# WHAT DOES A ROBOT LOOK LIKE?: A MULTI-SITE EXAMINATION OF USER EXPECTATIONS ABOUT ROBOT APPEARANCE

Elizabeth Phillips, Daniel Ullman, Maartje M. A. de Graaf, and Bertram F. Malle
Brown University

Robot design is a critical component of human-robot interaction. A robot's appearance shapes people's expectations of that robot, which in turn affect human-robot interaction. This paper reports on an exploratory analysis of 155 drawings of robots that were collected across three studies. The purpose was to gain a better understanding of people's *a priori* expectations about the appearance of robots across a variety of robot types (household, military, humanoid, generic, and AI). The findings suggest that people's visualizations of robots have common features that can be grouped into five broad components. People seem to distinguish between human-like and machine-like robots, with a default visualization of robots having a human-like appearance. In addition, expectations about robot appearance may be dependent on application domain.

## INTRODUCTION

The vision of robots in homes, workplaces, and schools as human companions, caretakers, and teammates is no longer confined to science fiction. Rapid technological advancement and reduced production costs of robots and robotic devices has led to a commercialized product environment in which robots are on the precipice of becoming ubiquitous home companions, health aides, and other social partners. In these roles, robots will take on a variety of forms that activate and guide people's expectations about the robot's capacities and limitations. We argue that with the increasing number of robots entering into human spaces and helping fill roles traditionally occupied by humans, it is important to gain a better understanding of the expectations that humans hold of these robot counterparts. Having such an understanding will facilitate an appropriate match between what future users expect their robotic helpers to be and what those robots will be in reality.

Many studies have illustrated that robot appearance powerfully influences how people interact with robots, and that this influence is strongly mediated by people's expectations about those robots. According to the expectation-confirmation theory, a person's continued use of a product, service, or technology relies on their level of satisfaction and the alignment of their experience with prior expectations (Oliver, 1980). The reference against which people make their evaluations of technology consists of their prior expectations about that technology (Bhattacherjee, 2001). With regard to robots, people's expectations will guide their decisions to interact with a robot in the first place and, if they do, their decisions to continue interacting with that robot or forego further use.

Research studying human-robot interaction has shown that people tend to hold overly high expectations about the capabilities of robots (Komatsu, Kurosawa, & Yamada, 2012). When expectations are not met by actual experiences with robots, an expectation gap arises (Lohse, 2011), which is one of the reasons why some people abandon the use of a robot in the longer-term (Graaf, Ben Allouch, & Dijk, 2016). Importantly, people's expectations about robot capacities are often influenced by the appearance of a robot (Goetz, Kiesler, & Powers, 2003; Sims et al., 2005; Woods, Dautenhahn, & Schulz, 2004). Human-like appearance in robots raises specific expectations about the capabilities of robots (Lee, Park, & Song, 2005; Rızvanoğlu, Öztürk, & Adıyaman, 2014). The more human-like a robot looks, the more human-like qualities people attribute to robots (Hegel et al.,2008; Sims et al., 2005). The appearance of a robot can also influence the types of roles and jobs people expect robots to fulfill. Goetz, Kiesler, and Powers (2003) found that people preferred human-looking robots over mechanical-looking robots for highly social roles like office clerks, hospital staff, and instructors. In contrast, their study showed that people preferred mechanical-looking robots for roles like lab assistants and security guards. Finally, robot appearance has also been shown to influence initial perceptions of robot trustworthiness. In a study by Schaefer, Sanders, Yordon, Billings, and Hancock (2012), robots that were judged as looking more "robot-like" as opposed to "human-like" were perceived as less trustworthy.

Robot designers cannot prevent humans from forming expectations about a robot's capabilities based on its appearance; what designers can do is adjust the physical appearance to shape the most realistic expectations for a given type of robot. And that is what designers of recent social robots seem to have done. A recent article on social robots exhibited at the annual U.S. Consumer Electronics Show (CES) highlighted that the social robots looked strikingly similar, with many outfitted with minimalist faces, small bodies, and disproportionately large heads, reminiscent of baby animals (Ackerman, 2017). The designers interviewed for the article explained that one reason for this design choice was to use robot form to help appropriately match user expectations to real-world robot capabilities. By keeping human-like features at a minimum, and diminishing the ones that are present, designers aim to help end users more accurately understand these robots and find appropriate ways of interacting with them. However, what is the basis of these design decisions? What expectations do people really have about the proper appearance of a social robot?

In this paper, we report on exploratory research into human expectations about robot appearance across several robot types (household, military, humanoid, generic, and AI).

We report on data collected from three independent studies conducted across three universities in two countries. Each study was designed to capture ordinary people's intuitive expectations about the appearance of future robots. As a set, the studies investigated two broad research questions:

**RQ1**: What are the features that characterize human expectations about robot appearance?

**RQ2**: Does the presence of certain features or feature sets differ by robot type (e.g., military, household, etc.)?

## METHOD

In each study, participants were asked to create a visualization of a robot that corresponded to what they imagined a certain type of robot might look like. In Study 1, 84 participants were asked to draw what they felt a future military robot might look like. In Study 2, 21 participants were asked to either draw or collage together images that represented what they imagined future household robots might look like. In Study 3, 58 participants were asked to draw one of three things: a robot, a humanoid robot, or an artificial intelligence. Thus, variation in the method for each study resulted in a "robot type" between-subjects factor, in which drawings were created to represent household robots, military robots, humanoid robots, robots in general (generic), and an artificial intelligence. Across the three studies, 163 drawings of robots were generated, of which 155 were included in the analyses reported in this paper. Justification for the exclusion of selected drawings is described in detail below.

### Method: Study 1

Study 1 was conducted at the University of Central Florida in Orlando, Florida. Illustrations of robots were collected as part of a larger study investigating human-robot interaction with a simulated military robot. A total of 84 undergraduate students (62 male, 22 female) with ages ranging from 18-34 ($M$ = 19.34, $SD$ = 3.41) participated in the study. Participants were asked to: "Please draw a picture of what you think a military robot teammate might look like." Participants were informed that drawing ability was not of interest to the study and that a simple sketch would be fine. To make their drawings, participants were given a letter size sheet of paper with a bounding box that measured approximately 16cm x 18cm. Within that bounding box, participants were asked to complete their drawings. Drawings took approximately 5 minutes to complete. All participants completed the study in return for course credit in the Department of Psychology.

### Method: Study 2

Study 2 was conducted at the University of Twente in Enschede, The Netherlands. A total of 21 participants (7 male, 14 female) were selected to participate in a long-term study investigating interactions with social robots in users' homes. During the first of a series of interviews, the participants were told: "I would like you to create your ideal robot that you would like to have in your house. Please focus on how that robot should look and what purposes it should be able to fulfill." Participants were provided with cutouts of different robots from magazines and other sources, as well as pencils, markers, and colored paper. At the time these images were created, the participants had not yet been provided with the details of the long-term study, nor had they been introduced to the robot that would be deployed in their homes.

### Method: Study 3

Study 3 was conducted at Brown University in Providence, Rhode Island. A total of 58 undergraduate students (21 male, 33 female, 1 other) with a mean age of 20 ($SD$ = 1.7) participated in a robot drawing task that was appended to the end of an unrelated (non-robot) study. Approximately half of the students participated in return for course credit in undergraduate psychology courses, while the other half were compensated $15 for completing the 1-hour study. Participants completed the drawing task in a few minutes at the end of the experiment. Participants were given a letter size sheet of paper with instructions printed at the top, without a bounding box. Participants read one of three possible instructions: "Draw a robot." (18 participants), "Draw a humanoid robot." (22 participants), or "Draw an Artificial Intelligence (AI)." (18 participants).

### Coding scheme

All of the drawings from the three studies were aggregated into a single data set for subsequent coding and analysis (while tracking their origin and the type of robot requested to draw). Eight drawings were excluded from further analyses: 6 because participants drew more than one robot, which made coding difficult, and 2 because the sketches were too light to be seen clearly by the raters. The final dataset included 155 drawings of robots.

To evaluate each of the drawings, we devised a coding scheme that modified a scheme created by Ezer (2008). To better suit the drawing tasks in the present studies, we added coding categories pertaining to the assumed purpose of the robot (e.g., does the robot have weapons, clothes?) and removed categories irrelevant to the present drawings (e.g., presence of tools for taking care of pets). Coding categories included robot appearance features related to embodiment (e.g., does the robot have a body, arms, legs, feet, hands, fingers, toes?), specific humanlike facial features (e.g., does the robot have eyes, ears, nose, mouth, head, face?), locomotion methods (does the robot have tracks or treads, wheels?), features related to the assumed purpose of the robot (e.g., does the robot have clothes, weapons, antennae, movement features?), superficial or interaction features (e.g., does the robot have buttons, screen?) as well as whether or not the robot's appearance was more human-like or mechanical. Two independent raters indicated whether each feature in the coding scheme was present in each participant's drawing. To handle coder disagreements, we adopted the following procedure. For coding categories with inter-rater reliability of $\kappa > .75$ (i.e., head, face, eyes, ears, nose, mouth, arms, hands, ability to move, legs, feet, wheels, tracks or treads, antenna, and weapons), the authors identified the drawings in which raters disagreed. Then the authors reconciled, by majority vote, whether the given feature was present. The same procedure was followed for coding categories with moderate agreement (presence of a body, $\kappa$ = .56; human-like appearance, $\kappa$ = .62). We recognized that the coding instructions for two features had been imprecise, resulting in very low kappas (mechanical appearance and gender); adopting sharper instructions, we therefore recoded these categories across all the robot drawings by majority vote as well. This resulted in 155 images of robots rated on 19 appearance features to be used in subsequent analyses.

Table 1.
*Results of Principal Components Analysis*

| **Coding category "Does the robot have…?"** | **Human-like Motion** | **Facial Features** | **Bodily Features** | **Gendered vs. Mechanical** | **Terrain / Military** |
|---|---|---|---|---|---|
| 1. Legs | **.883** | .211 | .186 | .041 | .015 |
| 2. Feet | **.863** | .171 | .174 | .111 | .015 |
| 3. Human-like appearance | **.798** | .359 | .190 | .187 | .051 |
| 4. Tracks or treads | **-.643** | .087 | .214 | -.075 | **.600** |
| 5. Hands | **.541** | .259 | .462 | .087 | .025 |
| 6. Eyes | .357 | **.789** | .231 | -.034 | -.073 |
| 7. Face | .431 | **.771** | .257 | -.032 | .021 |
| 8. Mouth | .369 | **.758** | .145 | -.025 | -.114 |
| 9. Nose | .029 | **.556** | .090 | .391 | -.370 |
| 10. Head | .417 | **.540** | .455 | .043 | .095 |
| 11. Ears | -.019 | **.521** | .011 | .020 | .105 |
| 12. Ability to move | .144 | .124 | **.880** | -.083 | .062 |
| 13. Body | .113 | .182 | **.877** | -.049 | .066 |
| 14. Arms | .432 | .229 | **.662** | .049 | .072 |
| 15. Gender | .135 | .233 | .072 | **.800** | -.022 |
| 16. Antennae | -.014 | .292 | -.016 | **-.652** | -.063 |
| 17. Mechanical appearance | -.163 | -.103 | .427 | **-.640** | -.014 |
| 18. Weapons | -.021 | -.106 | .191 | .042 | **.653** |
| 19. Wheels | -.399 | -.282 | .371 | -.053 | **-.627** |
| **Eigenvalue** | 6.763 | 2.477 | 1.452 | 1.349 | 1.309 |
| **% Cumulative Variance** | 35.60 | 48.63 | 56.28 | 63.38 | 70.26 |
| **Cronbach's α for Subscale** | .87 | .86 | .84 | .54 | .34 |

*Note:* Reported eigenvalues are those after varimax rotation. Cronbach's α values for the five component subscales are computed from the boldfaced high-loading variables (loadings > .50).

Table 2.
*Table of means, standard deviations, and results of one-way ANOVAs.*

| **Factor** | **Robot** | | **Humanoid** | | **Household** | | **Military** | | **AI** | | Omnibus | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *F* | *p* | *Eta*$^2$ |
| **Human-like Motion** | 0.67 | 0.34 | 0.90 | 0.22 | 0.79 | 0.30 | 0.47 | 0.39 | 0.46 | 0.37 | 8.51 | <.001 | 0.19 |
| **Facial Features** | 0.66 | 0.25 | 0.79 | 0.14 | 0.53 | 0.29 | 0.32 | 0.30 | 0.30 | 0.38 | 15.138 | <.001 | 0.29 |
| **Bodily Features** | 0.94 | 0.24 | 0.97 | 0.15 | 0.97 | 0.10 | 0.90 | 0.20 | 0.37 | 0.48 | 20.826 | <.001 | 0.36 |
| **Gendered vs. Mechanical** | 0.17 | 0.24 | 0.44 | 0.34 | 0.32 | 0.28 | 0.37 | 0.28 | 0.39 | 0.24 | 2.753 | .030 | 0.07 |

## RESULTS

### Research Question 1: Are there characteristic robot appearance features?

Principal Components Analysis (PCA) was used to investigate which features or feature sets best characterized human expectations about robot appearance in the entire dataset. After Varimax rotation, the PCA revealed the presence of five components with eigenvalues greater than 1, explaining 70.26% of the total variance (see Table 1). Similarly, the scree plot suggested a break after the fifth component. Component 1 revealed strong loadings (> .50) for five features characterizing human-like appearance, especially motion (legs, feet, human-like appearance, hands vs. tracks or treads). Component 2 revealed strong loadings for six features characterizing the robot's head and face (eyes, face, mouth, nose, head, and ears). Component 3 revealed strong loadings for three features characterizing the robot's body (locomotion features, body, and arms). Component 4 indicated whether the robot was drawn as gendered (which was predominantly male) vs. had mechanical elements (antennae, and mechanical appearance). Component 5 indicated rough-terrain or military use (weapons, tracks or treads vs. wheels). We formed five subscales from the respective sets of high-loading variables and computed α reliabilities. As the last component had a very low alpha and is mainly a result of drawings of military robots, we focused on the first four factors (still explaining 63% of the variance) in addressing research question 2.

### Research Question 2: Differences across robot types?

We conducted four separate one-way ANOVAs on the orthogonal component subscales to test whether the robot type (household, military, humanoid, generic, and AI) predicted scores on these component subscales (see Table 2). The results showed significant prediction for human-like motion features ($F(4, 149) = 8.51$, $p < .001$), facial features ($F(4, 149) = 15.138$, $p < .001$), bodily features ($F(4, 148) = 20.826$, $p < .001$), as well as gender vs. mechanical features ($F(4, 149) = 2.753$, $p = .030$).

Further examination of these relationships revealed a number of notable differences in the presence of features between robot types. On the component of human-like motion features, drawings made in the military condition had significantly fewer human-like motion features than drawings of robots made in the household condition and in the humanoid condition (each $p < .001$). Similarly, drawings made in the AI condition had fewer human-like motion features than drawings made in the household ($p = .053$) and humanoid ($p = .002$) conditions. On the component of facial features, we observed an ordering of drawing conditions, where humanoid robots had the highest number of facial features followed by generic robots, household robots, AI, and military robots. Drawings of humanoid robots, generic robots, and household robots were roughly equivalent to each other in their high scores on facial features, although humanoid robots had significantly more facial features than household robots ($p = .032$). In addition, the humanoid, generic, and household robots each had more facial features than military drawings ($p = .056$, $p < .001$, $p < .001$, respectively). Military and AI drawings did not significantly differ from one another on facial features. On the component of bodily features, drawings of AI showed significantly fewer bodily features than all the other drawing conditions (each $p < .001$). Finally, the gender vs. mechanical component showed that the generic robots had less prominent gendered features than the other robot drawings, but only the comparison to the humanoid robots reached statistical significance ($p = .025$).

## DISCUSSION

The current study offers exploratory insights on people's expectations about robot appearance across several robot types (household, military, humanoid, generic, and AI). The results of the PCA reveal that people's visualizations of robots have common features that can be grouped into five broad components: human-like motion, facial features, bodily features, gendered vs. mechanical appearance, and terrain/military domain.

Overall, the results showed some main trends in the data. First, the findings suggest that people tend to imagine two broad classes of robots as a function of application domain. One class representing human-like robots intended for household or human-assistance applications, and the other representing machine-like robots intended for military or search and rescue applications. Humanoid, household, and generic robots comprise a group with more human-like features than robots intended for military or search and rescue applications. Robots rated as mechanical-looking were less likely to be rated as having a gender, indicating that people may have more gender-neutral conceptualizations of mechanical robots. In addition, robots rated as mechanical-looking were more likely to be drawn with an antenna.

A second trend in the data is related to the application domain for robots. Overall, people across the conditions expected household and generic robots to be more human-like in shape, with facial features, and legs for locomotion. Military robots, on the other hand, had statistically fewer facial features than humanoid, household, and generic robots, and had few human-like motion features. This pattern might be a reflection of participants' conceptualizations of many modern unmanned ground vehicles (UGVs) used heavily in U.S. military and search and rescue domains (Pejic, 2016). Also, participants' representations of military robots and AI stood out as most different from the other robots. For most features, the humanoid, generic, and household robots were not different from one another but were different from the military robots and AI. Additionally, the category of AI produced highly diverse drawings, ranging from fully-embodied robots to abstract networks. These findings support earlier research suggesting that a robot's appearance should match its intended application domain (Goetz, Kiesler, & Powers, 2003; Graaf & Ben Allouch, 2015).

A third and final trend in the data represents a tension between a possible default for human-like appearance and a desire for differentiation by domain. Unsurprisingly, drawings representing a humanoid robot included the highest amount of human-like features (e.g., facial features). However, even when people were simply asked to “draw a robot” (i.e., generic robot condition) as well as specifically asked to draw a household robot, the robot visualizations still had many human-like features. This suggests that people may have a default conceptualization of robots resembling a humanoid appearance. On the other hand, visualizations of household robots and generic robots had fewer facial features and human-like motion features than humanoid robots. This may indicate that people’s expectations for generic and household robots include some humanoid features, but perhaps not as many as a near human-looking android. A similar tension was identified in the study by de Graaf and Ben Allouch (2015), which found that although people favored a humanoid appearance for a household robot, they expressed that robots should be easily distinguishable from humans. Thus, robot developers will want to consider the level of human-likeness of robot appearance in their designs, which might be dependent on application domain.

There are some limitations in this study that need to be addressed. First, although the participants in the household condition (Study 2) had the option to draw or otherwise create by hand their visualizations of robots, they mostly created collages of robots largely from existing materials (e.g., magazines) provided to them. This inherently limited their robot representations. The second limitation is that paper presents exploratory research. The different types of robot drawings were not produced under tightly controlled conditions, the studies relied on slightly different instructions, and they had uneven sample sizes. A final limitation is that our coding scheme was an adapted version of one used in previous work; thus, it must be validated by further research. Nonetheless, the covariance among many of the appearance features was surprisingly high and suggests compelling hypotheses about people’s systematic representations of future robots. Therefore, the current results offer a valuable starting point for integrating people’s expectations of robot appearance into successful robot design.

## CONCLUSION

In conclusion, as robots become increasingly more integrated into everyday life, we must account for people’s expectations about robots. In this inquiry we set out to capture what people expect future robots to look like. Understanding people’s expectations about robot appearance is critical to proper robot design, in turn contributing to successful human-robot interaction, and increased acceptance within society. In general, features that make up a robot’s appearance can be grouped together into higher level categories. Further, people expect human-like features in future robots but the expected human-likeness varies considerably across application domains. Discovering the optimal match between application domain and specific level(s) of human-likeness is therefore a promising and necessary direction of future research.

## REFERENCES


Ackerman, E. (2017, January 6). Why every social robot at CES looks alike. *IEEE Spectrum. Retrieved* from http://spectrum.ieee.org/tech-talk/robotics/home-robots/ces-2017-why-every-social-robot-at-ces-looks-alike

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, *25*, 351-370.

Graaf, M. M. A. de, & Ben Allouch, S. (2015). The evaluation of different roles for domestic social robots. *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication*, 676-681.

Graaf, M. M. A. de, Ben Allouch, S., & Dijk, J. A. G. M. van (2017). Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction.* Vienna, Austria, March 6-9.

Ezer, N. (2008). *Is a robot an appliance, teammate, or friend? Age-related differences in expectations of and attitudes towards personal home-based robots* (Doctoral dissertation, Georgia Institute of Technology). Retrieved from ProQuest.

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. *Proceedings of the 12th IEEE International Symposium on Robot and Human Interactive Communication*, 55-60.

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008). Understanding social robots: A user study on anthropomorphism. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication*, 574-579.

Komatsu, T., Kurosawa, R., & Yamada, S. (2012). How does the difference between users’ expectations and perceptions about a robotic agent affect their behavior? *International Journal of Social Robotics*, *4*, 109-116.

Lee, K., Park, N., & Song, H. (2005). Can a robot be perceived as a developing creature?: Effects of a robot's long-term cognitive developments on its social presence and people's social responses toward it. *Human Communication Research*, *31*, 538-563.

Lohse, M. (2011). Bridging the gap between users' expectations and system evaluations. *Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication*, 485-490.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 460-469.

Pejic, I. (2016, February 4). Military analysis: U.S. UAVs, UGVs, autonomous weapon systems and military robotics. *South Front: Analysis & Intelligence.* Retrieved from https://southfront.org/military-analysis-u-s-uavs-ugvs-autonomous-weapon-systems-and-military-robotics/

Rızvanoğlu, K., Öztürk, Ö., & Adıyaman, Ö. (2014). The impact of human likeness on the older adults’ perceptions and preferences of humanoid robot appearance. *Proceedings of the International Conference of Design, User Experience, and Usability*, 164-172.

Schaefer, K. E., Sanders, T. L., Yordon, R. E., Billings, D. R., & Hancock, P. A. (2012). Classification of robot form: Predicting perceived trustworthiness. *Proceedings of the Human Factors and Ergonomics Society*, *56*, 1548-1552.

Sims, V. K., Chin, M. G., Sushil, D. J., Barber, D. J., Ballion, T., Clark, B. R., ... & Finkelstein, N. (2005). Anthropomorphism of robotic forms: A response to affordances?. *Proceedings of the Human Factors and Ergonomics Society*, *49*, 602-605.

Woods, S., Dautenhahn, K., & Schulz, J. (2004). The design space of robots: Investigating children's views. *Proceedings of the 13th IEEE International Symposium on Robot and Human Interactive Communication*, 47-52.