

Characterizing the Design Space of Rendered Robot Faces

Alisa Kalegina
University of Washington
Seattle, Washington
kalegina@cs.washington.edu

Grace Schroeder
University of Washington
Seattle, Washington
grs8@uw.edu

Aidan Allchin
Lakeside School
Seattle, Washington
aidan.allchin@lakesideschool.org

Keara Berlin
Macalester College
Saint Paul, Minnesota
kearaberlin@gmail.com

Maya Cakmak
University of Washington
Seattle, Washington
mcakmak@cs.washington.edu

ABSTRACT

Faces are critical in establishing the agency of social robots; however, building expressive mechanical faces is costly and difficult. Instead, many robots built in recent years have faces that are rendered onto a screen. This gives great flexibility in what a robot's face can be and opens up a new design space with which to establish a robot's character and perceived properties. Despite the prevalence of robots with rendered faces, there are no systematic explorations of this design space. Our work aims to fill that gap. We conducted a survey and identified 157 robots with rendered faces and coded them in terms of 76 properties. We present statistics, common patterns, and observations about this data set of faces. Next, we conducted two surveys to understand people's perceptions of rendered robot faces and identify the impact of different face features. Survey results indicate preferences for varying levels of realism and detail in robot faces based on context, and indicate how the presence or absence of specific features affects perception of the face and the types of jobs the face would be appropriate for.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*;

KEYWORDS

Social robots, Robot face design

ACM Reference Format:

Alisa Kalegina, Grace Schroeder, Aidan Allchin, Keara Berlin, and Maya Cakmak. 2018. Characterizing the Design Space of Rendered Robot Faces. In *HRI '18: 2018 ACM/IEEE International Conference on Human-Robot Interaction, March 5–8, 2018, Chicago, IL, USA*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3171221.3171286>

1 INTRODUCTION

The importance of faces to human survival is undeniable: we read faces to infer emotional states [24], follow gaze to derive intention

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HRI '18, March 5–8, 2018, Chicago, IL, USA

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-4953-6/18/03...\$15.00

<https://doi.org/10.1145/3171221.3171286>



Figure 1: Collage of rendered robot faces from the data set collected in this work (Sec. 3).

[14], and examine a face for markers of personality and traits [31]. Robots with social faces exploit this human capacity for reading faces to establish agency [7, 17], personality [7, 26], and traits [12]; communicate intent [23, 26, 33]; and make their internal state transparent [2, 6]. Historically, most social robots have static physical faces, like Pepper, or mechanically actuated expressive faces, like Kismet [6] or Simon [12]. While the benefits of an expressive robot face is undebatable for many applications of social robots, building mechanically actuated faces is challenging and adds significantly to the cost of the robot.

Instead, many recent commercial and research robots have faces that are rendered onto a screen. This trend is fueled by the availability and affordability of tablets and their ease of programming. Having the face rendered on a screen gives complete flexibility over its design. Furthermore, it allows for the face to be easily animated for blinking, eye gaze, and facial expressions.

Despite the prevalence of robots with rendered faces, there are no existing surveys analyzing the variations of faces that have been designed for different purposes. Little is known about how people perceive these faces, as most previous studies in this vein have focused on physical faces [11, 16, 19]. In addition, there are no guidelines for designing such robot faces. Our work aims to fill that gap.

In this paper we first present a survey of 157 rendered robot faces (Fig. 1)¹ and an analysis based on 76 attributes of these faces. Next we present findings from two questionnaires that measured people's perception of (a) selected existing faces, and (b) synthesized faces that differ by only one feature from a reference face. We find

¹Images included in this paper under ACM guidelines on Fair Use.

that participants preferred less realistic and less detailed robots in the home, but highly detailed yet not exceedingly realistic robots for service jobs. The lack of key features like pupils and mouths resulted in low likability ratings and engendered distrust, leading participants to relegate them to security jobs. Most of the robots across both surveys were seen as most fitting for entertainment and education contexts. Robots ranked high in likability were also often ranked highly in positive traits like friendliness, trustworthiness, and intelligence.

2 RELATED WORK

The impact of a robot's face within human-robot interactions has been repeatedly documented. However, most research so far has focused on physical and mechanical faces [2, 4, 13, 18, 25]. For instance, Powers et al. [26] demonstrated that modifying a robot's physical gender cues such as voice pitch and lip coloration altered participants' perceptions of the robot's personality, specifically on the dimensions of leadership, dominance, compassion, and likability. Many researchers have also investigated the impact of a robot's general appearance, to which a robot's face contributes in crucial ways [5, 11, 12, 19, 20, 28, 32].

The literature on the topic of digital robot faces is sparse and centered on examining three-dimensional human-like virtual heads [15, 16]. In one such experiment, Broadbent *et al.* measure the participants' attribution of agency to a robot that employed either a human-like face display, a silver face display, or a no-face display. The study suggested that even the presence of an "uncanny" silver face can promote perceptions of agency, as compared to a no-face display [7]. However, as recognized by the authors, many iterations of robot faces exist between the limits of having no face and having a human-like silver face—abstract geometric faces, cartoon-like two-dimensional faces, and three-dimensional non-human faces, for instance—thus leaving a wide range of faces unexamined. The first questionnaire in our paper (Sec. 4) extends their findings with an exploration of how a diverse set of rendered faces, ranging in human-likeness and detail, is perceived by people.

Also related to our work is research on virtual and animated character faces focusing the Uncanny Valley effect [27, 29] as well as perceived attributes of faces [10, 21, 22, 30, 31]. One study investigated the impact of viewing cartoon faces and demonstrated a cross-over effect between cartoon and human faces. After viewing videos of either an animated show featuring large-eyed cartoon characters or a live-action show featuring human actors, participants showed a marked increase in preference for human eyes that were larger than normal after having watched the cartoon [9]. This illustration of the influence of animated faces on perception of human faces reinforces the need for thoughtful design of a robot face, even if it is cartoon-like in character, especially if there will be repeated exposure to the robot.

3 A SURVEY OF RENDERED FACES

3.1 Methodology

We identified 157 individual robots with screen faces by first performing a basic web search for web, image, and video results using the keywords "robots with screens," "robot screen faces," "touchscreen robot," "smartphone robot," and "telepresence robot." Each

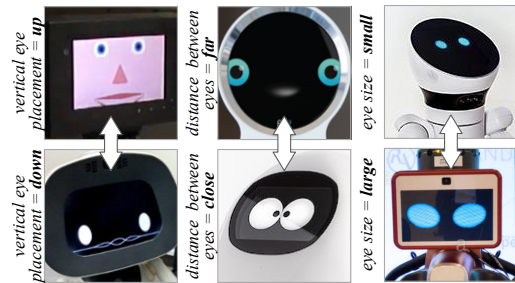


Figure 2: Faces that exemplify the two extremes of three continuous value features.

relevant result was explored and documented, using videos of the robot to gauge facial expressions and movement. This search was performed separately by multiple people, thus alleviating some bias in personalized search results. If a candidate robot was created by an organization featuring additional robots in its portfolio, those too were assessed for possible inclusion. Any candidate robots tangentially seen within an electronics conference video or in articles regarding such events were explored as well.

A host of robots were excluded from the data set due to insufficient data availability. Additionally, any robot utilizing a LED display in lieu of a full LCD one was not included, as the focus of the study is solely on screen-based robot faces. Robots which imitated a pixelated effect on a LCD display (Cozmo, Xibot) constitute an active design choice and were therefore retained. Fictional robots were also excluded from the data set.

3.2 Face dimensions

All faces in the dataset were coded across 76 dimensions. For the purposes of our data collection, "face" was defined as the top frontal portion of a robot that includes at least one element resembling an eye. The first 11 dimensions indicate whether a particular element is present on the face, *e.g.*, *mouth*, *nose*, *eyebrows*, *cheeks/blush*, *hair*, *ears*, *eyelids*, *pupils*, and *irises*. 19 dimensions record the color of these elements and the face, and any additional features such as *eyelashes*, *lips*, and *reflected eye glare*. Three dimensions record *eye size*, *nose size*, and *cheek size*, 14 dimensions indicate feature and face shape, and seven dimensions describe feature placement. Furthermore, we indicate whether any of the elements are animated or change in some way (*e.g.*, for facial expressions or blinking).

Also included are any physical features (*e.g.*, external ears) and embodiment properties, such as *screen type*, *screen size*, and *robot height*. The robot's embodiment *type* was coded as either humanoid, zoomorphic, or mechanical, in accordance with Rau *et al.* [19]. These categories encompass robots which, respectively, imitate human-like appearance in some way (adding a face, arms, *etc.*), imitate animal-like appearance (fur, animal face, animal body), or explicitly show mechanical parts (wires, wheels, treads) with an appearance dictated by the robot's function.

Properties that have continuous values are discretized based on the range and distributions observed in the data set. For example, *eye placement* can have the values *up*, *raised*, *center*, *low*, and *down*, which are all defined in relation to the center line of the face. Fig. 2 gives examples of faces to illustrate some of these features.

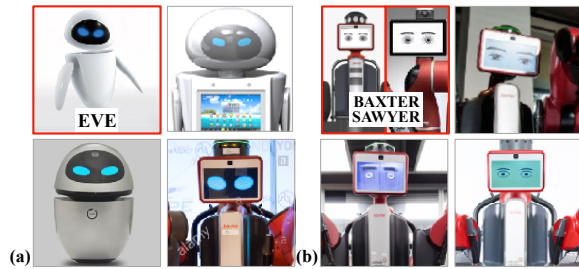


Figure 3: Example sets of rendered robot faces that mimic a popular face.

Three contextual dimensions were also recorded: the *year* of the robot’s make, its *region of origin*, and the *job category* of the robot (*i.e.*, what setting it is primarily made for).

3.3 Findings

3.3.1 *Summary Statistics.* The majority (33.8%) of the robots surveyed originated in the United States, with an additional 17.8% hailing from China, 10.2% from Japan, 8.3% from Korea, and 2.5% from Germany. All robots in the data set were created between the years 2001 and 2017, with majority concentrations occurring in 2015 and 2016 (22.4% and 21.8%, respectively). 23.6% of the robots were used primarily for research, 22.9% for entertainment (including toys), 21.0% for the home, and 14.6% for service (*e.g.*, waiting on tables, delivery, and receptionist). The most prevalent robot type classification was *humanoid* (60.5%), with 27.4% classified as *mechanical*, and 12.1% as *zoomorphic*.

3.3.2 *Feature distributions.* We analyzed the distribution of feature values over a pared down set of features that have no dependencies: *mouth, nose, eyebrows, cheeks/blush, hair, ears, face color, eye color, eye shape, pupil, lid, iris, eye size, eye placement, and eye spacing.* Properties like *iris color* that are only relevant for faces that have an *iris* were excluded. Properties of the eye were included since all faces were assumed to have eyes.

Overall, 34.4% of robots had a black face, 20.4% had a white face, and 14.0% had a blue face. 65.6% of the robot faces had a mouth, 40.8% had eyebrows, 21.7% had a nose, 21.0% had cheeks or blush, 8.3% had hair, and 3.8% had ears. The predominant eye color was white at 47.1%, followed by blue at 18.5% and black at 16.6%. 40.8% of eyes were circular, 28.0% were vertical ovals, and 11.5% were shaped similarly to the human eye. The majority (71.3%) of robots had pupils, while 65.6% had no eyelids and 59.9% had no irises. More than half of the robots (51.6%) featured eyes that took up between $\frac{1}{2}$ and $\frac{1}{20}$ of screen space, 38.2% had eyes centered on their face, and 43.9% of eyes were spaced evenly between the center and the edges of the face.

Eleven out of these 15 dimensions mentioned above had more than one feature value represented in 20% of robots. In order to analyze common feature value overlaps, we employed the Group By data aggregation function from pandas, a Python data analysis library. Few groups of faces had the exact same set of features; the largest group of faces with identical feature encoding had only three robots, representing the Baxter family of robots.

3.3.3 *Observed patterns.* We performed K-modes clustering with the Cao initialization [8] to identify common groups of faces. While



Figure 4: Example faces that include unique features represented in only one or two robot in our dataset.

some of the clusters were not coherent to the human eye, several were qualitatively distinct. One set of faces seemed to mimic the Disney robot character Eve Fig. 3(a). Another set, which were mostly rendered on Baxter Research Robot screens, were clearly influenced by the original Baxter and its cousin Sawyer Fig. 3(b). We also saw groups of very simple faces that only had two eyes with varying eye properties and details (*e.g.*, Otto in Fig. 5), as well as very complex, human-like faces that approximated full human renderings on a screen (*e.g.*, Valerie in Fig. 5).

3.3.4 *Unique features.* Some feature values that were represented in a very few faces in our dataset are worth mentioning as examples of creative design and demonstrations of the flexibility of rendered robot faces. Some examples shown in Fig. 4 are *dotted eyelashes, lower eyelashes, eye glasses, asymmetrical faces, pixelizations, speech bubbles, single eye, and sound wave shaped mouth.*

4 PERCEPTION OF RENDERED FACES

Understanding the dimensions of the design space helps designers discern the different possible variations of faces. However, it does not inform designers about how those variations might elicit different reactions from people. We conducted an online survey featuring faces from the dataset described in Sec. 3 to obtain empirical data that can inform the design of faces that elicit a particular response. Building upon Blow *et al.*’s work, which examined the perception of physical robot faces rooted in Scott McCloud’s triangular representation of the design space for cartoon faces [4], we chose rendered robot faces spanning across a subjective spectrum of realism and detail. The aforementioned K-modes clustering method was not employed in this analysis.

4.1 Questionnaire design

Twelve representative rendered robot faces were used in the survey (Fig. 5). These were hand-selected from the data set (Sec. 3) as to capture various points along the spectrum of realism and detail. While this spectrum is inherently subjective, we attempted to introduce a measure of objectivity by analyzing the number of binary features present on each face, *i.e.*, whether the face has a nose, eyebrows, cheeks, *etcetera*. The selected faces have from 0 to 10 missing features. A higher number of missing features indicates a lack of detail. Our analysis suggests that more detailed faces were also more realistic, and vice versa.

The robots included in the survey were: Aido, Buddy, Datouxi-fan (henceforth referred to as “Datou”), EMC, FURo-D, Gongzi Ziabai, HOSPI-R, Jibo, Otto, Sawyer, Valerie the Roboceptionist,

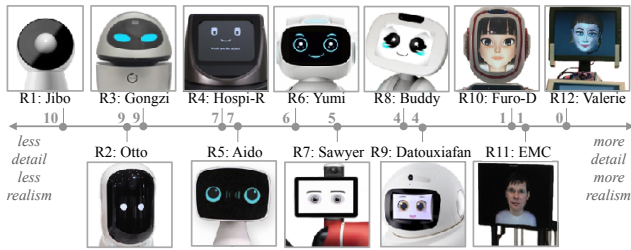


Figure 5: Faces used in the first survey in order of increasing detail. Number on the scale indicates the number of missing binary features on the face.

and Omate Yumi. More popular faces with higher quality images were preferred over others in the selection process.

The questionnaire presented participants with an image of a robot and asked them to rate the face across six 5-point semantic differential scales. Three of these scales were selected from the Godspeed questionnaires [1] (*Machinelike–Humanlike*, *Unfriendly–Friendly*, *Unintelligent–Intelligent*) and three were added to measure perceived trustworthiness, age, and gender (*Untrustworthy–Trustworthy*, *Childlike–Mature*, *Masculine–Feminine*). While a central tenet of the Godspeed questionnaires is that of increasing internal reliability, employing multiple indices in our questionnaire would have incurred increased survey fatigue in the participants due to the high number of faces, hence only one scale per measure was used. The order in which each scale appeared was randomized.

Participants were also asked to indicate how much they liked the robot’s face on a scale from 1 to 5 and asked to explain their answer in an optional free-form comment box. In addition, participants were asked to give each robot a short name, with the question providing the prompts of “robot with blue face” and “aggressive robot.” A final question asked participants to indicate which jobs or roles the robot would be most suitable for, selecting as many options as they felt were apt from the following list: *education, entertainment, healthcare, home, industrial (factory), research, service (hotel, restaurants, shops), and security (surveillance, security guard)*. The job options were a result of selecting the predominant types of jobs found in our data set and the types of jobs examined in previous work [16]. This set of questions remained the same for each of the 12 robot faces. The order of the faces was randomized for each participant.

The introduction to the questionnaire provided instructions on its structure, a link to a consent form, and contact information. The last page of the survey asked the participants to (optionally) provide demographic information: their birth year, gender, ethnicity, and level of education. The participants were also asked if any of the following selections applied to their previous experience with robots: *work with robots, study or have studied robots, own a robot, and watch or read media that includes robots*.

4.2 Data

The questionnaire was disseminated and administered through Amazon Mechanical Turk. The two criteria for eligible workers were having a HIT approval rate $\geq 97\%$, and an approved number of HITs ≥ 100 . The robot naming task was used as a control question in order to identify whether the participant was paying attention

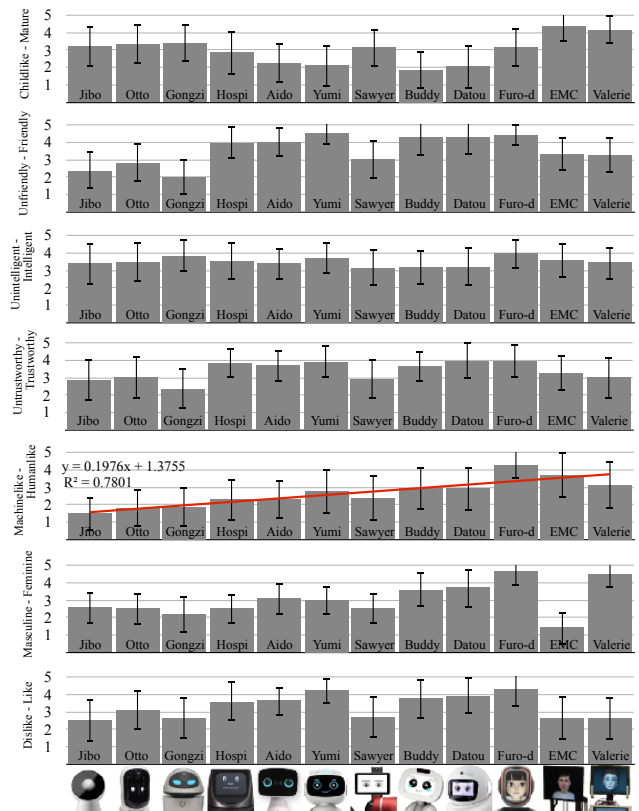


Figure 6: Ratings of the robot faces selected from our dataset on the six differential scales and one Likert scale question averaged across 50 participants (error bars indicate standard deviation).

to the image of the robot and thus providing relevant information. One participant’s work was rejected, due to them spending almost less than half the average time on each question and providing an overwhelming majority of “neutral” answers. The questionnaire stayed live until 50 completed surveys were approved. The average time per assignment was roughly 37 minutes.

The participant pool was comprised of 64.0% males and 36.0% females between the ages of 20 and 68. The ethnicity distribution was 72.0% White / Caucasian, 16.0% Asian or Pacific Islander, 6.0% Hispanic or Latino, 6.0% Black or African American. 81.7% of participants had some college education or higher, and 71.0% of participants were exposed to robots through media.

4.3 Results

Fig. 6 shows the ratings given by participants on the six differential scale questions and one Likert scale question. We performed paired *t*-tests (using Bonferroni correction for the number of hypotheses tested) for each pair of faces (66 pairs) on all dimensions; however, the full results are not displayed due to the high number of significant differences. Fig. 9(a) shows the distribution of participant votes on which jobs they thought were suitable for each face.

4.3.1 Friendliness. Yumi, FURo-D, Buddy, and Datou were perceived to be the friendliest robots. The latter three robots had relatively detailed faces, but were not exceedingly realistic, thus avoiding the Uncanny Valley effect. Jibo and Gongzi were perceived to be the most unfriendly.

4.3.2 Intelligence. The robots rated as most intelligent were FURo-D and Gongzi, while Sawyer, Buddy, and Datou were rated as least intelligent. Although these latter three robots were considered the least intelligent of the set, their ratings hovered around the “3 (Neutral)” mark; they were not overtly rated as “Unintelligent.”

4.3.3 Trustworthiness. Datou and FURo-D were deemed the most trustworthy, and Gongzi the least. Gongzi was frequently named “angry robot” or something to that effect (17/50), with respondents saying that it “seems almost mean”, “looks menacing”, and “this robot is intimidating, seems like it would be used by law enforcement.” Since the robot’s eyes are large they lend themselves to a pupil-less appearance; the respondents may have responded differently if pupils were present.

4.3.4 Human-likeness. The robot rated as most human-like was FURo-D, while Jibo was rated as the most machinelike. As robots increased in realism and detail, the ratings of human-likeness increased accordingly. A correlation of $R^2 = 0.75$ was observed between our subjective scale of realism and the measured human-likeness. Surprisingly the last robot on our spectrum, Valerie, was rated as significantly less human than the FURo-D ($p < .005$). This might be due to the different screen size and orientations; while FURo-D looks like a human wearing a helmet, Valerie is clearly a rendering of a floating human head on a larger screen.

Several respondents made explicit reference to viewing Jibo as some sort of mechanical device: “It just looks like a speaker, a bluetooth speaker”, “The robot looks like a satellite that connects to other devices that looks lifeless”, “looks like surveillance camera.”

4.3.5 Age. Buddy, Datou, and Yumi were deemed to be the most childlike: all three being the only cartoon robots with relatively detailed eyes and a smile. EMC and Valerie were rated as most mature, with multiple respondents including “man” as part of EMC’s name, and “lady” for Valerie’s.

4.3.6 Gender. EMC, the robot with the most explicitly male appearance, was rated as most masculine. FURo-D and Valerie were seen as the most feminine. Out of the set of robots that did not explicitly model the human appearance, Gongzi was considered to be the most masculine, which was surprising given the Eve character’s depiction as more feminine in the movie *Wall-E*. Since multiple comments regarding Gongzi described the robot as “angry” or “aggressive”, these traits which are often associated with males in many Western cultures may have influenced their ratings toward the “masculine” end of the scale. Datou and Buddy were seen as the most feminine. Several respondents noted Datou’s pink coloration (“Pink girl robot,” “Pink Faced robot”), which may have influenced their gender inference.

4.3.7 Overall preference. The robots with the highest likability were Yumi and FURo-D, while Jibo was the most disliked, alongside Gongzi, EMC, and Valerie. The four robots rated highest in likability were also the ones perceived to be the friendliest.

Jibo’s lack of features—indeed, it only has one eye—frequently flummoxed the respondents: “I don’t understand the face of it,” “There is not much that resembles a face,” “This isn’t a face. Not by any standards. It’s just a ball.” One key aspect to Jibo that is not represented in a static image is its dynamic emotive qualities, as signaled by its bodily movements and animated pupil.

EMC and Valerie’s low likability ratings appear to be influenced by the Uncanny Valley effect, with respondents stating: “This face [Valerie] looks too close to being a human face while also being far away enough to be creepy”, “The face [Valerie] is very creepy”, “The more realistic the faces are [EMC] the more creepy they look”, “This face [EMC] looks way too human. It seems like a somewhat odd cgi model that isn’t quite human yet. I don’t really like it for this reason and think it looks a little creepy.” Hence, the Uncanny Valley effect appears to have an influence even when the face is merely rendered on a screen and not housed in a human-like body.

4.3.8 Jobs. Robots most frequently picked for an education context, Yumi and Datou, were often described as being child-friendly; e.g., “It [Datou] seems like a perfect face to interact with children”, “She [Datou] is friendly and kids would love interacting with her”, “[Yumi] has a cuteness the kids would love.” Both of these robots had high ratings of friendliness and child-likeness, but were not deemed to be overly intelligent, possibly implying that sociability was the most important factor in selecting a robot for an education context. Some overlap appeared between the entertainment and education jobs: the five most frequently chosen robots for the entertainment category were also frequently selected for an education context.

The most unfriendly robot, Gongzi, was frequently selected for security jobs, while FURo-D and HOSPI-R were popular picks for service jobs. Since HOSPI-R’s face featured a line of text (“Would you like a Drink”) below the mouth, it most likely had a considerable effect on job selection, with multiple respondents giving it the name of “drink robot.” Both of these robots were actually created for a service context. Valerie, another service robot, was also most frequently assigned to the service category, possibly due to the presence of a receptionist headset featured in her picture. The research robot EMC was frequently picked for research jobs, alongside Otto. The robots embodying the least amount of realism and detail were the ones most frequently assigned to industrial and security jobs.

5 IMPACT OF FACE FEATURES

To characterize the impact of different features on people’s perception of the face, we conducted a second questionnaire using a controlled set of faces.

5.1 Questionnaire design

The second questionnaire used the same structure and set of questions as the previous one (Sec. 4.1); the sole difference between the two is the set of images used. In lieu of using real robot faces from the dataset, we created synthetic robot faces that were embedded in the same robot body. One of the faces, which we refer to as the *baseline*, was the *modal* face where each feature of the face has the value that is most common in the dataset. The rest of the faces in the set differed from the modal face on only one dimension.

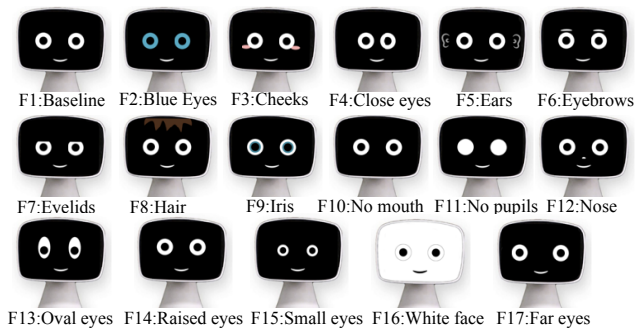


Figure 7: Faces used in the second survey. The face on the top left is the average face in our dataset and all other faces differ from it by one feature.

The set of dimensions that were changed included (i) the presence and absence of all face elements observed in our dataset (eyes, mouth, nose, eyebrows, cheeks/blush, hair, ears, eyelids, pupils and irises), (ii) the shape, size, and placement properties of the eye(s), and (iii) face color. For dimensions with more than one possible value, only values represented in more than 20% of all faces in our dataset were considered. To explore comparisons in feature values such as *eye color*, the second-most dominant feature value was used. In total, 17 different faces were used in this questionnaire (Fig. 7).

5.2 Data

The administration of the survey was identical to that of the previous survey. There were no rejected questionnaires and the average survey length was 44 minutes. The participant pool was comprised of 68.0% males and 32.0% females between the ages of 22 and 64. The ethnicity distribution was 80.0% White/Caucasian, 16.0% Asian or Pacific Islander, 10.0% Hispanic or Latino, 2.0% Black or African American. 84.0% of participants had college education or higher, and 82.0% of participants were exposed to robots through media.

5.3 Results

Fig. 8 presents the average ratings on the semantic differential scale questions and the Likert-scale likability question. Fig. 9(b) shows the distribution of participant votes on which jobs they thought were suitable for each face.

5.3.1 Friendliness. No faces were deemed significantly more friendly than the baseline face. The significantly less friendly faces were the ones lacking a mouth, lacking pupils, and possessing eyelids. The face with no mouth, rated as being most unfriendly, was frequently referred to as “creepy” by the participants, and that it gave an air of surveillance; *e.g.*, “[it] looks like it is watching my every move.” The face with eyelids may have suffered from an issue in design: the eyelids are depicted as lowered over the top portion of the eye, leading some responders to echo the sentiment that “cutting off the circular eyes makes it look suspicious.” All three of these robots were most frequently picked for security jobs.

5.3.2 Intelligence. The face deemed most intelligent featured eyebrows. The design of the eyebrows was such that they were lowered closer to the top of the eye, thus avoiding the intimation of a baby face: an effect that could induce the perception of increased

naivete and therefore of lower intellect [3]. This face was also rated as being the most mature.

The faces that were significantly less intelligent than the modal face were the ones with no mouth, closely spaced eyes, and cheeks. Several respondents referred to the closely spaced eyes face as being “dumb” and “goofy.” Accordingly, it was frequently relegated to the entertainment category.

5.3.3 Trustworthiness. No faces were ranked significantly more trustworthy than the baseline. If respondents interpreted trustworthiness to be equivalent to honesty, then it is possible that the symmetrical structure of the face and the large, even eyes could have played a role in their ranking, as those features have been illustrated to promote a perception of the face being honest [34].

The least significantly trustworthy faces were the same as the three rated as being least friendly: the face with eyelids, the face without a mouth, and the face without pupils. This phenomenon could be the inverse of a similar effect encountered by Li *et al.* [19]: the highly likable robots in their studies were consistently rated as being more trustworthy. Respondents expressed unease regarding the face with no pupils, frequently referring to it as being “creepy,” and giving it names akin to “dead eyes robot” and “soulless robot.” The face with eyelids was frequently referred to as “sly” and “smug.”

5.3.4 Human-likeness. The faces that were significantly more human-like than the baseline featured ears, eyebrows, hair, irises, and nose. While the robot with cheeks was rated as more human-like than baseline, it was not significant. These findings support the design recommendations of DiSalvo *et al.* [13]: increasing the complexity of the eyes (*e.g.*, adding irises) and having a face with four or more features increases the perception of humanness of a robotic head.

Faces without a mouth and without pupils were rated as significantly more machinelike, with the former eliciting the comment that “the fact that this robot has no mouth makes it seem very unemotional.” One consideration to make is that the stark contrast of a robot with no mouth, considered within a set of 16 robots that feature a smiling mouth, may have generated a stronger reaction in the responders.

5.3.5 Age. Robots that have cheeks, smaller eye distance, hair, and nose were perceived as significantly more childlike than the baseline robot face. The robots with eyebrows was perceived as significantly more mature than the baseline. Several respondents made note of the face with eyebrows – rated as most mature overall – appearing to be “older,” “elderly,” or as being “fun to spend time with for adults.” The evolutionary biology literature documents that infantile features, including large eyes, large head, and small mouth, evoke a nurturing response in observers [3, 30]. Our findings were partially consistent with that, although the robot with smaller eyes was not perceived as older.

A possible influencing factor of the nose face being seen as childlike may be the chosen design of the nose: several respondents made explicit reference to the nose being “cute,” “little,” and a “button nose,” with one respondent naming the robot “Buttons.” In her book *Reading Faces*, Zebrowitz notes that a pug or small nose is often considered an indicator of a baby face [34].

5.3.6 Gender. The robot with cheeks was perceived as significantly more feminine than the baseline. Since the design of the

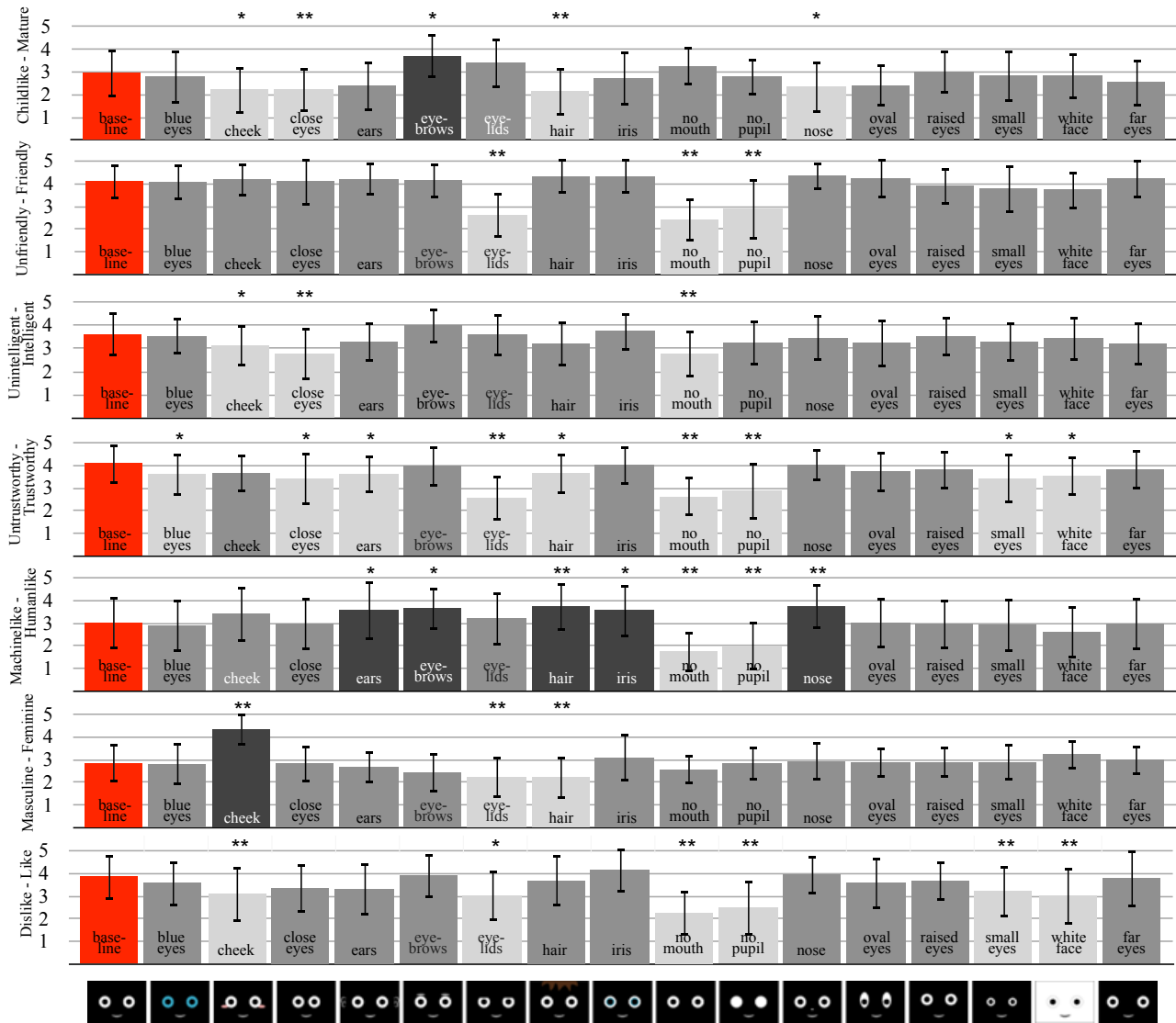


Figure 8: Ratings of the robot faces varied by one feature on the six differential scales and one Likert scale question averaged across 50 participants (error bars indicate standard deviation). Statistical significance between the baseline face (red) and each of the other faces based on paired *t*-tests with a Bonferroni correction for the number of hypotheses tested (16 for each scale), are shown with * for $p < 0.05$ and ** for $p < 0.005$.

cheeks was the appearance of pink blush, several respondents interpreted it as being makeup (e.g., “robot with eye makeup,” “happy face robot with pink eyeliner”), a concept traditionally associated with femininity. The face ranked second in femininity was the white face. The effect of it appearing more feminine could be attributed to the fact that women are biologically predisposed to have lighter skin than men [34].

Robots with eyelids and hair were perceived as significantly more masculine. Many respondents made reference to hair giving the robot the look of a male child (“Little Boy Robot”, “Kid Robot”) and noting its unkempt appearance (“disheveled robot”, “shaggy hair robot.”) This face was rated as being significantly childlike.

5.3.7 Overall preference. No faces were rated as significantly more likable than the baseline face, although the robot with irises was the most liked overall, with one respondent noting that “making the eyes a little more human with the color placement makes it feel quite friendly and approachable.” Robots with no mouth, no pupils, cheeks, small eyes, white face, and eyelids were significantly less likable than the baseline face, with the no mouth, no pupil, and eyelids faces receiving the lowest ratings of the set.

5.3.8 Jobs. The entertainment category was the most frequently assigned category overall, and the industrial category was the least frequently assigned. This trend indicates that the smiling, humanoid robot is deemed unfit for factory work by respondents but apt for entertainment, with the most machinelike robots receiving the

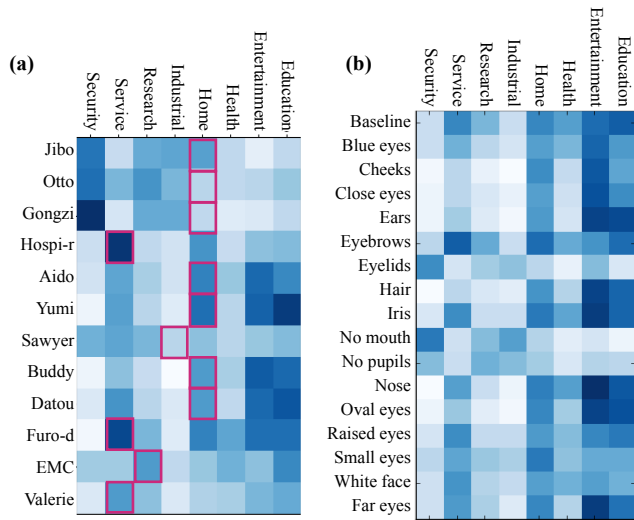


Figure 9: Frequency of responses in which the rendered robot faces in our (a) first and (b) second study were selected to be appropriate for different jobs or roles (dark blue: high frequency – white: low frequency). The actual jobs of the robots are indicated with squares.

highest frequency ratings in the category of industrial robot. Robots rated as most disliked and most unfriendly were overwhelmingly selected for security jobs. The robots most often chosen for the entertainment category had a nose, irises, and widely spaced eyes. The education category was most frequently chosen for the robot with ears, with respondents commenting that “it reminds me of a friendly teacher” and “the ears make the robot seem to be very good at listening.”

6 DISCUSSION

A similar effect emerged in both studies: the faces with no pupils and no mouth were consistently ranked as unfriendly, machinelike, and unlikable, and were overwhelmingly selected for security-type work. Respondents consistently cited surveillance for these types of robots: “This robot [no mouth robot] is kind of scary and just seems that it would be just watching as for security purposes,” “The bright eyes make the robot [no pupils robot] appear to be looking for everything and to be very observant,” “The robot [Gongzi] has angry eyes, possibly used for surveillance.”

Robots with pink or cartoon-styled cheeks were consistently ranked as feminine across both studies. The less detailed versions of these robots (Buddy, Datou, and robot with cheeks) were frequently rated as being childlike and friendly and were frequently selected for entertainment and education contexts.

Robots with somewhat detailed blue eyes (*i.e.*, eyes with at least a pupil), were frequently chosen for entertainment contexts and ranked as friendly and relatively trustworthy. Robots with mouths, especially in the form of a smile, were frequently relegated to entertainment and education categories across both studies.

Although most of the real robots in the first study were created for the home, they were more frequently placed within other contexts. Most of the robots chosen for the home category were of middling realism and detail, the most popular robot for the job being Yumi from the first study and the robot face with eyebrows from

the second study. Both of these robots had short, linear eyebrows placed just above the eyes. Multiple responders cited the simplicity of Yumi’s face being an attractive feature: “A very friendly robot. Reminds me of simple robots from the 80’s”, “I like this as a robot with machine and human like qualities”, “It maintains its role as a machine, but also emits a happy feeling.” These responses seem to indicate that, when it comes to placing a robot in their home, the respondents distrust highly realistic robot faces, and instead prefer a robot with several human-like features that imbue a feeling of sociability, while still explicitly remaining a machine. These results echo previous findings in which people preferred robots which are not fully realistic when interacting in a domestic setting [11, 32].

The most popular robot for the service category, FURo-D, is highly detailed but not fully realistic, thus managing to avoid the Uncanny Valley phenomenon. One respondent said that “it doesn’t try to go for a fully realistic approach, it stay[s] on a middle ground and makes it more friendly.” FURo-D was ranked as most human-like overall, and received high ratings in friendliness, intelligence, trustworthiness, and likability. These findings are in accordance with previous work [25, 32], which suggests that a high human-like rating for a robot correlates with higher rankings in sociability and intelligence. A possible reason for this robot’s frequent relegation to the service category is that the respondents assigned higher capabilities to the robot specifically because of its human-like appearance [7], and inferred that FURo-D embodies important qualities for service work: friendliness, intelligence, and trustworthiness.

Limitations. A key limitation of our study is that participants only looked at static images. With animated videos, a face without pupils may not appear as “soulless” if it is able to blink; or robots like Jibo could convey their emotive qualities through motion and thus be less reminiscent of a machine. The optimal examination of the effects of rendered faces would have participants interacting with varying robot faces in person, using a robot with a programmable face. Another limitation of this study is the potential co-dependency of features examined in Study 2. Future studies could examine the difference in people’s perception of robots with rendered faces in comparison with physical faces and examine the question of a robot’s perceived ethnicity.

7 CONCLUSION

Our work aims to characterize the design space of robot faces that are rendered on a screen and contributes the following:

- (1) A framework of 76 face features for specifying rendered robot faces and a dataset of 157 rendered robot faces coded in this framework.
- (2) Empirical findings on how people perceive a set of rendered robot faces varied on a scale of realism and level of detail;
- (3) Empirical findings on how individual face features impact people’s perception of a robot.

We plan to grow our data set dynamically as more social robots emerge on the market or in research publications. To that end, we created `robotfaces.org` which stores all robot face information in a database, provides up to date summary statistics about faces, and allows registered users to provide new entries by filling out a form, perform filtered searches, and download the latest data set in different formats.

REFERENCES

- [1] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. (2009). <https://doi.org/10.1007/s12369-008-0001-3>
- [2] Christian Becker-Asano and Hiroshi Ishiguro. 2011. Evaluating facial displays of emotion for the android robot Geminoid F. In *Affective Computational Intelligence (WACI), 2011 IEEE Workshop on*. IEEE, 1–8.
- [3] Diane S Berry and Leslie Zebrowitz Mcarthur. 1985. Some Components and Consequences of a Babyface. *Journal of Personality and Social Psychology* 48, 2 (1985), 312–323.
- [4] Mike Blow, Kerstin Dautenhahn, Andrew Appleby, Chrystopher L Nehaniv, and David Lee. 2006. The art of designing robot faces: Dimensions for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 331–332.
- [5] Cynthia Breazeal and Brian Scassellati. 1999. How to build robots that make friends and influence people. In *Intelligent Robots and Systems, 1999. IROS'99. Proceedings. 1999 IEEE/RSJ International Conference on*, Vol. 2. IEEE, 858–863.
- [6] Cynthia L Breazeal. 2004. *Designing sociable robots*. MIT press.
- [7] Elizabeth Broadbent, Vinayak Kumar, Xingyan Li, John Sollers 3rd, Rebecca Q Stafford, Bruce A MacDonald, and Daniel M Wegner. 2013. Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality. *PLoS one* 8, 8 (2013), e72589.
- [8] Fuyuan Cao, Jiye Liang, and Liang Bai. 2009. A new initialization method for categorical data clustering. *Expert Systems with Applications* 36, 7 (2009), 10223–10228.
- [9] Haiwen Chen, Richard Russell, Ken Nakayama, and Margaret Livingstone. 2010. Crossing the ÅTuncanny valleyÅT: adaptation to cartoon faces can influence perception of human faces. *Perception* 39, 3 (2010), 378–386.
- [10] Matthieu Courgeon, Stephanie Buisine, and Jean-Claude Martin. 2009. Impact of expressive wrinkles on perception of a virtual characterÅT's facial expressions of emotions. In *Intelligent Virtual Agents*. Springer, 201–214.
- [11] Kerstin Dautenhahn, Sarah Woods, Christina Kaouri, Michael L. Walters, Kheng Lee Koay, and Iain Werry. 2005. What is a robot companion - Friend, assistant or butler?. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*. <https://doi.org/10.1109/IROS.2005.1545189>
- [12] Carla Diana and Andrea L Thomaz. 2011. The shape of simon: creative design of a humanoid robot shell. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 283–298.
- [13] Carl F DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. 2002. All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, 321–326.
- [14] Chris D Frith and Uta Frith. 2006. How we predict what other people are going to do. *Brain research* 1079, 1 (2006), 36–46.
- [15] Rachel Gockley, Jodi Forlizzi, and Reid Simmons. 2006. Interactions with a moody robot. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 186–193.
- [16] Jennifer Goetz, Sara Kiesler, and Aaron Powers. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*. Ieee, 55–60.
- [17] F Hara and H Kobayashi. 1995. Use of face robot for human-computer communication. In *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on*, Vol. 2. IEEE, 1515–1520.
- [18] Ran Hee Kim, Yeop Moon, Jung Ju Choi, and Sonya S Kwak. 2014. The effect of robot appearance types on motivating donation. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 210–211.
- [19] Dingjun Li, PL Patrick Rau, and Ye Li. 2010. A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics* 2, 2 (2010), 175–186.
- [20] Manja Lohse, Frank Hegel, Agnes Swadzba, Katharina Rohlfling, Sven Wachsmuth, and Britta Wrede. 2007. What can I do for you? Appearance and application of robots. In *Proceedings of AISB*, Vol. 7. 121–126.
- [21] Karl F MacDorman, Robert D Green, Chin-Chang Ho, and Clinton T Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior* 25, 3 (2009), 695–710.
- [22] Rachel McDonnell and Martin Breidt. 2010. Face reality: investigating the uncanny valley for virtual faces. In *ACM SIGGRAPH ASIA 2010 Sketches*. ACM, 41.
- [23] Bilge Mutlu, Fumitaka Yamaoka, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 69–76.
- [24] Nikolaas N Oosterhof and Alexander Todorov. 2008. The functional basis of face evaluation. *Proceedings of the National Academy of Sciences* 105, 32 (2008), 11087–11092.
- [25] Aaron Powers and Sara Kiesler. 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*. ACM, 218–225.
- [26] Aaron Powers, Adam DI Kramer, Shirlene Lim, Jean Kuo, Sau-lai Lee, and Sara Kiesler. 2005. Eliciting information from people with a gendered humanoid robot. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*. IEEE, 158–163.
- [27] Jun'ichiro Seyama and Ruth S Nagayama. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and virtual environments* 16, 4 (2007), 337–351.
- [28] Dag Sverre Syrdal, Kerstin Dautenhahn, Sarah N Woods, Michael L Walters, and Kheng Lee Koay. 2007. Looking Good? Appearance Preferences and Robot Personality Inferences at Zero Acquaintance.. In *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*. 86–92.
- [29] Angela Tinwell, Mark Grimshaw, Debbie Abdel Nabi, and Andrew Williams. 2011. Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior* 27, 2 (2011), 741–749.
- [30] Alexander Todorov, Chris P Said, Andrew D Engell, and Nikolaas N Oosterhof. 2008. Understanding evaluation of faces on social dimensions. *Trends in cognitive sciences* 12, 12 (2008), 455–460.
- [31] Alexander Todorov and James S Uleman. 2003. The efficiency of binding spontaneous trait inferences to actorsÅT faces. *Journal of Experimental Social Psychology* 39, 6 (2003), 549–562.
- [32] Michael L Walters, Kheng Lee Koay, Dag Sverre Syrdal, Kerstin Dautenhahn, and Rene Te Boekhorst. 2009. Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. *Proc of New Frontiers in Human-Robot Interaction* (2009).
- [33] Yuichiro Yoshikawa, Kazuhiko Shinozawa, Hiroshi Ishiguro, Norihiro Hagita, and Takanori Miyamoto. 2006. Responsive Robot Gaze to Interaction Partner.. In *Robotics: Science and systems*.
- [34] L.A. Zebrowitz. 1997. *Reading faces: window to the soul?* Westview Press. <https://books.google.com/books?id=4fp9AAAAMAAJ>