Previous research on gender effects in robots has largely ignored the role of facial cues. We fill this gap in the literature by experimentally investigating the effects of facial gender cues on stereotypical trait and application ascriptions to robots. As predicted, the short-haired male robot was perceived as more agentic than was the long-haired female robot, whereas the female robot was perceived as more communal than was the male counterpart. Analogously, stereotypically male tasks were perceived more suitable for the male robot, relative to the female robot, and vice versa. Taken together, our findings demonstrate that gender stereotypes, which typically bias social perceptions of humans, are even applied to robots. Implications for design-related decisions are discussed.

Imagine the following scenario that takes place several decades in the future: By then, both authors of this work might be senior citizens, and the same might hold true for our readers. In spite of grieving lost youth, however, imagine that each senior citizen of this future society would be equipped with a personal robot assistant that would take care of everyday life chores, such as personal care, household maintenance, and other conveniences for you. Your personal robot assistant would facilitate everyday life by being able to support you in any possible way. What could your robot assistant look like, and why did designers opt for this particular appearance? Would your companion’s look affect your perceptions of its “personality” and capabilities? The present research focuses on exactly these questions, as we address the issues of design choices in robots and their consequences for the perception of those robots.

It is clear that, to date, the scenario outlined here has not yet been fully realized. Nevertheless, taking into account the interdisciplinary effort of international scientists in social robotics, engineering, computer sciences, psychology, and related fields, such a vision will sooner or later become...
Forming Impressions of Other Persons

Classic work in person perception and impression formation has documented the impact of visual cues that are sufficient in triggering social categorization and subsequent stereotyping processes (Brewer, 1988; Fiske, 1998; Fiske & Neuberg, 1990). This research has also shown that the most salient social categories are age, race, and sex. Categorizing others along these lines almost automatically makes information processing efficient and economical (e.g., Bargh, 1999; Devine, 1989). However, stereotyping and discrimination are the costs associated with such automatic, category-based information processing. In this paper, we focus on gender stereotyping in particular, aiming to show that social categorization processes are even applied to nonhuman targets. That is, we will show that not only humans, but even robots fall victim to gender stereotypes.

Facial Cues and Gender-Stereotyping

Why should robots be gender-stereotyped in the first place? Previous research on face processing has demonstrated that human faces are a primary source of social category information that indicate, for instance, a person’s sex (Zebrowitz, 1997; see also Mason, Cloutier, & Macrae, 2006). In determining a target’s sex, however, individuals seem to rely on hairstyle as a salient facial cue in particular (Brown & Perrett, 1993; Burton, Bruce, & Dench, 1993). Indeed, in their priming experiments, Macrae and Martin (2007) demonstrated that isolated hair cues are capable of activating stereotypical knowledge structures about men and women. Specifically, they showed that long hair led to an increased accessibility of knowledge
structures about the social category of females, whereas short hair activated stereotypical knowledge structures about men. Importantly, in a subsequent experiment, Martin and Macrae (2007) also demonstrated that priming effects of hairstyle as a sex-specifying cue even emerged under suboptimal information processing conditions. That is, even when hair cues were presented in a distorted, blurred manner, they affected subsequent categorization of target faces according to sex. However, such facilitating effects were not obtained in the absence of hair cues when cropped faces were used as primes.

Taking these results into account, we argue that not only in social interactions between humans might hairstyle as a physiognomic cue facilitate social judgment processes. We propose that feature-based categorization and stereotyping processes might also be triggered by visual facial cues when these are manipulated using humanoid robots as targets. Specifically, we investigated whether subtle gender cues (i.e., hairstyle) in anthropomorphic robots influenced the way users perceived them and their functional capabilities.

Social Perception and Judgments of Computers and Robots

Our work extends the prominent and classic computers as social actors (CASA) approach (Reeves & Nass, 1996; also see Nass, Moon, Morkes, Kim, & Fogg, 1997). According to this perspective, people interact with computers in ways that are comparable to human–human interaction and, usually, people are not even aware of the fact that they do so automatically. That is, in a series of experiments, Nass and colleagues (e.g., Nass & Moon, 2000) showed that people instinctively treat computers like humans, for instance by mindlessly applying human social categories to computers.

By now, this line of research has been extended to robots and, in part, analogous findings were obtained (see Echterhoff, Bohner, & Siebler, 2006). In human–robot interaction, the robot’s voice, demeanor, or its appearance function as social cues that provide information about the robot’s persona and subsequently guide social perception (Powers & Kiesler, 2006; Powers et al., 2005). Thus, such cues facilitate human–robot interaction in that users infer certain traits and functions from the physical design of the robot. This is in line with Powers et al. (2005), who proposed that people do not “approach the robot tabula rasa but rather develop a default model of the robot’s knowledge” (p. 159). To illustrate, the appearance and language of a robot affected the perceived knowledge of the robot when it was described in terms of its ethnicity, either stemming from China versus the United States. In that study, the Chinese-speaking robot of “Asian ethnicity” was assumed
to know more about landmarks in China, compared to an “American” robot that had been developed in the U.S. (Lee, Kiesler, Lau, & Chiu, 2005).

In human–human social cognition, not only ethnicity, but also gender represents a core category, which is heuristically used in impression formation. In human–robot interaction, however, gendering processes based on visual cues are still underresearched. Existing work on gender effects in social robotics heavily builds on previous research by proponents of the CASA approach and has primarily focused on the machine’s synthetic voice as the main cue to trigger gender stereotyping of machines. For example, in their groundbreaking research on the effects of gender-stereotypic responses toward computers, Nass, Moon, and Green (1997) demonstrated that participants attributed gender to computers that communicated in a low- versus a high-pitched synthetic voice. Subsequently, the low versus high frequency of the synthetic computer voice triggered gender-schematic judgments of the “male” versus the “female” computer. Specifically, the female-voiced computer in a dominant role was perceived more negatively than was the male-voiced dominant computer. Furthermore, evaluations provided by the male computer were taken more seriously than when praise was given by the female computer.

Lately, Nass’s work was extended from computers to anthropomorphic robots: Powers and Kiesler (2006) investigated the effects of physical appearance and voice frequency on attribution of sociability and competence to a robot. They tested the assumption that a baby-faced humanoid robot would be perceived as more sociable, but less competent than a mature-faced robot and that the robot’s appearance would influence advice-taking intentions of the perceivers. Indeed, the authors found that baby-facedness predicted perceived sociability of the robot, as well as participants’ intentions to take health advice from the robot. Furthermore, low voice frequency of the robot predicted perceived knowledgeability of the robot and participants’ willingness to take advice from it.

In the present research, our goals are twofold. On the one hand, we aim at extending the sparse literature on gendering of machines by investigating visual facial features as focal cues to trigger such processes. On the other hand, we go beyond classic CASA research by examining these effects in anthropomorphic robots. Specifically, we investigated the effects of visual gender cues (i.e., hairstyle) on gender-stereotypic perceptions of the targets. Gender-stereotypic perceptions were assessed using traits indicative of prototypically “male” agency and “female” communion. This draws on classic research by Bem (1974, 1981; see also Campbell, Gillaspy, & Thompson, 1997; Schneider-Düker & Kohler, 1988; Spence & Helmreich, 1978; Spence,

3For recent work on effects of vocal femininity in humans, see Ko, Judd & Blair (2006).
Helmreich, & Stapp, 1974, 1975), who identified stereotypically male and female personality traits and their role in gender-schematic information processing.

These traits can also be characterized along the dimensions of agency and communion. Numerous empirical studies have documented that these two dimensions are the core dimensions of social cognition (Abele, 2003; Cuddy, Fiske, & Glick, 2008; Fiske, Cuddy, Glick, & Xu, 2002; Williams & Bargh, 2008). That is, whereas women are stereotypically ascribed traits related to interpersonal warmth, men are perceived predominantly in terms of agentic features (also see Eagly, 1987; Eagly & Mladinic, 1989; Eagly, Wood, & Diekman, 2000; Glick & Fiske, 1999; Schneider-Düker & Kohler, 1988, Spence & Helmreich, 1978; Spence et al., 1974, 1975).

Does this also hold true for gendered robots? In recent years, advancements in the technical development of robots have resulted in an increasing interest in determinants that facilitate human–robot interaction. Designers of more or less humanlike robots have particularly been interested in the study of features in the appearance of a robot that not only facilitates a smooth robot–user interaction, but also indicates fields of applications (e.g., DiSalvo et al., 2002; Duffy, 2003). Thus far, however, little is known about the social psychological implications of such design-related decisions. For example, this concerns the systematic study of applications of specific types of robots (Hinds et al., 2004; Kanda et al., 2004; Lohse, Hegel, & Wrede, 2008; Pineau et al., 2003).

In this paper, therefore, we address issues of potential applications and the extent of stereotypical trait ascriptions to anthropomorphic robots, thereby extending the literature in both domains. We propose the following:

**Hypothesis 1.** There will be a Robot Type × Trait Type interaction such that participants will ascribe more agency-related traits to the male robot than to the female robot, whereas they will ascribe more communal traits to the female robot than to the male robot.

**Hypothesis 2.** There will be a Robot Type × Task Type interaction such that participants will perceive the male robot as being more capable of stereotypically male tasks than the female robot, whereas they will perceive the female robot as being more capable of stereotypically female tasks than the male robot.

**Hypothesis 3.** Participants will choose a task of verbal ability over a task of mathematical ability when having to interact with the female robot, whereas they will choose a task that requires mathematical ability over the one requiring verbal ability when
having to interact with the male robot. The effects of gender stereotypes will also affect participants’ choice of team task that must be completed in cooperation with the male robot versus the female robot, respectively.

We tested the hypotheses in a computerized experiment.

Method

Participants

There were 60 (30 males, 30 females) participants who completed the study. The participants’ mean age was 24 years ($SD = 3.86$; range = 19–38 years). Most of the participants were German (90%) students who were recruited on the campus of the University of Bielefeld. The students were, on average, in their fifth semester of study ($SD = 3.66$), and they majored in a variety of fields.

Materials

Robot stimuli. Participants were told that we were interested in their attitudes towards “modern technologies of the future.” Specifically, they were informed that we would be interested in their opinion about two newly developed robots (Hegel, 2010). As part of a within-subjects design, participants rated two robot types. Three $5 \times 5$ cm headshots of the targets were presented on the computer screen during the ratings (see Figures 1 and 2). The items were presented one after another on separate screens.

Figure 1. Female robot type: long hair.
With all other aspects held constant, the two sets of stimulus pictures differed only in terms of the robots’ hairstyles and the shape of the lips. As shown in Figures 1 and 2, the “female” robot was presented with long hair, whereas the “male” robot was presented with a short haircut to indicate masculinity.

In a pilot study, 20 participants recruited at the University of Bielefeld were presented with the pictures and judged the “gender” of two robot types. They were asked to indicate the extent to which the robot appeared “rather male” versus “rather female” using a 7-point Likert-type scale. The scale midpoint indicated gender neutrality of the target. As predicted, the long-haired “female” ($M = 1.78$, $SD = 0.97$) robot was perceived as significantly more feminine than was the short-haired “male” robot ($M = 4.67$, $SD = 1.80$), $t(16) = 4.23$, $p = .001$.

**Trait attributions.** Because this study was a within-subjects design, both targets had to be evaluated. Participants began with their judgments of stereotypically feminine and masculine traits that were taken from a German version of the Bem Sex-Role Inventory (Schneider-Düker & Kohler, 1988). To make their evaluations, the participants were presented with a fixed randomized list of 12 adjectives per dimension, and they rated the adjectives on a 7-point scale, with higher ratings representing greater agreement. The list of traits consisted of 12 attributes that tap the dimensions of communion (e.g., *affable, friendly, polite, affectionate*) and agency (e.g., *assertive, dominant, determined, authoritative*), respectively. The complete set of traits is listed in the Appendix.

**Applications.** Following the trait ratings, participants were asked to judge the extent to which they would utilize the robot in a specific application area. Importantly, the application areas were selected on the basis of their gender stereotypicality for men or women, as indicated by pretesting.
(e.g., transport goods, monitoring technical devices, prepare meals, elderly care, childcare). To illustrate, repairing technical equipment was judged as a prototypically male task, whereas taking care of children was perceived as a stereotypically female task. Thus, the participants rated a list of six typically male tasks and six typically female tasks. Responses were rated on a 7-point scale to indicate the extent to which the robots could be used for the respective tasks. Again, higher values represent greater perceived suitability of the robots for each task. The gender-stereotypic tasks are listed in the Appendix.

Choice of team task. To examine the effects of gender cues on team task choice, participants were asked to indicate for each of the two robots if they would choose the robot as interaction partner for (a) a (stereotypically male) math task; or (b) a (stereotypically female) verbal task.

Additional variables. Participants were asked to provide demographic data (gender, age, major, number of semesters of study, nationality, mother tongue). At the end of the survey, we asked participants to indicate the perceived “gender” of the robot they had judged previously on a 7-point scale with the labels masculine, neutral, or feminine. This item was included to serve as a manipulation check.

Procedure

Participants were tested individually using desktop computers. They rated both types of robots with regard to traits and tasks as part of a within-subjects design. After completion of the questionnaires, participants received a chocolate bar and were debriefed, thanked, and dismissed.

Results

Preliminary Analyses

Internal consistencies (Cronbach’s $\alpha$) were calculated for the stereotypically male, agency-related traits and the stereotypically female, communal traits to form indexes that consisted of 12 items each. Furthermore, Cronbach’s alpha was calculated for the typically male and female task indexes for each robot. Overall, the composite scores show good to very good reliability, as illustrated in Table 1. Consequently, mean scores of each composite were computed, with high values indicating a strong attribution of traits or task suitability to either of the robots.
Test of Main Hypotheses

Manipulation check. In order to ensure that participants had correctly perceived the robot types as “male” versus “female” targets, we analyzed responses regarding perceived robot gender. We predicted that the robot with short hair would be perceived as more masculine than would the robot with long hair. To test this hypothesis, we ran independent-sample \( t \) tests, comparing male versus female robot type with respect to perceived gender of the robot. The results indicate that, as hypothesized, the robot with short hair was perceived as more masculine (\( M = 4.60, SD = 2.04 \)) than the long-haired robot (\( M = 1.65, SD = 0.92 \)), \( t(59) = 11.26, p < .001 \).

Ascription of stereotypically male and female traits. Hypothesis 1 predicted a Trait Type × Robot Type interaction. To test the hypothesis, we conducted a repeated-measures ANOVA with agency and communion indexes as repeated measures. A significant main effect of trait type was obtained, \( F(1, 59) = 34.68, p < .001 \). That is, overall, participants provided higher communion ratings than agency estimates. The main effect of robot type was not significant (\( F < 1 \)). Most importantly, as predicted, the Trait Type × Robot Type interaction was significant, \( F(1, 59) = 6.15, p = .02 \). Subsequently, we conducted one-sided \( t \) tests and found that the male robot was perceived as possessing more agency-related, stereotypically masculine traits (\( M = 2.92, SD = 1.03 \)) than was the female robot (\( M = 2.68, SD = 0.92 \)), \( t(59) = 2.36, p = .01 \), whereas the female robot was perceived as slightly warmer (\( M = 4.17, SD = 1.31 \)) than was the male robot (\( M = 3.96, SD = 1.32 \)), \( t(59) = 1.93, p = .03 \). Figure 3 displays the pattern of results.

Potential applications for the robots. With regard to possible fields of application of both robots, we also predicted a Task Type × Robot Type
interaction. That is, Hypothesis 2 predicted that participants would be more likely to choose stereotypically male applications for the male robot than for the female robot; whereas stereotypically female tasks would be perceived as more suitable for the female robot, relative to the male robot.

Again, we conducted a repeated-measures ANOVA on the dependent variables. Both main effects of task type and robot type were not significant: task type, $F < 1$; robot type, $F(1, 59) = 1.95, p = .17$. However, the interaction pattern predicted by Hypothesis 2 was obtained, $F(1, 59) = 19.32, p < .001$. That is, as hypothesized, participants perceived the male robot as more suitable for “male” tasks ($M = 4.54, SD = 1.21$) than the female robot ($M = 3.91, SD = 1.32$), $t(59) = 4.26, p < .001$; whereas the female robot was perceived as more suitable for stereotypically “female” tasks ($M = 4.38, SD = 1.73$) than was the male robot ($M = 3.92, SD = 1.54$). This is illustrated in Figure 4.

**Choice of team task.** We analyzed participants’ responses regarding their choice of team task they would like to complete with the two robots. To do so, we conducted chi-square tests on team task as a function of robot gender. The results show that participants were equally likely to choose a math task or a verbal task (52% vs. 48%) to work on with the female robot. In this case, the chi-square test was not significant, $\chi^2(1) = 0.07, p = .80$. With regard to team task choice for the male robot, however, the participants significantly more often opted for the math task (73%) than the verbal task (27%), $\chi^2(1) = 13.07, p < .001$.
Discussion

In the present study, we experimentally investigated the effects of visual gender cues on the perception of anthropomorphic robots. Specifically, we examined whether hair length of a target biased the ascriptions of stereotypically male versus female traits and applications to two gendered robot types. Apparently, our manipulation of robot “gender” by varying hair length was successful: The manipulation check confirmed that the short-haired robot was perceived as more masculine than was the long-haired robot. These visual gender cues, in turn, affected social perceptions of the robot types.

Importantly, the data support our main hypotheses. First, as predicted, short hair versus long hair resulted in differential stereotypic judgments of the gendered robots in that participants perceived the male robot as more agentic than the female robot. The female robot, on the other hand, was perceived as more communal than the male robot. Thus, obviously, in line with findings on the effects of human hairstyle on judgments and behavior (e.g., Macrae & Martin, 2007; Martin & Macrae, 2007), the robots’ hair cues activated participants’ knowledge structures about males and females, and gender stereotypes subsequently biased the evaluations of the robots. In addition, our findings show that basic dimensions of human social cognition—namely, agency and communion—were equally applied to non-human objects (Abele, 2003; Cuddy et al., 2008).

Second, with regard to perceived suitability for sex-typed tasks, we found that, again, the male robot was perceived as more suitable for typically male
tasks (e.g., repairing technical devices, guarding a house) than was the female robot. The female robot was perceived as more suitable for gender-stereotypically female tasks (e.g., tasks related to household and care services).

Third, stereotypes also affected participants’ choice of team task to be completed with the robots. That is, with regard to the male robot as a potential interaction partner, participants were significantly more likely to choose a task that required mathematical ability than verbal ability. On the other hand, participants did not differentiate between task types for the female robot. Because of social desirability concerns, participants might have refrained from openly “discriminating against” the female robot by choosing it as team partner for tasks that require stereotypically female verbal ability. This, however, would need to be tested in future studies.

Taken together, our research shows that people apply gender stereotypes that typically characterize human–human social cognitive processes to robots. Even though the trait attributions may be interpreted in terms of anthropomorphism (Epley, Waytz, & Cacioppo, 2007), our findings also document that gender stereotypes seem to be so deeply ingrained that people even applied them to machines with a male or a female appearance.

Given such findings, issues related to the ethics of robot design inevitably arise: Should gender stereotypes be used and exploited in the design of robots or virtual agents to manipulate the user’s mental models? Or should designers instead construct counterstereotypical machines (e.g., female service robots to help a mechanic, male CareBots)? These questions seem to be worth considering by developers and designers of robots because of their social and societal consequences.

Since Allport (1954), many researchers have attempted to change stereotypes and prejudice. Consequently, designers should rather develop gender-neutral or counterstereotypical machines to counteract the stability of personal and cultural stereotypes (Devine, 1989). On the other hand, opponents of this view might argue that design choices depend on when and where the machine is to be used. To illustrate, if the ultimate goal is to reduce errors produced by machines, to minimize risks and dangers for the people using them (e.g., as rescue assistants), and to improve pleasantness of human–robot interaction, one might recommend the construction of stereotype-congruent robots. For instance, our research showed that the male robot was perceived as more dominant and agentic than the female robot. Thus, if one would decide to use this robot to assist in elderly care, it might well be possible that the male robot could evoke fear in elderly human users. On the other hand, such negative and hindering effects could be avoided by exploiting human mental models and expectations pertaining to female robots, because our data showed that female robots were not only perceived as
communal, but also as more suitable as care assistants than the male robots. However, further research in applied contexts is needed to provide insight into how our robot prototype is perceived by senior citizens living in care facilities, for example.

Overall, the present research not only raises questions of high relevance from the applied perspective. Our work is equally important with regard to basic research because we extended both classic research by Nass, Moon, and Green (1997), as well as more recent work that has applied the CASA approach to social robotics (Powers et al., 2005). We do so in two ways: On the one hand, the present research is innovative in that it examined visual cues instead of voice cues to indicate robot gender. On the other hand, our work introduces the first results regarding the newly developed anthropomorphic robot “FloBi” (Hegel, 2010; Hegel, Eyssel, & Wrede, 2010). Its modular industrial design allows for clear-cut variations in robot character, and thus represents an optimal tool for prospective experimental research in social robotics and social psychology. This has important implications for its future use in basic research regarding determinants and consequences of anthropomorphism. Specifically, the modular construction of the robot head allows for maximal flexibility in usage. This aspect can prove particularly advantageous in the context of human–robot interaction studies with embodied gendered robots.

Once the robot’s technical system is advanced enough, we will conduct interaction studies to replicate our findings in a more ecologically valid setting. We would hypothesize that the opportunity for participants to see the robot embodied and to interact with it might reinforce stereotyping processes (also see Krach et al., 2008). In a less artificial interaction scenario, we could also study observable “discriminatory” behavior displayed toward the female robot type, relative to the male robot type (e.g., decreased willingness to choose the female robot as an interaction partner in tasks that require “typically male” skills; also see Powers & Kiesler, 2006). This would go beyond the assessment of mere behavioral intentions, as in the current research.

Moreover, future research could investigate the combined effects of visual and auditory gender cues. This way, we would integrate aspects of the classic CASA paradigm (Nass & Moon, 2000), adapt it to anthropomorphic robots (Powers et al., 2005), and then extend it by researching additional social cues that might be used in inferring traits and functionalities from a robot’s physical design. It is through this inference process that human–robot interaction can be facilitated.

Equally important is the aspect of the consequences of social categorization processes of nonhuman machines. Obviously, the data indicate the important role of visual gender cues in social cognitions about the robot targets’ potential applications and functionalities. Our findings have
important implications for the design of robots because it is through aesthetic form that social cues are transmitted. These, in turn, provide users with information regarding the capabilities of the social robot. Research by Gong (2008) also supports our argument about the physical aspects of robots. In a study on computer agents, he demonstrated that with an increasing degree of anthropomorphism, computer agents received increasingly more social responses (also see Parise, Kiesler, Sproull, & Waters, 1999).

To facilitate human–robot interaction, the appearance of a robot should be congruent with the estimations of users regarding the robot’s functionalities. We were able to show that male versus female physiognomic appearance affected the social perception of the robot target and influenced judgments of the robot’s persona and its fit to specific tasks. Further research is under way to shed more light on the social-cognitive mechanisms underlying anthropomorphism and gendering processes of robots, so this knowledge can be used to advance the development of robots.

Thus, given the combined interdisciplinary efforts of basic and applied researchers, the scenario we described in the introduction will become a future reality (Kanda et al., 2004; Pineau et al., 2003). Further research outcomes will determine whether our personal robot assistant would have a male, female, or gender-neutral appearance. At present, our participants’ social judgments of gendered robots seem to be grounded in rather gender-stereotypical conceptions about social roles. Perhaps sometime in the future, however, not only will robots be as technically refined as envisioned, but gender prejudice will also be a thing of the past.

References


Appendix

*Stereotypical Traits and Tasks Used in the Present Research*

Stereotypically male traits:

*Authoritative, speaks his mind, assertive, determined, aggressive, cold, organized, confident, hard-hearted, dominant, tough, has leadership skills*

Stereotypically female traits:

*Affectionate, empathetic, delicate, friendly, sincere, family-oriented, sensitive, cooperative, affable, polite, sentimental, romantic*

Stereotypically male tasks:

*Transporting goods, repairing technical equipment, guarding a house, steering machines, handcrafting, servicing equipment*

Stereotypically female tasks:

*Childcare, household maintenance, after-school tutoring, patient care, preparing meals, elderly care*