first asking whether the robot in their study was gendered and then asking participants to specify how feminine and masculine the gendered robot was.

When Likert scales were used to measure the robot’s genderedness (first and second approach), the mean scores on the items female/feminine and male/masculine were only rarely close to the end points of the corresponding gender. As an example, for Ghazali et al. [28], the manipulation check was significant. However, the difference between the male and female robot was not marked (male robot: $M = 5.50$, $SD = 1.60$; female robot $M = 6.07$, $SD = 0.83$). When the manipulation of the robot’s genderedness was performed with nominal scales (third approach), the difference between the robot’s “genders” was obviously more marked. However, female robots were more difficult to categorize across studies. This was particularly evident in [63] where the robot with the dampened female voice was miscategorized by 73% of the participants and given a male name by 70% of them.

Overall, 79% of the studies performing a statistical manipulation check ($N = 15$) were successful in manipulating the robot’s genderedness. Sixteen percent of them ($N = 4$) were only partially successful. Finally, 5% of them ($N = 1$) did not report the results of the statistical manipulation check [82] (see Table 3 and Fig. 5). The only instances where the manipulation check was only partially successful were the studies with a gender neutral or gender incongruent condition [33, 54, 55], or an altered gendered voice [62].

4.7 Results: Effects of Robot’s Genderedness (RQ2)

4.7.1 Methodological Note

The studies we reviewed employed 132 dependent variables. These could be nested into 17 groups based on conceptual similarity (e.g., warmth and mildness were nested under communion). For convenience, we refer to the group variables when reporting main and interaction effects. This grouping was merely done to clearly summarize the results and draw conclusions from them.

4.7.2 Main Effects

In the reviewed studies, only 17% of the dependent variables (22 dependent variables out of 132) were affected by the manipulation of the robot’s genderedness in terms of main effects. The genderedness of the robot did not yield any significant effect on the dependent variables nested under *competence* (10 dependent variables), *likability* (15 dependent variables), *credibility* (3 dependent variables),

acceptance (8 dependent variables), *task-related robot evaluations* (4 dependent variables), *proximity* (1 dependent variable), *closeness* (2 dependent variables), and “other” (2 dependent variables). Moreover, it had seldom main effects also on the dependent variables in the other groups.

When the results were significant, participants tended to perceive the robot in line with *gender stereotypes* (see Sect. 4.3.2). They attributed more communal traits to female robots than to male robots [5, 22] ([4] marginally significant) and more agentic traits to male robots than to female robots [22]. They showed higher affective trust towards female robots than towards male robots [4, 5], and rated the female robot as more suitable for stereotypical female tasks [4, 5, 22] and the male robot as more suitable for stereotypical male tasks [22]. Moreover, they donated more money [44, 74], said more words [63], and smiled more to female robots than to male robots [72]. The only studies that were counterintuitive in terms of gender stereotypes were Chita-Tegmark et al.’s [14] where, in contrast with the authors’ expectations, the male robot was perceived as more emotionally intelligent than the female one, and Bernotat et al.’s [4, 5], where, as opposed to the author’s assumptions, the female robot elicited more cognitive trust than the male robot.

Very few studies disclosed a significant main effect of the robot’s genderedness on *crucial HRI constructs* (see Sect. 4.3.2). In [33], the female robot was rated significantly higher in animacy and anxiety than the male one, and in [36], it was trusted significantly less. Interestingly, some of these studies report conflicting evidence. For instance, the male robot was perceived as more anthropomorphic than the female robot in [33], while it was perceived as more machine-like in [54].

4.7.3 Interaction Effects

The reviewed studies showed a significant interaction effect of the robot’s genderedness and (an)other independent variable(s) on 24.24% of the dependent variables (32 of the 132 dependent variables). Fifty percent of these effects resulted from the interaction between the robot’s genderedness and participant’s gender. The other half of these effects resulted from the interaction between the robot’s genderedness and a further independent variable (i.e., severity of moral infraction [30], interaction modality [54], type of emotion [13], childlikeness of the robot [62], stereotypically gendered task [35], or learning material [59]).

Robot’s Genderedness and Participant’s Gender. Among the studies that found an interaction effect between the robot’s genderedness and the participants’ gender, 50% (8 out of 16 dependent variables) showed a significantly positive effect of the *matching* between the robot’s genderedness and the participant’s gender, and 50% (8 out of 16) the opposite, a sig-

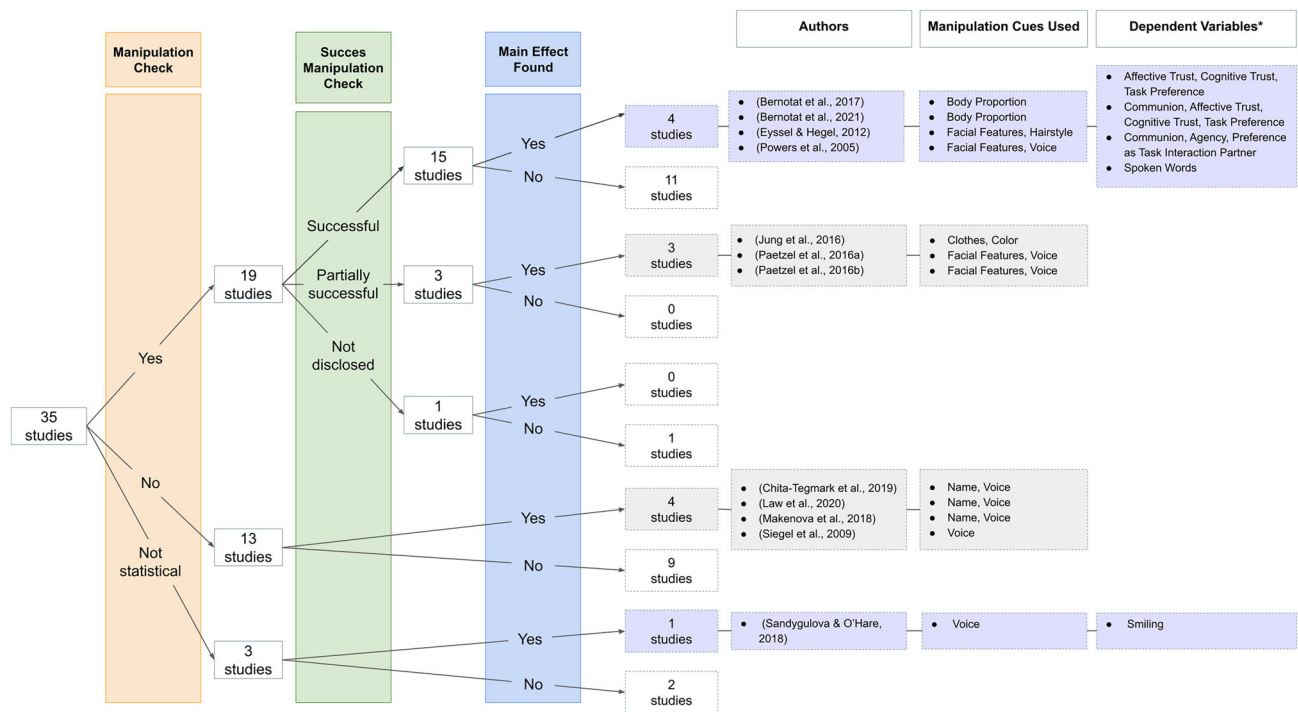


Fig. 5 Diagram summarizing the results of the scoping review. The orange column displays which of the included studies enlists a manipulation check, the green column shows how many of the studies performing a manipulation check actually succeeded in manipulating the robot’s genderedness, and the blue column highlights the studies finding a main effect of the robot’s genderedness on the dependent vari-

ables. The purple boxes on the right enlist the papers featuring main effect of gender on the dependent variables, the gender cues used when such effect was found, and the dependent variables influenced by robot’s genderedness. *= the dependent variables reported here are only those significantly affected by the robot’s genderedness

nificantly positive effect of the mismatch between the robot’s genderedness and the participant’s gender. With regards to the former results, adults seemed to perceive a robot with the same gender as them as significantly less harsh [30], more anthropomorphic [21], more psychologically close [21], and eliciting less negative cognition [28]. Further results disclosed that children were in a significantly better mood [72], smiled more [69], played more [71], and got more physically close [71] to a robot that shared the same gender as them, which lends support to the gender segregation hypothesis for cHRI. No evidence was found in support of the use of female robots as role models for women learning computer science topics [59].

With regards to the positive effect of a human-robot gender mismatch, women seemed to attribute higher emotional intelligence to male robots [14] and men found female robots more trustworthy [74], credible [86] (although [74] find this effect for both men and women), and engaging [74] and were willing to donate them more money [74]. Furthermore, men and women uttered more words to the robot of the opposite “gender” in [63], and younger children showed more happiness in the opposite gender than in the same gender condition in [72]. In general, the results of the studies exploring human-

robot gender (mis)match on the perception and interaction with robots are conclusive when it comes to children but inconclusive when it comes to adult participants.

Robot’s Genderedness and Further Independent Variables. Fifty percent of significant interaction effects were due to the joint effect of the robot’s genderedness and another independent variable (for an overview- of the results, see Table 4). Calvo-Barajas et al. [13] discovered that adolescents liked a female robot less when it expressed negative emotions, but they liked a male robot more when it expressed the same emotions. Similarly, Jackson et al. [30] disclosed that participants liked when a male robot rejected morally problematic requests, and they did so in several situations. However, male participants did not like when a female robot issued a strong rejection to a morally problematic request. In the same line, Powers and Kiesler [62] found out that all the participants in their study would follow the advice of a childlike male robot, whereas only half of them would follow the advice of an adultlike female robot.

On the opposite, Paetzel et al. [54] and Reich-Stiebert & Eyszel [66] revealed counter-stereotypical findings. In Paetzel et al., the female robot elicited more positive perceptions when it could express itself through multiple modalities,

Table 4 Summary of the significant interaction effects between robot's genderedness and further independent variables

Authors	Results
Calvo-Barajs et al. [13]	<p>The male robot was perceived as more likable in terms of appearance when it expressed high anger and low happiness (instead of medium anger, and low anger).</p> <p>The female robot was perceived as less likable in terms of appearance when it expressed high anger (instead of all other emotions: low, medium, and high happiness, and low and medium anger) and low anger and medium happiness (instead of low happiness).</p>
Jackson et al. [30]	<p>The male robot was perceived as too direct in the pre-test but not when responding to norm violating commands.</p> <p>Male participants liked male robots when rejecting commands from male humans for severe norm violations, but did not like female robots rejecting commands from female humans for weak norm violations.</p> <p>Male participants liked male robots but not female robots when they issued strong rejections.</p> <p>Female participants preferred when robots did not comply with the requests of a human with the same gender as the robot.</p>
Paetzel et al. [54]	<p>The female robot was perceived as more responsible, intelligent, pleasant, relaxed, and content than the male robot, but only in the multimodal condition (i.e., when the robot used both facial expressions and voice to interact).</p> <p>The male robot was perceived as more familiar and trustworthy than the female robot but only in the unimodal condition (when it used only facial expressions).</p>
Powers and Kiesler [62]	<p>100% of the participants said they would be willing to follow the advice of the childlike male robot.</p> <p>91% of the participants said they would be willing to follow the advice of the adultlike male and childlike female robots.</p> <p>Only 50% of the participants said they would be willing to follow the advice of the adultlike female robot.</p>
Reich-Stiebert and Eyssel [66]	<p>Participants who were instructed to solve a stereotypically male task with a female robot reported higher contact intentions with respect to participants involved in conditions where the genderedness of the task and the genderedness of the robot matched each other.</p>

whereas, in Reich-Stiebert & Eyssel, the participants interacting with a female robot in a stereotypical male task were more willing to interact again with the robot than the others.

5 Addendum: Papers 2021-2022

To conclude our Results section, we would like to report a short addendum on the studies manipulating the robot's genderedness between May 2021 and May 2022. To identify the studies in this addendum, we used the same search strings and databases detailed in Sect. 3.1 and followed the same selection pipeline discussed in Sect. 3.1.1. However, we did not perform the full process of coding and information extraction described in Sect. 3.2. The present section only aims at indicating the most recent developments in the investigation of robots' genderedness and highlighting whether novel results have been disclosed. The short review we performed returned 40 papers, of which 7 met the inclusion criteria after reading the abstract, and only 5 after reading the entire article [26, 48, 58, 61, 73]. In Table 5, we give more details about these papers.

Neuteboom and de Graaf (2021) [48] looked into the effects of robot's genderedness (female and male robot) and task (analytical and social) on the robot's perceived trustworthiness (i.e., capacity trust and moral trust), as well as on its social perception (i.e., agency and communion), and humanness (i.e., human uniqueness and human nature). In line with previous studies, they did not find any significant effect of robot's genderedness and performed task on people's perceptions.

Perugia et al (2021) [58], instead, explored how people attribute gender (femininity and masculinity) and stereotypical traits (communion and agency) to Furhat. Most Furhat's faces were attributed a "gender" in line with their names. Interestingly, the robot's genderedness influenced people's perceptions of the robot's agency but not of its communion. This study confirms that the robot's genderedness can influence the attribution of stereotypical traits to humanoid robots in agreement with [4, 5, 22].

The other three studies focused on the genderedness of service robots. Forgas-Coll et al. [26] investigated the effects of gender-personality congruity on customers' intention to use a service robot. They discovered that while the

congruous gender-personality robots (female-cooperative and male-competitive) did not differ from the incongruous ones (female-competitive and male-cooperative) in promoting intention to use, they did differ between each other: the female-cooperative robot performing significantly better than the male-competitive one in promoting intention to use.

With a slightly similar objective, Pitardi et al. (2022) [61] looked into the effects of matching robot's genderedness and participant's gender on people's perceived comfort and control in a service encounter, as well as on their brand attitude (i.e., positive and negative evaluations of the service provider). The study disclosed that human-robot gender congruity has a significant positive influence on perceived control and comfort, but not on brand attitude, and that the cultural value of masculinity mediates the effect of human-robot gender congruity on participant's perception of control.

Again in a service context, Seo (2022) [73] investigated the effects of robot's genderedness on pleasure and customer satisfaction in a service encounter and took into account the robot's anthropomorphism as an additional independent variable. The results showed that a female service robot leads to higher satisfaction and pleasure than a male service robot and that the robot's anthropomorphism plays a key role in positively influencing the results.

To sum up, the five studies in the addendum did not introduce novel ways to manipulate the genderedness of humanoid robots (except from personal titles, which can be equated to pronouns, see Table 5). In terms of results, however, they do disclose some interesting insights. They show a preference for female robots and human-robot gender congruity in service contexts [26, 61, 73]. Interestingly, they also reveal that values of masculinity play a role in this preference. It might be that service contexts are much more powerful than others in eliciting stereotypical knowledge of male and female roles, and especially so for those participants with more conservative views of gender.

6 Discussion

In the following, we are going to summarize the main findings of the literature review, answer the research questions, and identify gaps in the literature that warrant further attention. Then, we will discuss the results of the review and provide guidelines that the HRI community could follow when *gendering* or studying the gendering of robots. In doing so, we combine our epistemological backgrounds in Social Robotics and Gender Studies.

6.1 Summary of Results and Answers to RQ1 and RQ2

To summarize the results of the scoping review, **the HRI scholarship most often manipulated the robot's genderedness through voice, name, and facial features (RQ1)**. These cues were mostly used in interactive studies enlisting the use of a physical robot (see Fig. 4). In the majority of cases, the manipulation of the robot's genderedness with voice, name, and facial features yielded the expected results in terms of gendered perceptions (i.e., successful manipulation check). However, it often failed to produce a main effect of the robot genderedness on the dependent variables. Indeed, if we take a look at Fig. 4b and the purple boxes in Fig. 5, we realize that the most successful gender cues in influencing people's perceptions of robots were body proportions [4, 5], and facial features [22, 63]. If we pay close attention to the results of this scoping review, what becomes apparent is that the studies enlisting a significant main effect of the robot's genderedness on the dependent variables are predominantly picture-based (e.g., communion, agency, task preference). Moreover, we can see that, in these studies, **robot's genderedness is mostly successful in eliciting gender stereotypes of communion, agency and task preference/suitability, but does not yield notable significant effects on crucial HRI constructs, such as competence, likability, and acceptance (RQ2)**.

Given that robot's genderedness seems to be more harmful than useful as a design feature (it affects stereotyping but does not improve HRI), robotic companies might want to carry out user studies at different points of robot development to understand which perceptions the robots they are developing generates (e.g., in terms of stereotypes) and whether the cues they used to suggest "gender" (whether voluntarily or not) could have a role in prompting stereotypes. *Perugia et al. [56] have already started investigating which design cues in a robot are more likely to elicit stereotyping. However, more research in this direction is needed (GAP 1)*.

Given stereotypes towards gendered robots are prevalent but mostly studied with static images and in short-term studies, *future HRI research should also investigate if stereotype attribution is influenced by a robot's embodiment (GAP 2) and whether it changes over time (GAP 3)*. In a repeated interaction study, Paetzel et al. [53] discovered that participants develop stable perceptions of a robot's warmth and competence (concepts similar to communion and agency) after two minutes of interaction and do not update them over time. Longitudinal perceptual studies like Paetzel et al.'s are needed also in the context of gendered HRI, to disclose whether stereotypes are formed once and for all a few minutes after

Table 5 Details about the studies in the addendum: authors, cues used to manipulate the robot’s genderedness, and dependent variables (in **bold**, the significant main effects)

Authors	Cues used	Main effect
Neuteboom and de Graaf [48]	Name, Apparel	Trustworthiness, social perception, humanness
Perugia et al. [58]	Voice, Facial features	Agency , communion
Forgas-Coll et al. [26]	Voice, Facial features	Intention to use, usefulness, ease of use enjoyment, social influence, adaptiveness, sociability
Pitardi et al. [61]	Personal titles, color	Cultural values, perceived control, feelings of comfort, service brand attitude, familiarity
Seo et al. [73]	Name	Pleasure, satisfaction

meeting a robot or can modify with repeated interactions. In addition, since many studies focused on explicit stereotyping *it might be worth performing implicit bias studies* [52] *investigating people’s automatic, pre-reflective stereotyping of gendered robots (GAP 4)*. Finally, since the main concern of Roboethics and Robophilosophy is that people’s behaviors towards robots might eventually generalize to humans [27], *the HRI scholarship is in need of research paradigms and studies that explore whether and how gender stereotyping towards robots can influence people’s attitudes towards humans (GAP 5)*.

6.2 Discussion of Methodological Pitfalls

None of the studies we reviewed included non-binary, transgender, gender non-conforming, and gender fluid participants. Thirty-nine out of 3902 participants taking part in the reviewed studies (i.e., 1%) selected the option other/undisclosed. We can only assume that part of these participants identified with a gender falling outside of the binary. We consider the lack of gender-diverse participants a huge gap when studying the process of gendering robots, especially considering that the studies in this review brought to light the complex interweavings of participants’ gender and robot’s genderedness. This might have happened because participants’ gender is oftentimes asked with check-boxes providing only two options, “female” and “male”, but it might have also happened due to the lack of a proactive effort in including more gender identities. We advocate for this effort, hence we propose a first guideline for research on gendering robots:

Guideline 1: *Include transgender, gender fluid, gender non-conforming, and non-binary people, not just cisgender people, in the studies investigating robot’s genderedness.*

This guideline also urges to drop the biologized and essentialist way of asking about sex on a female/male categorical binary. The distinction of sex/gender and the deterministic understanding of sex as a binary biology is highly criti-

cized within the neuro- and biofeminist field [6]. Instead, understanding the terminology of the variety of gender identities that are actually relevant for social interaction as well as actively employing diverse recruiting efforts are needed. Scheuerman et al. drafted a living document “[HCI Guidelines for Gender Equity and Inclusivity](#)” containing a section on gender inclusive research methods which gives valuable insights into how to perform inclusive research. For instance, they suggest using the following options to ask about participants’ gender: woman, man, non-binary, prefer not to disclose, prefer to self-describe and explain how to carry out in-person studies in a way that is respectful of all gender identities (see also [78]).

The studies we reviewed not only lacked of heterogeneity in terms of participant’s gender, but also often omitted a definition of “gender”. Only Bryant et al. [8] attempted a description of the gendering process as related to robots, and Rea et al. [64] and Reich-Stiebert and Eyssel [66] provided a definition of the terms sex and gender as referred to participants. Given that people interpret “gender” in many different ways (e.g., some conflate it with sex), providing a definition of human gender and robot “gender” in papers focusing on the manipulation of the robot’s genderedness could help circumvent any confusion deriving from the word’s polysemy, as well as improve a paper’s clarity and generalizability. A practical way to do so is for authors to reflect on their understanding and experience of “gender” and how this is translated into gendered robots. According to the definitions included in this scoping review, for instance, human sex refers “to biological and physiological features” [66], whereas human gender is “a social construction” [66]. These definitions, provided by Reich-Stiebert and Eyssel, are in line with the definitions of sex and gender of the American Psychological Association (APA) [2]. Robot’s “gender”, instead, is defined by Bryant et al. as the result of “robot gendering, the attribution of gender onto a robotic platform via voice, name, physique, or other features” [8]. While we consider this definition absolutely fitting, we deem it incomplete as it only depicts the designer’s side of the gendering process and overlooks the participant’s side, as to say the way participants

attribute gender to the robot as a result of its “voice, name, physique and other features” (see *decoding* in Sect. 1.3). We do not advocate for a universally fixed definition of gender that could fit all research and researchers. However, we think it is important to:

Guideline 2: *Provide a definition of human gender and clarify what is meant with robot “gender” to avoid giving rise to opaque interpretations of the paper’s results and implications.*

Another methodological pitfall we observed in some of the studies, which is unfortunately endemic to HRI research, is the uniformity of participants’ characteristics. Most of the reviewed studies resorted to a sample of young participants (probably university students). The main drawback of the homogeneity in participants’ characteristics is that it makes it difficult to address context- and user-specific differences. We acknowledge that resorting to students as participants is oftentimes dictated by the research complexity or by the lack of funding to recruit a more diverse set of participants. However, in the specific context of gendering robots, this might give one-sided results, as individual participants’ characteristics might disclose relevant insights into how gendered robots are perceived. For instance, Bernotat et al. disclosed a significant effect of people’s benevolent sexism and tendency to act in a social desirable way on their tendency to stereotype gendered robots [5], a result confirmed by additional studies [29, 56]. Individual characteristics are more likely to differ and bring meaningful results in heterogeneous participants’ samples. While we put forward a caveat in this sense, we do not feel like enforcing a guideline, as the use of university students as participants might depend on the economic availability of each research group.

From a methodological perspective, we need to mention another aspect we observed in the reviewed studies, which might constitute a limitation to the generalizability of this review, namely the richness of robots, tasks, and activities. The studies we reviewed used many different robotic platforms and envisioned many different tasks (e.g., observing pictures, watching videos, interacting with the robot), activities (e.g., educational activities, casual conversations) and participants’ roles (e.g., remote observer, co-present observer, interactant). This complexity is not bad in principle, but is risky when not followed by replication as it makes comparability and generalizability difficult, thus hindering the possibility of drawing conclusions on the role of robot’s genderedness as a whole. Except for Siegel et al.’s paradigm [74], consisting of a robot introducing itself or a research project and asking for donations, which has been replicated by You and Lin [86] and Makenova et al. [44], and Eyssel and Hegel’s study [22], in which participants were shown a picture of a robot and were asked to evaluate it in terms of

stereotypical traits and tasks, which was replicated by Bernotat et al. [4, 5], the many studies we presented in this review were very rarely replicated and further explored. Hence, we would like to suggest the HRI scholarship to:

Guideline 3: *Perform replication studies where existing experimental designs and activities on gendered robots are incrementally modified (e.g., change robotic platform or gender manipulation) to check if results still hold.*

6.3 Discussion on Manipulation of Robot’s Genderedness (RQ1)

Through this scoping review, we discovered that the robot’s genderedness has been manipulated by the HRI scholarship using cues such as the robot’s voice, name, facial features, apparel, colors, body proportions, and hairstyle. Some of these cues are fruit of social conventions and socio-cultural schemata (e.g., names, hairstyle, apparel), others refer to the physical and physiological characteristics of gendered bodies (e.g., the waist-to-hips ratio and the voice frequency). Nevertheless, most of them tap into a binary understanding of gender. Indeed, in 89% of the reviewed studies, the gendering of the robot has been manipulated within the female/male binary. As a result, we draw the following guideline:

Guideline 4: *Include gender neutral or gender ambiguous robots in the studies to understand whether less binary gendering is possible or even meaningful.*

While it is clear from our review, that the studies attempting a manipulation of gender neutrality [8, 54] and gender ambiguity in robots [55] led to non-significant manipulation checks, we nonetheless encourage researchers to investigate whether and how it would be possible to design gender expressions for robots that go beyond the binary and what these gender expressions entail in terms of robot’s perception. In their paper, Paetzel et al. [55] were close to successfully manipulate gender neutrality by providing the Furhat robot with a face whose gender did not match the gender of its voice. This is an important result as it shows that multimodality can be exploited to obtain more diverse robot designs.

As a non-negligible aspect of the gendering process observed in the reviewed studies, most of the gender cues were used in combination with others and only rarely in isolation, as if the layering of these cues could strengthen the gender attribution. However, overdoing gender cues and/or using extremely stereotypical cues—like the pink ribbon/blue bow-tie in Thellman et al. [82], the pink earmuffs in Jung et al. [33], or the flower hair clip/bow tie in Zhumabekova et al. [88]—might make the gender manipulation too obvious, thus revealing the purpose of the study. From the results of the manipulation checks in the reviewed studies,

it is apparent that gender is attributed to robots on the basis of the tiniest gender cues. As an example, Rea et al. [64] managed to manipulate the robot's genderedness only with pronouns and the robot's voice. Besides, in a study not featured in this review, Perugia et al. discovered that the higher the number of appearance cues used to suggest gender in a robot, the higher the consequent stereotyping (especially for female robots) [56]. Hence, unless specifically motivated by the research questions and experimental design (e.g., intention to study the effects of stereotypical gender designs), we suggest the HRI scholarship to:

Guideline 5: *Avoid overdoing gender cues and use as little gender cues as possible, and as subtle gender cues as feasible, when manipulating a robot's genderedness as an experimental variable.*

On this note, Paetzel et al. also showed how the head of the Furhat robot alone, without any projected face on it, already leads to gender attributions [55]. Since the same authors showed how multimodal cues can change people's attribution of gender to the robot, the gender attributed to the robot at baseline (without any gender cues added) can be highly influential in determining the final gender attribution (when the additional gender cues are implemented, e.g., the robot's voice and name). Hence, it is worth to:

Guideline 6: *Perform a pre-test of the genderedness of the robotic platform one plans to use and account for it when discussing the robot gender manipulation.*

Tools like the humanoid ROBOTS—Gender and Age Perception (ROBO-GAP) dataset [57]—which provide ratings of perceived femininity, masculinity, and gender neutrality for all the 251 robots in the Anthropomorphic roBOT (ABOT) dataset [60]—can come in handy in this context as they can help researchers checking the gender attributed to the robot they plan to use at baseline.

Another striking result of this scoping review was that almost half of the studies did not perform any statistical analysis to assess whether the manipulation of the robot's genderedness actually succeeded. This is particularly problematic as it makes it difficult to establish whether the lack of significant effects of the robot's genderedness on the dependent variables is actually due to the robot's genderedness or to other lurking variables. Future studies should:

Guideline 7: *Always perform a manipulation check to test whether the robot's genderedness is perceived by participants in the expected way.*

Measuring the robots' genderedness is not exempt from shortcomings. A research concept is necessarily entangled with the questionnaire that asks the participant about it [38]. Meaning, if the concept is a binary understanding of gender,

then a question about feminine or masculine aspects in one or different items, will ontologically reproduce a binary idea of gender. Besides, asking people to attribute gender to a robot might result in a gender attribution even when the robot is not perceived as gendered in the first place.

In this scoping review, we identified several ways to measure the robot's genderedness. One way to measure the robot's genderedness in a sound way is to use multidimensional assessment tools (see Sect. 4.6) as in [4, 5, 51, 55, 81]. In multidimensional approaches, different gender dimensions are employed to assess the perceived gender of the robot (e.g., feminine and masculine), hence participants are free to rate a robot as predominantly masculine, while at the same time recognizing in it some feminine characteristics, but they can also rate the robot as high in both femininity and masculinity. Unidimensional assessment tools, instead, directly tap into a binary understanding of gender and force participants to choose between two gender categories visually represented as opposite, masculine/male and femininity/female, while some robot designs might have features of both (e.g., Pepper). Besides, in these scales, it is unclear what the midpoint means. In a 7-point Likert scale with 1-male and 7-female as end points, what does 4 stand for? Some participants might interpret the midpoint as gender neutral, others as gender ambiguous, and this might lead to unreliable ratings. Another sound, but more qualitative, assessment tool is proposed by Powers & Kiesler [62], who asked participants to give a name to the robot and inferred the perceived gender based on it. This is an interesting approach as it explores the process of gender attribution in a more implicit way and gives participants the possibility to not just give traditional names to robots, but also more technical and object-oriented ones [68]. However, this approach might fall short if studies involve participants with different nationalities, as naming conventions might change across countries (e.g., Simone is a male name in Italian but a female name in German).

On top of the assessment tools used, it is always important to check whether significant differences in the perceived gender of the robot actually represent differences in attributed gender. In Ghazali et al. [28], the manipulation check is deemed successful since the female and male robot conditions significantly differ in terms of ratings. However, when taking a look at the descriptive statistics reported by the authors, the robot's perceived genderedness did not differ in terms of gender. The authors adopted a unidimensional assessment scale spanning from masculine (1) to feminine (7). The female robot had a mean gender rating of 6.07, while the male robot a mean gender rating of 5.50, thus indicating that both robots fell on the feminine side of the Likert scale. Based on this, we recommend researchers to perform a manipulation check, but also:

Guideline 8: *Check the descriptive statistics of each gender condition as part of the manipulation check, as a significant difference between conditions does not necessarily grant a different categorization of the robot's genderedness.*

6.4 Discussion on Effects of Robot's Genderedness on Perceptions of and Interactions with Robots (RQ2)

When taking the results as whole, it becomes quite clear that gendering robots has a strong effect on stereotyping. We cannot help but wonder whether the effects that robot's genderedness has on stereotyping might have been due to the way the robot was gendered in the first place. As to say, if we imbue robots with stereotypical gender cues, it might become difficult for participants to not stereotype them as a result.

In general, one of the clear-cut outcomes of this scoping review is that genderedness does not have an effect on crucial constructs for the HRI, such as acceptance and likability, as it perhaps does for voice assistants. In this regard, however, the studies published between 2021 and 2022 paint a different picture. They disclose that in service contexts, female robots and gender "congruity" (i.e., the match between participant's gender and robot's genderedness) are almost always preferred. Comparing these results with the research on voice assistants, it seems that there is something in the service context that makes the female genderedness of artificial agents immediately relevant. As if the fact that we as humans are used to see women in service roles makes the suitability of female robots in the same role immediately glaring. From a feminist standpoint, a question arises: do we have to second the preference of the user for female service robots even if we know it stems from a discriminatory understanding of a gendered society? We as authors argue that we do not have to, and present the HRI community with a guideline that could serve as a design opportunity:

Guideline 9: *Use gendered robots to offer occasions of defamiliarization with normative gender roles and disrupt binary conceptualizations of human gender and tasks.*

In the context of interaction effects, two results caught our attention in the papers we reviewed. Calvo-Barajas et al. [13] discovered that children perceived a female robot as less likable when it expresses high anger instead of more positive or less intense emotions, while Jackson et al. [30] disclosed that male participants like male robots but not female robots when they issue strong rejections. These results seem to suggest that female robots, like women, are liked less when they are not compliant or not consensual. This follows the problematic narrative that wants women submissive and aware of "their place" in the world. In a real-life environment, how should a female robot react to people issuing annoyance for their lack of compliance or consent? Should they maintain a jokey vibe of servitude as voice assistants originally did [84] or react resolutely as in Winkle et al. [85]? We consider Winkle et al.'s work [85] a valid and viable option. Aside from this, however, the HRI scholarship should start reflecting on the ethical implications of gendered robots and their (mis)treatment, especially given the highly symbolic meaning human-humanoid interactions entertain with human-human interactions [57, 76, 77, 87]. As such we suggest a last guideline:

Guideline 10: *Critically reflect on the results of your research on gendered robots and engage with a discussion of the ethical implications of your findings, especially considering the highly symbolic value of human-humanoid interactions for human-human relations.*

For future robot designs, the challenge remains whether we could come close to a gender neutral or even genderless humanoid robot, and whether this would help to circumvent gender biases and stereotypes. As authors, we think that gendering robots is not a problematic process *per se*, it is the way robots are gendered following normative and binary views of female and male gender that is problematic. As such, we urge roboticists to shake binary and normative views of gender from the core, and identify more inclusive and less stereotyped configurations of gender in robotics that do not reinforce, borrowing Balsamo's words [3], traditional narratives about the gendered body.

Table 6 Experimental information about the studies included in the scoping review: authors (Date), independent variables (*bs* = between subjects; *ws* = within subjects), dependent variables (in bold), the significant main effects of robot’s genderedness on the dependent variables, i.e., $p < .05$, and summary of findings

Author (date)	Variables		Dependent*	Summary
	Independent	Dependent		
Bernotat et al. (2017) [4]	<i>bs</i> <i>Robot Genderedness</i> (female, male)	affective trust, cognitive trust, task preference	communion, agency, affective trust, cognitive trust, task preference	The female robot was perceived as more suitable for stereotypically female tasks and induced more affective and cognitive trust than the male robot.
Bernotat et al. (2021) [5]	<i>bs</i> <i>Robot Genderedness</i> (female, male)	communion, agency, affective trust, cognitive trust, task preference	communion, agency, affective trust, cognitive trust, task preference	The female robot was perceived as more communal and suitable for stereotypically female tasks and induced more affective and cognitive trust than the male robot.
Bryant et al. (2020) [8]	<i>bs</i> <i>Robot Genderedness</i> (female, male, neutral)	occupational competency, trust in occupational competency	occupational competency, trust in occupational competency	The perceived gender of the robot did not have an impact on perception of occupational competency nor on trust in the robot’s occupational competency.
Calvo-Barajas et al. (2020) [13]	<i>ws</i> <i>bs</i> <i>bs</i> <i>Robot Genderedness</i> (female, male) Emotion Type (happiness, anger) Emotion Intensity (low, medium, high)	competence (smartness, helpfulness), likability (appearance, friendliness), trust (keeping secrets, telling the truth, exerting influence, being trustworthy)	competence (smartness, helpfulness), likability (appearance, friendliness), trust (keeping secrets, telling the truth, exerting influence, being trustworthy)	The robot’s genderedness did not influence the perception of competence, trust, and likability of the robot. However, it interacted with emotion intensity to affect the robot’s perceived likability (appearance).
Chिता-Tegmark et al. (2019) [14]	<i>bs</i> <i>bs</i> <i>bs</i> <i>bs</i> <i>ws</i> <i>Robot Genderedness</i> (female, male) Agent Type (human, robot) Participant Gender (female, male) Emotional Intelligence (low, high)	emotional intelligence	emotional intelligence	Male agents (human/robot) were perceived to have higher levels of emotional intelligence (EI) than female agents. Female participants perceived female agents as having lower EI than male agents in the low EI condition.
Eyssel & Hegel (2012) [22]	<i>bs</i> <i>bs</i> <i>bs</i> <i>ws</i> <i>Robot Genderedness</i> (female, male)	communion, agency, preference as task interaction partner	communion, agency, preference as task interaction partner	Participants ascribed more agentic traits to the male robot and more communal traits to the female robot. The male robot was rated as more suitable for stereotypical male tasks and the female robot was evaluated as more suitable for stereotypically female tasks.
Eyssel et al. (2012a) [21]	<i>bs</i> <i>bs</i> <i>bs</i> <i>Robot Genderedness</i> (female, male) Participant Gender (female, male) Voice Type (humanlike, robotlike)	anthropomorphic inferences (attribution of mind), psychological closeness, acceptance	anthropomorphic inferences (attribution of mind), psychological closeness, acceptance	Participants revealed greater HRI acceptance and felt more psychological closeness for the same-gender robot compared to the opposite-gender robot. Participants anthropomorphized the same-gender robot more strongly than the opposite-gender robot.
Eyssel et al. (2012b) [23]	<i>bs</i> <i>bs</i> <i>bs</i> <i>Robot Genderedness</i> (female, male) Participant Gender (female, male) Voice Type (humanlike, robotlike)	likability (warmheartedness), anthropomorphism (human essence), contact intentions	likability (warmheartedness), anthropomorphism (human essence), contact intentions	Participants perceived the same-gender robot as more likable and psychologically close than the opposite-gender robot.
Ghazali et al. (2018) [28]	<i>bs</i> <i>bs</i> <i>Human-Robot Gender Match</i> (mismatch, match) Trust Face (least, most trustworthy)	perceived intelligence, anthropomorphism, trust beliefs (trust, perceived trust, individualized trust), trusting behaviors, feeling of anger, negative cognitions	perceived intelligence, anthropomorphism, trust beliefs (trust, perceived trust, individualized trust), trusting behaviors, feeling of anger, negative cognitions	Psychological reactance was lower towards the same-gender robots compared to the opposite-gender robots.

Table 6 continued

Author (date)	Independent	Variables	Dependent*	Summary
Jackson et al. (2020) [30]	bs bs bs ws ws	<i>Robot Genderedness</i> (female, male) Participant Gender (female, male) Human Requester Gender (female, male) Severity Moral Infraction (low, high) Face Threat Robot Response (low,high)	politeness, harshness, directness, likability	There was no significant main effect of the robot's genderedness on the dependent variables, but many significant interaction effects between robot's genderedness and the other independent variables. These interaction effects seemed to point to the fact that male robots are perceived more favorably than female robots when rejecting commands.
Jung et al. (2016) [33]	bs bs	<i>Robot Genderedness</i> (female, male) Cue Location (body, screen)	anthropomorphism, anxiety, animacy	The male robot was perceived as more anthropomorphic, more animate and less anxious than the female robot regardless of the gender cue location.
Kraus et al. (2019) [34]	bs bs ws	<i>Robot Explicit Genderedness</i> (female, male) Robot Personality (female, male) Task Scenario (female, male)	competence, likability, trust, reliability, acceptance, predictability	The male personality was perceived as significantly more trustworthy and predictable in the taxi ordering scenario. The female personality was perceived as more likable in both scenarios.
Kuchenbrandt et al. (2014) [35]	bs bs bs	<i>Robot Genderedness</i> (female, male) Gender Typicality Task (female, male) Participant Gender (female, male)	task-related competence, follow directions, acceptance, speed of completion, errors	In the female task context, participants made more errors, were less willing to accept help from the robot on a future task, and anthropomorphized the robot to a smaller degree than in the male task.
Law et al. (2020) [36]	bs bs bs	<i>Robot Genderedness</i> (female, male) Emotional Intelligence (low, high) Presentation Style (text, voice)	trust	Robots in the high emotional intelligence (EI) condition were trusted more than the robots in the low EI condition. Male robots were trusted more than female robots.
Law et al. (2020) [36]	bs bs	<i>Robot Genderedness</i> (female, male) Performance-based Trust (low, high)	trust	The robot's genderedness did not affect participants' trust.
Lugrin et al. (2020) [42]	bs bs	Presentation Style (text, voice) <i>Robot Genderedness</i> (female, male) Language Type (standard, accent, dialect)	perceived competence, likability, social skills	The robot's genderedness did not affect people's perception of the robot's competence, likability and social skills.
Makenova et al. (2018) [44]	bs bs	<i>Robot Genderedness</i> (female, male) Participant Gender (female, male)	trust, donations , credibility, engagement	The female robot received more donations from participants compared to the male robot.
Nomura & Kinoshita (2015) [50]*	bs	<i>Robot Genderedness</i> (female, male)	20 pairs of adjectives, only 4 specified: pretty-hateful, familiar-unfamiliar , mild-terrible, pleasant-implesant	The female robot was perceived as prettier and more familiar than the male robot.
Nomura & Takagi (2011) [51]	bs bs bs	<i>Robot Genderedness</i> (female, male) Participant Gender (female, male) Participant Background (tech., social sciences)	politeness, mildness, ambitiousness, assertiveness	The robot's genderedness did not have a significant effect on politeness, mildness, ambitiousness, assertiveness.
Paetzel et al. (2016a) [54]	bs bs	<i>Robot Genderedness</i> (female, male) Modality (unimodal, multimodal)	intelligence, likability, uncanniness, anthropomorphism , trust, relaxation, contentment	The male robot was perceived as less anthropomorphic than the female robot, and especially so in the multimodal condition. The female robot was perceived as more intelligent, likable for entertaining a conversation, pleasant, relaxed and content than the male robot in the multimodal condition.

Table 6 continued

Author (date)	Independent	Variables	Dependent*	Summary
Paetzl et al. (2016b) [55]	bs bs	<i>Genderedness Robot Face</i> (female, male) <i>Genderedness Robot Voice</i> (female,male)	likability, uncanniness, anthropomorphism, trust	On the opposite, the male robot was perceived as significantly more familiar and trustworthy than the female robot in the unimodal condition. The robot's genderedness did not affect the perception of pleasantness (i.e., likability); strangeness (i.e., uncanniness), anthropomorphism. The congruent female robot was perceived as more trustworthy than the congruent and incongruent male robots.
Pfeifer & Lugrin (2018) [59]	bs bs	<i>Robot Genderedness</i> (female, male) Learning Material (female, male)	objective learning success	Robot's genderedness did not have any significant main effect on learning success. Participants who took part in the learning activity with a female robot explaining male learning materials obtained the highest learning success.
Powers & Kiesler (2006) [62]	bs bs bs	<i>Robot Voice Genderedness</i> (female, male, female dampened, male dampened) chin height (short, long) forehead height (short, long)	following robot's advice (<i>not the only dependent variable but the only one used to test differences across conditions</i>)	100% of the participants who saw the robot with a short chin (childlike) and heard the robot speak in the undampened male voice said they would take the robot's advice. When the chin was long (adult) and had a male voice, or the voice was female and the chin was short, the percent went down to 91%. only 50% of the participants who saw the robot with a long chin and heard it speak with a female voice said they would take its advice.
Powers et al. (2005) [63]	bs bs ws	<i>Robot Genderedness</i> (female, male) Participant Gender (female, male) Target of Conversation (female, male)	compassion, dominance, knowledge, likability, personality, spoken words	Participants said more words to the male robot than to the female robot. Men uttered more words than women in the female robot condition, whereas women said more words than men in the male robot condition.
Rea et al. (2015) [64]	bs bs	<i>Robot Genderedness</i> (female, male) Participant Gender (female, male)	communion, politeness, agency, engagement, task preference, relaxation	Female participants showed more politeness towards the robot than male participants. Participants did not apply gender stereotypes to the robots based on their genderedness. The robot's genderedness did not affect engagement.
Reich-Stiebert & Eyssele (2017) [66]	bs bs bs	<i>Robot Genderedness</i> (female, male) Participant Gender (female, male) Task Gender Typicality (female, male)	warmth, agency, competence, willingness to engage, obj. learning outcome, subj. learning outcome, mood, intrinsic motivation, perception as learning companion, satisfaction	Participants showed not to apply gender stereotypes to robots when interacting in gender stereotypical tasks. A mismatch between robot's genderedness and gender typicality of the task led to increased willingness for future learning with the robot.
Sandygulova & O'Hare (2018) [72]	ws bs bs	<i>Robot Gendered Voice</i> (female, male) Children Gender (female, male) Children Age (5-8 yo, 9-12 yo)	likability, robot type, robot as social actor, mood, valenced affective state, happiness, smiling	Young children preferred to interact with a same-gender robot compared to a opposite-gender robot, whereas older children reported no difference in their preference. Children smiled more during the interaction with the female robot than during the interaction with the male robot.
Sandygulova & O'Hare (2016) [71]	bs	Human-Robot Gender Match (mismatch, match)	duration of interaction, proximity	Children spent more time with the same-gender robot and their proximity with the robot decreased or remained stable when interacting with a male robot and increased when interacting with a female robot.

Table 6 continued

Author (date)	Variables		Summary
	Independent	Dependent*	
Sandygulova & O'Hare (2015) [70]	ws <i>Robot Voice Genderedness</i> (female UK, male UK, female US, male US)	voice preference	Children preferred the robot with the English UK accent over the robot with the English US accent. No effect of gender on voice preference is reported.
Sandygulova et al. (2014) [69]	bs <i>Robot Voice Genderedness</i> and Age (female child, male child, female adult) Children Age (8, 9, 10, 11, 12 yo)	happiness	Female children expressed more happiness towards the female robot. Male children expressed more happiness towards the male robot
Siegel et al. (2009) [74]	bs <i>Robot Genderedness</i> (female, male) Participant Gender (female, male) Participant alone (not alone, alone)	trust, donations , credibility, engagement	Men trusted the female robot more, donated more money to it, considered it more credible, and felt more engagement with it. Women showed no preference for any of the robots in terms of trust and donations, but rated the male robot as more credible and more engaging than the female robot.
Tay et al. (2014) [81]	bs <i>Robot genderedness</i> (female, male) Robot Personality (introverted, extroverted) Occupational role (female, male)	perceived trust, acceptance, attitude towards robots, subjective norms, affective evaluation, cognitive control, perceived behavioral control	The perceived gender of the robot did not affect perceived trust, acceptance, attitude towards the robot, subjective norms, cognitive evaluations. Participants showed higher affective evaluations and perceived behavioral control in the female healthcare robot condition, and higher affective evaluations in the male security robot condition. Some of the non-significant results were marginally significant.
Thellman et al. (2018) [82]	bs <i>Robot Genderedness</i> (female, male) Participant Gender (female, male)	persuasiveness	The robot's genderedness did not have a significant effect on persuasiveness, nor did participants' gender.
You & Lin (2019) [86]	bs <i>Robot Voice Genderedness</i> (female, male) <i>Robot Appearance Genderedness</i> (female, male, neutral)	trust, donations, credibility, engagement	The genderedness of the robot did not affect trust, the amount of donations, credibility, and engagement. While men rated the female robot as more credible (safety) than the male robot, women did not show any difference in their ratings.
Zhumabekova et al. (2018) [88]	ws <i>Robot Genderedness</i> (female, male) Participant Gender (female, male)	liked interacting with robot	Children liked the interaction with a same-gender robot more than the interaction with a robot with a different gender.
Steinhaeusser et al. (2021) [79]	bs <i>Robot Voice Genderedness</i> (female, male, neutral) Participant Gender (female, male)	anthropomorphism, transportation, attitudes towards robots	There was no significant effect of robot's genderedness on anthropomorphism, the attitude towards robots, and the transportation participants felt when the robot told the story, nor were there any interaction effects between participant's gender and robot's genderedness

Acknowledgements The authors would like to thank Laura van der Bij for performing the strings search for the papers in the addendum, and Latisha Boor for checking the correctness of Tables 1 and 2, contributing to the creation of Table 4, and designing the visualizations in Figures 2, 3, 4, and 5.

Author Contributions GP formulated the research questions, devised the inclusion and exclusion criteria, performed the string search, went through the selection pipeline, read all the papers, extracted the information from the papers, summarized all the results, wrote sections 2, 3, 4, 5, and 6 of the paper, contributed to the writing of section 1, and prepared all the tables. DL formulated the research questions, devised the inclusion and exclusion criteria, performed the string search, read part of the papers, extracted information from part of the papers, wrote section 1 of the paper and contributed to the writing of section 6.

Funding GP's work is partly funded by the research programme Ethics of Socially Disruptive Technologies (ESDiT), which is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031). DL's work is funded by the Wallenberg AI, Autonomous Systems and Software Program-Humanities and Society (WASP-HS) project "The Ethics and Social consequences of AI and caring Robots."

Declarations

Conflict of interest The authors declare that they have no conflict of interest or competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alesich S, Rigby M (2017) Gendered robots: implications for our humanoid future. *IEEE Technol Soc Mag* 36(2):50–59
- Association AP, et al (2021) Inclusive language guidelines
- Balsamo AM et al (1996) *Technologies of the gendered body: reading cyborg women*. Duke University Press, New York
- Bernotat J, Eyssel F, Sachse J (2017) Shape it—the influence of robot body shape on gender perception in robots. In: *International conference on social robotics*. Springer, pp 75–84
- Bernotat J, Eyssel F, Sachse J (2021) The (fe)male robot: how robot body shape impacts first impressions and trust towards robots. *Int J Soc Robot* 13(3):477–489
- Bluhm R, Maibom HL, Jacobson AJ (2012) *Neurofeminism: issues at the intersection of feminist theory and cognitive science*. Springer, Berlin
- Brown E, Perrett DI (1993) What gives a face its gender? *Perception* 22(7):829–840
- Bryant D, Borenstein J, Howard A (2020) Why should we gender? The effect of robot gendering and occupational stereotypes on human trust and perceived competency. In: *Proceedings of the 2020 ACM/IEEE international conference on human–robot interaction*, pp 13–21
- Burton AM, Bruce V, Dench N (1993) What's the difference between men and women? evidence from facial measurement. *Perception* 22(2):153–176
- Butler J (1988) *Performative acts and gender constitution: an essay in phenomenology and feminist theory*. *Theater J* 40:519–531
- Butler J (1994) *Against proper objects: introduction*. *Differ J Fem Cult Stud* 6(2):3
- Butler J (1999) *Gender trouble: feminism and the subversion of identity*. Routledge, London
- Calvo-Barajas N, Perugia G, Castellano G (2020) The effects of robot's facial expressions on children's first impressions of trustworthiness. In: *2020 29th IEEE international conference on robot and human interactive communication (RO-MAN)*. IEEE, pp 165–171
- Chita-Tegmark M, Lohani M, Scheutz M (2019) Gender effects in perceptions of robots and humans with varying emotional intelligence. In: *2019 14th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, pp 230–238
- Cockburn C (1992) The circuit of technology: gender, identity and power. In: *Consuming technologies: media and information in domestic spaces*, pp 33–42
- de Graaf M, Perugia G, Fosch-Villaronga E, Lim A, Broz F, Short E.S, Neerinx M (2022) Inclusive hri: equity and diversity in design, application, methods, and community. In: *Proceedings of the 2022 ACM/IEEE international conference on human–robot interaction*, pp 1247–1249
- Denner J, Werner L, Bean S, Campe S (2005) The girls creating games program: strategies for engaging middle-school girls in information technology. *Front A J Women Stud* 26(1):90–98
- D'Ignazio C, Klein LF (2020) *Data feminism*. MIT press, New York
- Ehrberger K, Räsänen M, Börjesson E, Hertz AC, Sundbom C (2017) The androchair: performing gynaecology through the practice of gender critical design. *Des J* 20(2):181–198
- Ehrberger K, Räsänen M, Ilstedt S (2012) Visualising gender norms in design: meet the mega hurricane mixer and the drill dolphin. *Int J Des* 6(3)
- Eyssel F, De Ruyter L, Kuchenbrandt D, Bobinger S, Hegel F (2012) 'if you sound like me, you must be more human': on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In: *2012 7th ACM/IEEE international conference on human–robot interaction (HRI)*. IEEE, pp 125–126
- Eyssel F, Hegel F (2012) (s) he's got the look: gender stereotyping of robots 1. *J Appl Soc Psychol* 42(9):2213–2230
- Eyssel F, Kuchenbrandt D, Hegel F, De Ruyter L (2012) Activating elicited agent knowledge: how robot and user features shape the perception of social robots. In: *2012 IEEE RO-MAN: The 21st IEEE international symposium on robot and human interactive communication*. IEEE, pp 851–857
- Faulkner W (2001) The technology question in feminism: a view from feminist technology studies. In: *Women's studies international forum*, vol 24. Elsevier, pp 79–95
- Fine C, Rush E (2018) "why does all the girls have to buy pink stuff?" the ethics and science of the gendered toy marketing debate. *J Bus Ethics* 149(4):769–784
- Forgas-Coll S, Huertas-Garcia R, Andriella A, Alenyà G (2022) The effects of gender and personality of robot assistants on customers' acceptance of their service. *Serv Bus* 1–31
- Friedman MTD, Puzio A, Babushkina D, Löhr G, Gwagwa A, Kamphorst BA, Perugia G, IJsselsteijn W (2023) 3. social robots

- and society. In: Ethics of socially disruptive technologies: an introduction
28. Ghazali AS, Ham J, Barakova EI, Markopoulos P (2018) Effects of robot facial characteristics and gender in persuasive human–robot interaction. *Front Robot AI* 5:73
 29. Guidi S, Boor L, van der Bij L, Foppen R, Rikmenspoel O, Perugia G (2023) Ambivalent stereotypes towards gendered robots: the (im)mutability of bias towards female and neutral robots. In: Social robotics: 14th international conference, ICSR 2022, Florence, Italy, December 13–16, 2022, proceedings, part II. Springer, pp 615–626
 30. Jackson RB, Williams T, Smith N (2020) Exploring the role of gender in perceptions of robotic noncompliance. In: Proceedings of the 2020 ACM/IEEE international conference on human–robot interaction, pp 559–567
 31. Johnson E (2008) Simulating medical patients and practices: bodies and the construction of valid medical simulators. *Body Soc* 14(3):105–128
 32. Johnson KL, Tassinari LG (2005) Perceiving sex directly and indirectly: meaning in motion and morphology. *Psychol Sci* 16(11):890–897
 33. Jung EH, Waddell TF, Sundar SS (2016) Feminizing robots: user responses to gender cues on robot body and screen. In: Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems, pp 3107–3113
 34. Kraus M, Kraus J, Baumann M, Minker W (2018) Effects of gender stereotypes on trust and likability in spoken human–robot interaction. In: Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)
 35. Kuchenbrandt D, Häring M, Eichberg J, Eyssel F, André E (2014) Keep an eye on the task! how gender typicality of tasks influence human–robot interactions. *Int J Soc Robot* 6(3):417–427
 36. Law T, Chita-Tegmark M, Scheutz M (2021) The interplay between emotional intelligence, trust, and gender in human–robot interaction. *Int J Soc Robot* 13(2):297–309
 37. Lippa R (1983) Sex typing and the perception of body outlines. *J Pers* 51(4):667–682
 38. Lisy D (2019) The sexual continuum, a diffractive analysis, and our apparatuses of investigation. In: *Feminist philosophy of technology*. Springer, pp 265–285
 39. Loideain NN, Adams R (2020) From alexa to siri and the gdpr: the gendering of virtual personal assistants and the role of data protection impact assessments. *Comput Law Secur Rev* 36:105366
 40. Lugones M (2007) Heterosexuality and the colonial/modern gender system. *Hypatia* 22(1):186–219
 41. Lugones M (2010) Toward a decolonial feminism. *Hypatia* 25(4):742–759
 42. Lugrin B, Ströle E, Obremski D, Schwab F, Lange B (2020) What if it speaks like it was from the village? effects of a robot speaking in regional language variations on users' evaluations. In: 2020 29th IEEE International conference on robot and human interactive communication (RO-MAN). IEEE, pp 1315–1320
 43. Lykke N (2010) *Feminist studies: a guide to intersectional theory, methodology and writing*. Routledge, London
 44. Makenova R, Karsybayeva R, Sandygulova A (2018) Exploring cross-cultural differences in persuasive robotics. In: Companion of the 2018 ACM/IEEE international conference on human–robot interaction, pp 185–186
 45. Mehta CM, Strough J (2009) Sex segregation in friendships and normative contexts across the life span. *Dev Rev* 29(3):201–220
 46. Moss G, Gunn R, Heller J (2006) Some men like it black, some women like it pink: consumer implications of differences in male and female website design. *J Consum Behav* 5(4):328–341
 47. Nass CI, Brave S (2005) *Wired for speech: How voice activates and advances the human–computer relationship*. MIT press, Cambridge
 48. Neuteboom SY, Graaf M, et al (2021) People's perceptions of gendered robots performing gender stereotypical tasks. In: International conference on social robotics. Springer, pp 24–35
 49. Nomura T (2017) Robots and gender. *Gender Genome* 1(1):18–25
 50. Nomura T, Kinoshita Y (2015) Gender stereotypes in cultures: experimental investigation of a possibility of reproduction by robots in japan. In: 2015 International conference on culture and computing (culture computing). IEEE, pp 195–196
 51. Nomura T, Takagi S (2011) Exploring effects of educational backgrounds and gender in human–robot interaction. In: 2011 International conference on user science and engineering (i-user). IEEE, pp 24–29
 52. Nosek BA, Hawkins CB, Frazier RS (2011) Implicit social cognition: from measures to mechanisms. *Trends Cogn Sci* 15(4):152–159
 53. Paetzel M, Perugia G, Castellano G (2020) The persistence of first impressions: the effect of repeated interactions on the perception of a social robot. In: Proceedings of the 2020 ACM/IEEE international conference on human–robot interaction, pp 73–82
 54. Paetzel M, Peters C, Nyström I, Castellano G (2016) Congruency matters-how ambiguous gender cues increase a robot's uncanniness. In: International conference on social robotics. Springer, pp 402–412
 55. Paetzel M, Peters C, Nyström I, Castellano G (2016) Effects of multimodal cues on children's perception of uncanniness in a social robot. In: Proceedings of the 18th ACM international conference on multimodal interaction, pp 297–301
 56. Perugia G, Boor L, van der Bij L, Rikmenspoel O, Foppen R, Guidi S (2023) Models of (often) ambivalent robot stereotypes: content, structure, and predictors of robots' age and gender stereotypes. In: Proceedings of the 2023 ACM/IEEE international conference on human–robot interaction, pp 428–436
 57. Perugia G, Guidi S, Bicchi M, Parlangeli O (2022) The shape of our bias: perceived age and gender in the humanoid robots of the abot database. In: Proceedings of the 2022 ACM/IEEE international conference on human-robot interaction, pp 110–119
 58. Perugia G, Rossi A, Rossi S (2021) Gender revealed: evaluating the genderedness of furhat's predefined faces. In: International conference on social robotics. Springer, pp 36–47
 59. Pfeifer A, Lugrin B (2018) Female robots as role-models?-the influence of robot gender and learning materials on learning success. In: International conference on artificial intelligence in education. Springer, pp 276–280
 60. Phillips E, Zhao X, Ullman D, Malle B.F (2018) What is human-like?: decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. In: 2018 13th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 105–113
 61. Pitardi V, Bartikowski B, Osburg V.S, Yoganathan V (2022) Effects of gender congruity in human–robot service interactions: the moderating role of masculinity. *Int J Inf Manag* 102489
 62. Powers A, Kiesler S (2006) The advisor robot: tracing people's mental model from a robot's physical attributes. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, pp 218–225
 63. Powers A, Kramer A.D, Lim S, Kuo J, Lee S.I, Kiesler S (2005) Eliciting information from people with a gendered humanoid robot. In: ROMAN 2005. IEEE International workshop on robot and human interactive communication, 2005. IEEE, pp 158–163
 64. Rea DJ, Wang Y, Young JE (2015) Check your stereotypes at the door: an analysis of gender typecasts in social human–robot interaction. In: International conference on social robotics. Springer, pp 554–563
 65. Reich SM, Black RW, Foliaki T (2018) Constructing difference: lego® set narratives promote stereotypic gender roles and play. *Sex Roles* 79(5):285–298

66. Reich-Stiebert N, Eyssel F (2017) (ir) relevance of gender? on the influence of gender stereotypes on learning with a robot. In: 2017 12th ACM/IEEE international conference on human–robot interaction HRI. IEEE, pp 166–176
67. Robertson J (2010) Gendering humanoid robots: Robo–sexism in Japan. *Body Soc* 16(2):1–36
68. Roesler E, Naendrup-Poell L, Manzey D, Onnasch L (2022) Why context matters: the influence of application domain on preferred degree of anthropomorphism and gender attribution in human–robot interaction. *Int J Soc Robot* 1–12
69. Sandygulova A, Dragone M, O’Hare GM (2014) Investigating the impact of gender development in child–robot interaction. In: Proceedings of the 2014 ACM/IEEE international conference on human–robot interaction, pp 284–285
70. Sandygulova A, O’Hare GM (2015) Children’s perception of synthesized voice: robot’s gender, age and accent. In: International conference on social robotics. Springer, pp 594–602
71. Sandygulova A, O’Hare GM (2016) Investigating the impact of gender segregation within observational pretend play interaction. In: 2016 11th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 399–406
72. Sandygulova A, O’Hare GM (2018) Age-and gender-based differences in children’s interactions with a gender-matching robot. *Int J Soc Robot* 10(5):687–700
73. Seo S (2022) When female (male) robot is talking to me: effect of service robots’ gender and anthropomorphism on customer satisfaction. *Int J Hosp Manag* 102:103166
74. Siegel M, Breazeal C, Norton M.I (2009) Persuasive robotics: the influence of robot gender on human behavior. In: 2009 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 2563–2568
75. Søråa Roger Andre (2017) Mechanical genders: how do humans gender robots? *Gender Technol Dev* 21(1–2):99–115. <https://doi.org/10.1080/09718524.2017.1385320>
76. Sparrow Robert (2017) Robots, rape, and representation. *Int J Soc Robot* 9(4):465–477. <https://doi.org/10.1007/s12369-017-0413-z>
77. Sparrow R (2020) Robotics has a race problem. *Sci Technol Human Values* 45(3):538–560
78. Spiel K, Haimson OL, Lottridge D (2019) How to do better with gender on surveys: a guide for hci researchers. *Interactions* 26(4):62–65
79. Steinhäusser SC, Schaper P, Bediako Akuffo O, Friedrich P, Ön J, Lugin B (2021) Anthropomorphize me! effects of robot gender on listeners’ perception of the social robot nao in a storytelling use case. In: Companion of the 2021 ACM/IEEE international conference on human–robot interaction, pp 529–534
80. Tanevska A, Chandra S, Barbareschi G, Eguchi A, Han Z, Korpan R, Ostrowski AK, Perugia G, Ravindranath S, Seaborn K, Winkle K (2023) Inclusive hri ii: equity and diversity in design, application, methods, and community. In: Companion of the 2023 ACM/IEEE international conference on human–robot interaction, pp 956–958
81. Tay B, Jung Y, Park T (2014) When stereotypes meet robots: the double–edge sword of robot gender and personality in human–robot interaction. *Comput Hum Behav* 38:75–84
82. Thellman S, Hagman W, Jonsson E, Nilsson L, Samuelsson E, Simonsson C, Skönvall J, Westin A, Silvervarg A (2018) He is not more persuasive than her: no gender biases toward robots giving speeches. In: Proceedings of the 18th international conference on intelligent virtual agents, pp 327–328
83. Wajcman J (2004) TechnoFeminism. *Polity*
84. West M, Kraut R, Chew HE (2019) I’d blush if i could: closing gender divides in digital skills through education
85. Winkle K, Melsión G.I, McMillan D, Leite I (2021) Boosting robot credibility and challenging gender norms in responding to abusive behaviour: a case for feminist robots. In: Companion of the 2021 ACM/IEEE international conference on human–robot interaction, pp 29–37
86. You HC, Lin KW (2019) Gendered tour-guide robots and their influence on user attitude and behavior. In: Proceedings of the 5th international ACM in-cooperation HCI and UX conference, pp 32–35
87. Zhao S (2006) Humanoid social robots as a medium of communication. *New Media Soc* 8(3):401–419
88. Zhumabekova K, Ismailova A, Kushkinbayev D, Sandygulova A (2018) Exploring the effects of robot gender on child–robot interaction. In: Companion of the 2018 ACM/IEEE international conference on human–robot interaction, pp 287–288

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dr. Giulia Perugia is an Assistant Professor at the Human-Technology Interaction Group of Eindhoven University of Technology (TU/e). She holds a double degree PhD in Assistive Technologies from the Technical University of Catalonia (UPC) and TU/e and was a postdoctoral researcher at Uppsala Social Robotics Lab from 2018 to 2021. As a social scientist, her research lies at the intersection of Social Robotics, Social Psychology, and Ethical and Inclusive HRI. Among others, she is interested in investigating bias and stereotyping in the design and perception of humanoid robots, translating ethical issues (e.g., deception) into testable experiments and design problems, and identifying ways to co-design social robots and AI technologies with their intended stakeholders.

Dominika Lisy is a PhD candidate at the Department of Thematic Studies, Division of Gender Studies at Linköping University under the supervision of Dr. Katherine Harrison, Prof. Harald Wiltsche, and Prof. Ericka Johnson. Her research is part of the Wallenberg AI, Autonomous Systems and Software Program - Humanities and Society (WASP-HS) project on “The ethics and social consequences of AI and caring robot”. She obtained a BSc in Psychology at University of Groningen, Netherlands, an MA in Gender studies at University of Gothenburg, Sweden, and an MSc in Cognitive Neuroscience at University of Skövde, Sweden. Her interdisciplinary PhD project is exploring metaphysical boundaries and affectivity in human-robot-relationship through a feminist philosophy of materiality, bodies, and the skin.