Judging a Socially Assistive Robot by Its Cover: The Effect of Body Structure, Outline, and Color on Users’ Perception

ELA LIBERMAN-PINCU, YISRAEL PARMET, and TAL ORON-GILAD, Ben-Gurion University of the Negev

Socially assistive robots (SARs) aim to provide assistance through social interaction. Previous studies contributed to understanding users’ perceptions and preferences regarding existing commercially available SARs. Yet very few studies regarding SARs’ appearance used designated SAR designs, and even fewer evaluated isolated visual qualities (VQs). In this work, we aim to assess the effect of isolated VQs systematically. To achieve this, we first conducted market survey and deconstructed the VQs attributed to SARs. Then, a reconstruction of body structure, outline, and color scheme was done, resulting in the creation of 30 new SAR models that differ in their VQs, allowing us to isolate one character at a time. We used these new designs to evaluate users’ preferences and perceptions in two empirical studies. Our empirical findings link VQs with perceptions of SAR characteristics. These can lead to forming guidelines for the industrial design processes of new SARs to match user expectations.

CCS Concepts: • Human-centered computing → User studies; Empirical studies in interaction design;

Additional Key Words and Phrases: Socially assistive robots, visual qualities, user experience, taxonomy, design

ACM Reference format:
Ela Liberman-Pincu, Yisrael Parmet, and Tal Oron-Gilad. 2023. Judging a Socially Assistive Robot by Its Cover: The Effect of Body Structure, Outline, and Color on Users’ Perception. ACM Trans. Hum.-Robot Interact. 12, 2, Article 23 (April 2023), 26 pages. https://doi.org/10.1145/3571717

1 INTRODUCTION

The COVID-19 crisis affected the lives of many. People chose or were directed to stay home during lockdowns, except for essential activities. This new reality pointed toward a growing need for socially assistive robots (SARs) for various contexts, functions, and users, in the para-medical field, for domestic uses, supporting the elderly and people with disabilities, or helping in a variety of activities with children. SARs can be defined as the intersection of assistive robotics and socially interactive robotics, as they provide assistance through social interaction [1]. SARs may provide

This research was supported by Ministry of Innovation, Science and Technology, Israel (grant 3-15625), and by Ben-Gurion University of the Negev through the Helmsley Charitable Trust, the Agricultural, Biological and Cognitive Robotics Initiative, the W. Gunther Plaut Chair in Manufacturing Engineering, and by the George Shrut Chair in Human Performance Management.

Authors’ address: E. Liberman-Pincu (corresponding author), Y. Parmet, and T. Oron-Gilad, Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be’er Sheva 84105, Israel; emails: elapin@post.bgu.ac.il, {iparmet, orontal}@bgu.ac.il.

This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2023 Copyright held by the owner/author(s).
2573-9522/2023/04-ART23
https://doi.org/10.1145/3571717

ACM Transactions on Human-Robot Interaction, Vol. 12, No. 2, Article 23. Publication date: April 2023.
motivation, encourage training, and support rehabilitation to various users using nonphysical interaction [2]. Heerink [3] divided SARs into two subcategories: companions and service robots, and included functionalities such as a butler, a guide, and an information provider. Since there may be many functional differences among SAR types, user experience is expected to vary with the context of use, functionality, user characteristics, and environmental conditions. However, as we present in our targeted market survey, SAR manufacturers often designate the same embodiment for diverse contexts (i.e., one design fits all). Moreover, our literature review revealed a gap between designers’ and manufacturers’ perceptions and users’ perceptions in the question of what the user needs and wants [4].

During a new SAR’s development process, attention centers primarily on specifications that guarantee proper functionality and compliance with safety regulations. Unfortunately, technical and functional specifications usually do not fully consider user needs nor guarantee SAR acceptance or users’ attachment to the SAR. Hence, although researchers have acknowledged the difficulties of integrating user requirements into robotic designs for many years, we argue that now, with the market growth, it must be done (SAR presence grew by 44% in 2019 and continues to grow [5]). In the para-medical field, the adoption of SARs can significantly reduce the risk of infectious disease transmission to medical professionals by allowing them remote diagnosis, monitoring, and treatment of patients [6]. In the domestic arena, SARs can support people’s daily tasks within or outside their home, with chores, monitoring, providing companionship and supporting older adults, attending to the needs of the disabled, or helping with children. Since SARs can be classified in more than one way, based on the task and environment they perform in the work of Honig [7], their level of autonomy [1], or proactive and motion abilities [8], the industrial design of a SAR must account for these functional differences, also considering the users, the expected user experience, and the environment. The physical form of the SAR can help users understand its nature and capabilities [9]. Its appearances can differ by its human-likeness, structure, color, dimensions, and so on. Matching a robots’ appearance to its task/s can improve users’ acceptance [10] is a long agreed-upon statement, but how this can be done is less studied.

Industrial design is an essential part of the process that leads to the development of new products, devices, and objects that have not existed before. Industrial designers focus on the visual appearance, usability, and manufacturability of a product, to achieve a high-value product and enhance the user experience of end users. During the design process, the design team must make numerous decisions regarding the product’s aesthetics and how it is going to be used by its users. Designers create a message encoded in a product by its geometry, dimensions, textures, and other visual qualities (VQs), leading the user to cognitive, affective, and behavioral responses [11]. Designing a SAR brings additional variables to the equation, first related to human-robot interaction (HRI). As opposed to static objects, or other kinds of robots, SARs must interact with users, and the quality of the interaction impacts the user experience [12]. Second, SARs use motion as nonverbal communication means beyond visual appearance [13, 14]. Third, SARs operate in dynamic changing circumstances, which may impact the dynamic relationships with them and how they are perceived by users, even within domains.

Within the world of robotics, and SARs particularly, the suitable and desirable user perception may vary according to the robot’s context of use and the dynamics of the interaction; therefore, we previously proposed two models of relationships to lead to a better selection of VQs [15]. Onnasch and Roessler [16] had linked morphology (anthropomorphic, zoomorphic, technical) to robot tasks and human roles, suggesting different HRI scenarios deserve other embodiments. Other studies pointed to the need to adjust the robot’s display screen to its task, understanding the importance of establishing different user perceptions for different domains. Broadbent et al. [17] suggested using a human-like face display only in cases where the robot should be perceived as sociable and
amiable and avoid it in other cases. Kalegina et al. [18] explored the effect of 17 different faces on participants’ perception of the robot’s characteristics and role ascription.

Yet our literature review revealed that very few studies dealt with isolated visual features using designated purposefully designed SARs rather than existing off-the-shelf SARs. Most of the visual appearance studies follow the uncanny valley theory and focus on anthropomorphism, ranging from androids to cartoon-like robots [19–23]. Except for a few studies (e.g., [24, 25]), detailed design factors have rarely been investigated. None of the studies addressed the effects of body structure and inner proportions (e.g., screens or LEDs’ dimension and location).

Our current work aims to evaluate the effect of SARs’ VQs on user perceptions of the SAR’s characteristics. To do so, we first deconstruct the VQs attributed to SARs as detailed in Section 3. This stage is then followed by a reconstruction process. The reconstruction design ended with the creation of 30 new SARs that differ from one another by one VQ or more. These SAR design models allow us to isolate and evaluate each VQ. We then use these SAR models to assess users’ preferences and perceptions in two different studies, as detailed in the following sections. To narrow down the application scope and robot type, we chose to focus on one type of SARs, a machine-like personal assistant for home use.

2 RELATED WORK

2.1 SAR Design

Design plays an important role in creating trust in a robotic system. The visual aesthetics of a product form the user’s initial judgments (first impressions). Based on visual information, consistent first impressions can be formed within the first 39 milliseconds. In most cases, the first impression affects human behavior, attitude, and relationship with the product in the mid- and long term [26]. Studies have found that even the color alone has a psychological effect on the user by triggering the human arousal system and affecting product perception and trust in the product [27, 28]. Designers use the word desirability to describe how attractive the visibility of a product is and whether the product is perceived as “worth having or seeking” and as being useful, advantageous, or pleasing [29].

SARs designs can be generally categorized into human-like and androids, animal-like, and machine-like [30–32]. Studies have found that robots with a more human-like design were perceived as more intelligent, yet participants tend to prefer working with a less human-like social robot [33]. Studies conducted among the elderly population have found that the elderly prefer having a robot that looks like a familiar object for their home setting [34, 35]. Anthropomorphic robots were found to be less socially acceptable compared to machine-like robots [36, 37]. The dimensions of the SAR affect the way it is perceived; smaller SARs were found to be more attractive to older adults [35].

2.2 Product Personality and VQs

People tend to attribute personality to different objects using human-like characteristics such as credible, friendly, and professional. The product’s VQs determine the user’s first impression and contribute to the user’s emotions, associations, and perception of the product [38]. The VQs of a product may express its innovativeness or otherwise a perception of a familiar old friend and have four semantic functions: describing its purpose and the way it should be used, expressing its values and qualities, signaling the user to specific reaction, and identifying the product’s nature, origin, and product range, among others. [38]. The product’s human-like characteristics may help the users anticipate the product’s behavior and capabilities [39]. The product’s structure, color, and materials are different tools for the designer to achieve the desired look and feel and user ex-
perience to communicate the intended product character and affect the user’s interaction with the product [40, 41]. Earlier studies have found that people prefer products with a similar personality to their own, as these products may help them confirm and express themselves [39]. Earlier studies linked VQs to users’ perceptions of different products. Demirbilek and Sener [38] explored evoking emotions such as happiness and joy using the product’s proportions and colors. Perez Mata et al. [42] explored the effect of vases’ geometric features, such as curves, vertical-horizontal aspect ratio, and color, on the user’s perception of the vases as being beautiful, aggressive, expensive, dynamic, and so on.

In the field of HRI, a robot’s perceived personality and even its perceived gender affect the user’s expectations and help define the interaction. Studies have found that even subtle gender cues integrating into a robot’s embodiment can significantly affect the HRI and the perception of the robot’s identity, behavior, and suitable stereotypical tasks [43, 44].

3 DECONSTRUCTION OF ROBOTS’ VQS FOLLOWING A TARGETED MARKET SURVEY

For our targeted survey of robotic products, we used the Google Images search engine to collect images of 90 commercial robots using the search words ‘socially assistive robot’ (the search was conducted in May 2020). After collecting various images and videos of SAR models, we deconstructed them into building blocks by visual components. The deconstruction phase is similar to describing the robot’s appearance to someone who can’t see it; first, we describe the robot’s general appearance, its figure (i.e., Is it human-like? Pet-like? How realistic is it?), structure, colors, dimensions, and so on. Then we look at smaller details and the use of the SARs’ main components in the design (i.e., how the designer used the screen or the wheels in the design). The deconstruction process revealed a variety of different elements in the designs of existing SARs. Together with the literature review, we identified recurring and noticeable elements that are likely to affect users’ perceptions and behaviors at two VQs levels: (1) general appearance and (2) main components use, as detailed in the following and summed in Figure 1. Naturally, some exceptional robots could not fit one or more classifications (e.g., unique color or shape). Nevertheless, the following sections represent the most common VQs found in our survey.

3.1 General Appearance: The Packaging

Morphology. SARs’ visual appearance can be classified based on their human likeness. For this, we classified it into three groups as common in the HRI literature: human-like, pet-like, and unrecognized (including machine-like). Human-like and pet-like feature an inner classification from very realistic to more abstract models (e.g., a highly detailed figure of a dog or an abstract representation of a donkey).

Shape and structure. The shape of the body can be classified by five models of structure (rectangle, A shape, V shape, hourglass, and diamond) based on the classification of human body shapes [45], as well as by a division into one whole unit as opposed to two separate units (considering the wheels and screen). Shapes affect people’s perceptions in many design and engineering disciplines [46]. Squares are considered to represent stability, trust, honesty, order, conformity, security, equality, masculinity, maturity, balance, and stubbornness, whereas rectangles are considered to represent action, aggression, energy, sneakiness, conflict, tension, masculinity, and force. Circles are thought to represent completeness, gracefulness, playfulness, comforting, unity, protection, childishness, innocence, youth, energy, and femininity [47]. The orientation of the shape also affects the perception; pointing-down triangles (V shape) are perceived as more threatening than pointing-up triangles (A shape) [48, 49].

ACM Transactions on Human-Robot Interaction, Vol. 12, No. 2, Article 23. Publication date: April 2023.
Judging a Socially Assistive Robot by Its Cover

Outline. We identified two groups: rounded-edge and chamfered-edge. Previous studies have found that people tend to prefer rounded-edge objects compared to more angular in various contexts [50–52].

Dimensions. The dimensions of the SAR affect the way it is perceived [35]. In general, SARs’ dimensions can be placed on a scale from pet size to human size.

Material. Materials can be described by the material itself (e.g., plastic, synthetic fur) or on a scale from hard to soft materials. A study conducted to evaluate the effect of material choice on user perception [25] is a good example of isolating one feature. This study found that wood and fur rated higher on the aspect of warmth than plastic did. Fur was perceived as more discomforting than wood.

Color. Commercial robots can be found in various color schemes; still, in our targeted market survey, three main color schemes emerged: (1) mainly white with a small area of light gray and an additional light color, (2) white with different shades of blue, and (3) dark colors like black, dark gray, and dark blue. Color is the most noticeable design characteristic at first sight [53]. Color is well known as an arouser of various psychological reactions and feelings; the effect of color has been investigated in many disciplines, such as psychology, architecture, and website design. Studies on user preferences for colors have found inconsistent results for several reasons;

Fig. 1. A classification of general appearance components of SARs.
first, color must be associated with a particular object [54], and second, color preferences are not universal and vary across individuals and groups [55]. Figure 1 illustrates our findings and presents a preliminary classification of general appearance components.

3.2 The Use of Components

Two recurring components were found in most of the designs: a screen and wheels. The design team can use these components in different ways expressing varied characteristics. The screen is often used as the robot’s head [56]. Hence, it has the potential to be an essential tool to create the SAR’s character; its orientation, location, and relative dimensions can affect users’ perception. Previous studies investigated the effect of the human-like robot’s head design and suggested that features such as the head shape, dimensions, and head-to-body proportions heavily influence the users’ perception of the robot’s personality [24, 57–59].

Wheels are an essential feature of mobile robots since they are easier to design and program than treads or legs, and in addition, in most cases, much cheaper [60]. The robots’ wheels can be classified by their type (steering wheel, caster wheel, etc.), the number of wheels, and the platform configuration [60, 61]. There is not much in the literature regarding wheels’ design impact on users’ perception, but it was found that the use of wheels rather than legs reduces human-likeness [62]. In our survey, we classified the use of the wheels in the design into three groups: hidden, shown, and emphasized, assuming these groups have the potential to affect the overall look of the SAR and the user’s perception. Figure 2 illustrates our findings and presents a preliminary classification of the use of main components.

3.3 A Taxonomy Based on the Deconstruction of Components of SARs

Our targeted market survey explored the design space of commercial SARs. Using a deconstruction process, we isolated common robots’ VQs and created a preliminary taxonomy. Figure 3 summarizes the deconstruction phase and presents a preliminary taxonomy of VQs and the components to be later systematically evaluated. We then present two evaluation studies exploring three fundamental VQs: structure, outline, and color.

4 BUILDING BOXES FOR SARS, RECONSTRUCTION, AND EVALUATION STUDIES

Two studies, using online questionnaires, were run in parallel between August and October 2020, during the COVID-19 lockdown. To help participants relate to the SARs, we defined a context of a personal assistant to the elderly population in times of social isolation (during the COVID-19
outbreak) that can offer support with the user’s daily tasks, communicating with family and friends, tracking vitals, and more. First, we describe the building boxes for the SAR designs and models, which are common to both studies. We then describe the measurement and assessment tools. Study 1 aimed to evaluate the effect of VQs on users’ perception of the SAR, whereas study 2 explored users’ preferences when choosing the building boxes and assembling their own SAR. Three basic VQs were selected for the investigation: body shape, outline, and color (as detailed in the following section).

4.1 Study Design

4.1.1 Creation of Building Boxes for New SARs Models. We created our own 30 (5 × 2 × 3) abstractive SAR designs using the deconstructed VQs, allowing us to isolate one character at a time (shape, color, outline). A 3D detailed design using a CAD software model was created for each abstractive design. All models are the same height, having four small wheels that are used to support their movement around the house, and all contain a large horizontal screen allowing watching videos, conducting video chats, and playing, among others. Table 1 details the selected VQ levels; Figure 4 shows the building boxes and three examples of the self-constructed 3D robotic figures. These designs were used as the building boxes for our two studies to evaluate the effect of each specific factor on users’ perceptions and preferences.

4.1.2 Tools and Materials. Measuring attitude toward robots and technology. Two standard questionnaires were used for measuring attitudes toward robots and technology: the Negative Attitudes towards Robots

---

Table 1. Deconstructed VQs Selected for the Evaluation Studies

| Visual Quality | Levels                                                                 |
|---------------|----------------------------------------------------------------------|
| Body structure| Rectangle, A shape, V shape, hourglass, diamond                      |
| Outline       | Rounded, chamfered                                                    |
| Color scheme  | White and blue, mainly white, dark                                   |

---

Fig. 3. A preliminary taxonomy of VQs.
Scale (NARS) [63] and Technology Adoption Propensity (TAP) [64]. NARS consists of 14 statements classified into three subordinate scales: S1: “Negative Attitude toward Situations of Interaction with Robots,” S2: “Negative Attitude toward Social Influence of Robots,” and S3: “Negative Attitude toward Emotions in Interaction with Robots.” TAP is a multiple-item scale developed to measure consumers’ propensities to adopt new technologies. It uses a 14-item index that combines assessments of consumers’ positive and negative attitudes toward technology.

Measuring design perception. Perception of visual appearance can be measured in different ways: ascribing product reaction words to different designs [29], using semantic differential scales where respondents are asked to indicate their position on a scale between two bipolar words, or using Likert scales where subjects are asked to state their agreement with different statements. In the world of robotics, two common questionnaires are used for the evaluation of users’ perceptions: Godspeed [65] and the Robotic Social Attributes Scale (RoSAS) [66]. Godspeed measures five distinct dimensions: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety, using a semantic differential scale. RoSAS defines three central evaluation factors: warmth, competence, and discomfort using a Likert scale. Based on these standard questionnaires and our preliminary studies, we create an alternative set of describing words that is more suitable for non-human-like robots to be used in our following studies: friendly, childish, innovative, threatening, intelligent, reliable, professional, massive, elegant, medical, and old-fashioned. In our preliminary studies, we tried two different measurement approaches. The first approach utilized selecting suitable words out of a word bank; participants only chose the words they found to match their perception of the design—either yes or no—with no levels of agreement [67, 68]. In the second approach, using a 5-point Likert scale, participants were asked to react and provide their level of agreement for each describing word (e.g., how well does the adjective friendly describe the above robot? Very little, little, medium, high, or very high) [15]. We concluded that a Likert scale would be the most appropriate for our studies.

4.2 Research Questions
We aim to investigate the effect of VQs on users’ perceptions and preferences of a SAR. The following questions were addressed: (1) What impact do the specific VQs, body structure, outline, and color scheme have on user perception? (2) Are there shared preferences regarding the VQs of personal assistant robots for home use? Three research hypotheses were formulated and are detailed next.

Hypothesis 1 [H1] SARs’ VQs affect user perception. Since there is not enough literature regarding the effect of isolated VQs in the field of robotic design, we based our hypothesis on previous studies in other design fields. Hsiao and Chen [69] found relationships between the object’s shape features
and user responses in three product categories: automobile, sofa, and kettle, representing large, medium, and small products.

[H1a] The body shape of the SAR affects the user’s perception. People respond to shapes, ascribing different characteristics to square, rectangle, and circle objects. [48, 49]. Therefore, we assume that links between the robot’s body structure and users’ perception of its characterization and gender ascription will be found.

[H1b] The SAR’s outline affects the user’s perception. Previous studies have found that people prefer rounded edges to more angular objects in various contexts [50–52]. We assume that respondents will prefer rounded-edge robots and that a definite positive link will be found between rounded-edge SARs and positive perception.

[H1c] The SAR color scheme affects user perception. Color triggers many psychological reactions and feelings in various disciplines. These reactions are also affected by context and personal preferences [54, 55]. We expect to find different attributes to each color scheme and identify differences in the preferences according to the personal preferences of the respondents.

[H1d] VQs that share similar perceptions will be selected together. The VQs of an object create its design language; we hypothesize that participants will choose VQs that share similar characteristics. This assumption is based on the authors’ previous findings in the medical design field, where the targeted market survey deconstruction revealed three common design languages composed of different color schemes and outlines [70].

Hypothesis 2 [H2] personal characteristics affect design perception and preferences of SARs. Many fields of design use personas as part of the development process. Personas refer to fictional persons created based on a survey of users and represent a group of users with common characteristics, attitudes, and behaviors regarding their interaction with a particular product or service [71, 72]. We hypothesize that a link between personas and design preferences will be found. Since, in the SAR field, most studies related to design preferences were using human-like and machine-like off-the-shelf robots, there is no clear way to evaluate VQ preferences but as a preferred choice out of several possible designs.

Hypothesis 3 [H3] positive perceptions will correlate with design preferences. Participants will tend to choose VQs related to positive perceptions (e.g., favor friendly over threatening).

4.3 Study 1: Evaluating the Effect of VQs on User Perception

4.3.1 Aim and Scope. In study 1, we used our self-designed SARs to evaluate adults’ perception of three VQs in the context of a personal assistant to the elderly population using an online questionnaire.

4.3.2 Method and Online Questionnaire Design. Thirteen robots out of the 30 designed robots were chosen. The selection was made so that comparisons could be analyzed for each VQ. All building boxes were maintained, but the number of alternatives was reduced to avoid lengthy questionnaires and the abandonment of participants. Figure 5 shows the 13 selected designs and illustrates the possible comparisons. For instance, the white and blue color scheme can be compared across four body structures and so forth.

Using Qualtrics, we designed an online questionnaire constructed of two parts: (1) personal characteristics, in which participants were asked to fill in demographic information, TAP, and NARS questionnaires, and (2) user perception, which contained a randomly selected 7 out of the 13 robots. Showing one image at a time, robots were presented from the same angle and dimensions using similar backgrounds. Following each image, respondents were asked to describe their reactions toward the robot’s design based on their first impression using a list of 11 describing words (friendly, childish, innovative, threatening, intelligent, reliable, professional, massive, elegant, medical, and
Fig. 5. Selected designs for user perception questionnaires. Thirteen robots were rendered, and the lines illustrate how their distribution covered the possible combinations of body shape (rectangle, A shape, V shape, hourglass, and diamond), color scheme (white and blue, mainly white, and dark), and outline (chamfered or rounded).

Fig. 6. Right: Workflow of the perception questionnaire. Left: Example of a robotic image, as seen in the questionnaire (in this case, a mainly white, A-shaped rounded-edge model).

old-fashioned) and a 5-point Likert scale. In addition, they were asked if they could ascribe gender to the robots. Note that in Hebrew, the word robot is assigned to the male gender; hence, there is an inherent bias toward ascribing the male gender to a robot. Images and describing words were shuffled within the questionnaire. Figure 6 shows the questionnaire structure and one example of a robot’s image.

4.3.3 Participants. The online questionnaire was distributed between August and October 2020 (in the midst of the COVID-19 restrictions), using social media (Facebook and WhatsApp) as a rolling snowball where participants were asked to further refer and post. In total, data from 160 adult respondents were collected (54% female and 46% male), 18 years and older (M = 33.6 years, SD = 13.8).
Table 2. VQs' Effect on Self-Designed SAR Characteristics

| Structure | Friendly | Childish | Innovative | Threatening | Old-Fashioned | Massive | Elegant | Medical | Robot Gender |
|-----------|---------|----------|------------|-------------|---------------|--------|---------|---------|-------------|
| Color     |         |          |            |             |               |        |         |         |             |
| Outline   |         |          |            |             |               |        |         |         |             |

Significance level p < .05

Note: Dark boxes represent significant effects.

Table 3. Predicted Impact of the Structure: Estimated Means

| Rectangle (R) | A Shape (AS) | V Shape (VS) | Hourglass (H) | Diamond (D) | Post Hoc Comparisons with Bonferroni Correction |
|---------------|--------------|--------------|---------------|-------------|-----------------------------------------------|
| Mean SE       | Mean SE      | Mean SE      | Mean SE       | Mean SE     | Mean SE |
| Massive       | 2.1 0.065    | 2.2 0.055    | 2.2 0.051     | 2.0 0.052   | 2.1 0.066 | VS,AS>>H |
| Friendly      | 1.8 0.066    | 2.0 0.056    | 1.7 0.052     | 2.0 0.053   | 1.9 0.068 | AS,H>>VS |
| Threatening   | 1.3 0.054    | 1.4 0.046    | 1.5 0.043     | 1.3 0.043   | 1.3 0.055 | VS>>R,H,D |
| Elegant       | 2.0 0.067    | 2.1 0.058    | 2.2 0.053     | 2.3 0.054   | 2.2 0.069 | H>>R |
| Innovative    | 2.0 0.063    | 2.3 0.051    | 2.3 0.051     | 2.4 0.053   | 2.2 0.064 | H,V,S,AS>>R |

(>) signifies p < .05, (>>>) signifies p < .01.

Note: Only significant comparisons are presented in the right column.

4.3.4 Analysis and Design. The original Likert scale in the questionnaire consisted of five ticks (levels of response). For some of the characteristics, there was less than 4% of the responses in the extreme cases (e.g., old-fashioned, childish, threatening (tick 5)) and equal to or less than 8% for the not at all (e.g., innovative, reliable, professional, wise (tick 1)). We therefore decided to adopt a three-tick Likert scale—negative, neutral, and positive—where ticks 4 and 5 and ticks 1 and 2 were merged. We are aware of arguments that this type of merging should not be done [73], but we argue that in the case of design, we need to take a more lenient approach—that is, even if the effect is marginal or less dominant (i.e., middle ticks), it is better to include and consider it in the design than ignore.

A mixed within-between design was used where each characteristic was defined as a dependent variable. A general linear mixed model analysis was conducted to determine which VQs and user characteristics significantly influence each perceived characteristic. Robot-id and participant-id were included in the model as random effects to account for individual differences among respondents and design differences among robots. Fixed effects were the participants’ age and gender and the robots’ VQ: structure, color, and outline.

4.3.5 Results. All three VQs were found to affect user perception of the SAR. Each VQ affected different characteristics perceptions; adjusting these VQs may help the designer present the SAR as being massive, friendly, threatening, elegant, childish, old-fashioned, medical, and innovative. Furthermore, we have found a significant effect on the perception of robot gender. Table 2 summarizes the results. Details are given in the following paragraphs. Estimated means are given in Tables 3 through 5, respectively, for structure, color, and outline. Ranges are from 1 to 3, with values closer to 3 indicating a higher probability that this robot would be assigned this characteristic.

Structure. The structure significantly affected user perception of five characteristics: friendly (F(4, 1054) = 2.97, p = .019), innovative (F(4, 1054) = 4.4, p = .002), threatening (F(4, 1054) = 
Table 4. Predicted Impact of Color Scheme: Estimated Means

|                  | White and Blue (WB) | Mainly White (MW) | Dark (D) | Post Hoc Comparisons with Bonferroni Correction |
|------------------|---------------------|-------------------|----------|------------------------------------------------|
| Massive          | Mean 2.0, SE 0.046  | Mean 2.1, SE 0.048| Mean 2.2, SE 0.043| D>MW>>WB                                             |
| Friendly         | Mean 2.0, SE 0.047  | Mean 1.9, SE 0.048| Mean 1.7, SE 0.044| WB>MW>>D                                             |
| Threatening      | Mean 1.2, SE 0.039  | Mean 1.3, SE 0.04  | Mean 1.5, SE 0.036| D>MW>>WB                                             |
| Elegant          | Mean 2.0, SE 0.051  | Mean 2.2, SE 0.052 | Mean 2.2, SE 0.044| MW,D>>WB                                             |
| Childish         | Mean 1.4, SE 0.035  | Mean 1.3, SE 0.035 | Mean 1.2, SE 0.03 | WB>MW>D                                              |
| Medical          | Mean 2.2, SE 0.046  | Mean 2.3, SE 0.046 | Mean 1.9, SE 0.04  | MW>>WB>>D                                            |

(>) signifies p < .05, (>>) signifies p < .01.

Note: Only significant comparisons are presented in the right column.

Table 5. Predicted Impact of the Line: Estimated Means

|                  | Sharp ($) | Rounded ($) | Main Effect |
|------------------|-----------|-------------|-------------|
|                  | Mean      | SE          | Mean        | SE          |             |
| Elegant          | 2.1       | 0.04        | 2.2         | 0.038      | R>>S       |
| Old-Fashioned    | 1.6       | 0.035       | 1.5         | 0.032      | S>>R       |
| Innovative       | 2.2       | 0.039       | 2.3         | 0.038      | R>>S       |

(>) signifies p < .05, (>>) signifies p < .01.

Note: Only significant comparisons are presented in the right column.

Fig. 7. Structure’s effect. A comparison of three robots sharing the same color and outline. Left: Robot designs #3, #5, and #9. Right: A radar chart of participants’ agreement to the describing words. The scale indicates the number of participants (out of 80) who indicated the robot has a specific characteristic (i.e., rated it 4 or 5 on the Likert scale).

2.375, p = .05), massive (F(4, 1054) = 3.23, p = .012), and elegant (F(4, 1054) = 4.34, p = .002). V-shaped robots were perceived as more massive and threatening than all others, along with being the least friendly. But they were also perceived as innovative. Rectangle robots were perceived as the least innovative and elegant but also as not threatening. A-shaped and hourglass robots were perceived as the friendliest. Hourglass robots were also perceived as the most elegant and innovative and the least massive. Figure 7 shows three robots (#3, #5, #9) sharing the same
color and outline. Designs #3 and #5 were both perceived as elegant and innovative, but #3 was perceived as more massive.

The body structure was found to affect the perception of gender; hourglass-shaped robots were perceived as more feminine than others (54% of the respondents addressed them as female), whereas V-shaped robots were perceived as most masculine (57% of the respondents addressed them as males) followed by the diamond shape (51%). Figure 8 shows the perception of gender ascribed by the participants.

Color. Color significantly affected user perception of six characteristics: massive (F(2, 1054) = 4.89, p = .008), friendly (F(2, 1054) = 7.95, p < .001), threatening (F(2, 1054) = 13.37, p < .001), elegant (F(2, 1054) = 3.62, p = .027), childish (F(2, 1054) = 13.92, p < .001), and medical (F(2, 1054) = 29.49, p < .001). Dark SARs were perceived as more massive and threatening than white and blue or mainly white ones, and less friendly, but also elegant and less childish. The combination of white and blue SARs was found to be perceived as the friendliest color scheme. Mainly white SARs were perceived as elegant and medical. Figure 9 demonstrates three sets of robot pairs sharing the same structure and outline.

In addition, color affects the perception of robots’ gender. Participants tend to ascribe the male gender to robots in general, but the white color scheme may affect their tendency toward the female gender, as shown in Figure 10.

Outline. Chamfered or rounded outlines affected the user’s perception of three characteristics: elegant, innovative, and old-fashioned. Chamfered edges made the robot be perceived as old-fashioned (F(1, 1054) = 4.83, p = .028), less innovative (F(1, 1054) = 5.18, p = .023), and less elegant (F(1, 1054) = 8.21, p = .004). Figure 11 demonstrates the outline’s effect by comparing two robots sharing the same structure and color. The chart presents differences in the perception of two more characteristics in addition to elegant, innovative, and old-fashioned that were found significant in the general linear mixed model: medical and friendly. This implies that a rounded outline is perceived as friendlier and more appropriate for the medical domain. Possibly, the outline is more effective than found in the model, and this was not reflected in the results due to few comparisons among the 13 chosen robotic designs.

4.3.6 Summary. We found that a designer can lead users to a perception of different robots’ characteristics. These findings support new SAR design processes by providing insights regarding the effect of structure, color, and outline. For example, to achieve the perception of a friendly SAR, a designer should consider using A-shaped or hourglass structures and avoid the V shape, choose
Fig. 9. Color’s effect. Three sets of robot pairs sharing the same structure and outline. The radar charts present participants’ agreement to the describing words. The scale indicates the number of participants (out of 80) who indicated that the robot has a specific characteristic (i.e., rated it 4 or 5 on the Likert scale).
light colors (a combination of white and blue is preferred), and avoid dark colors. Using rounded edges would make the SAR look more elegant and innovative, whereas chamfered edges would contribute to an old-fashioned device’s perception. However, a V-shaped structure and dark colors may be suitable for a more authoritative SAR; this appearance may be beneficial for tutoring robots and in cases where discipline is needed (e.g., a supervisor robot, an Inspection porter robot, or a gatekeeper robot).

To confirm these findings, we explored the two specific cases presented earlier to verify that these combinations are indeed perceived as expected. We searched for the robots perceived as the friendliest and robots perceived as the most threatening; we expected to find that combining VQs that stimulate the same perception would increase the perception of the specific characteristic. Our study’s average friendliness ranking was 2.7 out of 5: two of the 13 robots ranked highest as friendly—robot #11 and robot #12, both with an average ranking of 3.1 out of 5. Their appearance matched our expectations; robot #11 is a white and blue, rounded hourglass-shaped robot, and
robot #12 is a white, rounded A-shaped robot. The average ranking of threatening was 2 out of 5: one of the 13 robots ranked highest as threatening—robot #7 had an average of 2.4 out of 5. Its appearance matched our expectations; it is a dark, rounded V-shaped robot. Note that not all 30 combinations were presented and evaluated (see Figure 5), and this may be the cause for resulting in three rounded robots for the two categories. For example, there was no option for a dark, chamfered V-shaped robot. Table 6 presents the two friendliest models and the most threatening one.

### Table 6. Two Friendliest Models and the Most Threatening One

| Perceived characteristic | Robot      | Visual qualities |
|--------------------------|------------|------------------|
| Friendliest robots       | Robot #11  |                  |
|                          | Robot #12  | A shape          |
|                          |            | Rounded          |
|                          |            | Mainly white     |
| Most threatening robot   | Robot #7   | V shape          |
|                          |            | Rounded          |
|                          |            | Dark             |

### 4.4 Study 2: Exploring Users’ Preferences

#### 4.4.1 Aim and Scope.
In study 2, we sought to explore users’ preferences regarding body structure, outline (chamfered or rounded), color scheme, and the display screen–graphic user interface (GUI) of personal assistant robots for home use. We used our building boxes to let participants assemble their own personal SAR using an online questionnaire.

#### 4.4.2 Method and Online Questionnaire Design.
Using Qualtrics, we designed an online questionnaire constructed of three parts: (1) personal characteristics, in which participants were asked to fill in demographic information, TAP, and NARS questionnaires; (2) participants were asked to design their own personal assistant robot by choosing four elements that compose the robot’s appearance: the robot’s body structure (V shape, A shape, diamond shape, hourglass, or rectangle), outline (rounded or sharp), color scheme (dark, mainly white, and blue and white), and finally the
GUI utility icons or face display; and (3) participants watched a short animation presenting their final design. Figure 12 illustrates the preferences questionnaire structure.

4.4.3 Participants. The online questionnaire was distributed between August and October 2020, using social media (Facebook and WhatsApp) as a rolling snowball where participants were asked to further refer and post. In total, data from 110 adult respondents were collected (66% female and 34% male), 18 years and older (divided into three age groups). Participants were classified into two groups for each of the three subordinate scales of the NARS: low negative attitude toward robots with a score below the mean of 3.4 or high negative attitude toward robots for above mean scores [74]. Table 7 summarizes the participants’ information.

4.4.4 Results. Descriptive and chi-square analysis. Most of the participants preferred a V-shaped structure or rectangle over all three others. Although a chi-square test of independence (\(X^2(n) = XX\),

Table 7. Summary of Participants’ Information

| Total | N = 110 | 100% |
|-------|---------|------|
| Gender |         |      |
| Male   | N = 37  | 33.6%|
| Female | N = 73  | 66.4%|
| Age (years) |         |      |
| 18–35  | N = 36  | 32.7%|
| 36–55  | N = 38  | 34.6%|
| 56+    | N = 36  | 32.7%|
| TAP    |         |      |
| Very low | N = 0   | 0%   |
| Low    | N = 19  | 17.3%|
| Moderate | N = 44  | 40%  |
| High   | N = 36  | 32.7%|
| Very high | N = 11  | 10%  |
| NARS   |         |      |
| Negative attitude toward situations of interaction with robots | Low | N = 75 | 68.2% |
|        | High    | N = 35 | 31.8% |
| Negative attitude toward the social influence of robots | Low | N = 38 | 34.6% |
|        | High    | N = 72 | 65.4% |
| Negative attitude toward emotions in interaction with robots | Low | N = 46 | 61.8% |
|        | High    | N = 64 | 58.2% |
p = XX) showed that there was no significant association between gender and the selection of body structure, results showed that male participants tend to prefer a V shape over all other forms (46%), whereas within the female participants the most frequently selected structure was the rectangle (29%) followed by the hourglass structure and V shape (22%, each). Most participants (59%) chose to use a rounded outline in their designs, and a chi-square test of independence was performed to examine the relationship between gender and outline preferences. The relation between these variables was significant, $X^2(1, N = 110) = 4.4447, p = .035$. Male participants were more likely to choose a rounded outline than female participants (males = 73%, females = 52%); a chi-square test of independence showed no significant association, $X^2(2, N = 110) = 1.47, p = .48$, between age and outline design preferences.

Gender was found to have no effect on color scheme selection, $X^2(2, N = 110) = 0.252, p = .88$, and GUI selection, $X^2(1, N = 110) = 0.052, p = .82$; the most frequently selected color scheme was the white and blue combination (48%). Most of the participants preferred to use the icons screen (64%) rather than the face display regardless of their gender or age group. The NARS score significantly affected participants’ preferences regarding the robot’s GUI, $X^2(1, N = 110) = 98.337, p < .00001$; all participants with a low negative attitude toward the social influence of robots (S2) preferred the face display (100%), and most of the participants with a high S2 preferred the icons display (97%). TAP was found to have no effect on the participants’ preferences.

**Fitting a hierarchical cluster analysis.** Summarizing the descriptive results can bring us to the conclusion that an optimal design would be a rounded, white-and-blue V-shaped robot with an icons screen display, but this may be oversimplistic and nonrepresentative of the respondents’ preferences. We turned to further data-driven analysis, utilizing a hierarchical cluster analysis (HCA) to identify alternative clusters and detect how the robotic qualities group. Since all of our variables are nominal (categorical) and the common methods for unsupervised learning (cluster analysis) are suitable for continuous variables, choosing a method suitable for nominal variables is necessary. We used the nomclust function from the nomclust package for R. This function runs HCA with objects characterized by nominal (categorical) variables. In our analysis, we used the combination of the LIN index to measure similarity along with the mean linkage method as recommended in the work of Šulc and Řezanková [75]. Since both evaluation criteria (PSFM and PSFE) suggested working with three clusters (as noted in Figure A.1 of the appendix), we continued with three clusters for the analysis.

Attempts to characterize the three clusters with the help of four characteristics—body structure, outline shape, color scheme, and GUI—had brought us to the following clusters: cluster 1 is a white rounded A shape with a face display, cluster 2 is a white and blue rounded V shape with icons display, and cluster 3 is a white and blue chamfered rectangle with a face display. The mosaic plot (each rectangle height represents the percentage within a column (cluster)) in Figure 13 illustrates the characterization of the three clusters.

Descriptive statistics for examining the relationships between demographic variables of age and gender, and cluster affiliation, show that gender has no relation. All three clusters contain members of the three age groups, but cluster 2 has more participants at the ages of 36 to 55 years and fewer participants at the ages of 56 years and older compared to the two other clusters. The mosaic plot in Figure 14 illustrates the association between clusters’ affiliation and the demographic variables.

Descriptive statistics for examining the relationships between personality variables, S1: “Negative Attitude toward Situations of Interaction with Robots,” S2: “Negative Attitude toward Social Influence of Robots,” and S3: “Negative Attitude toward Emotions in Interaction with Robots” and cluster affiliation revealed that (S1) and (S2) have no relationship to the cluster affiliation. (S3) affects cluster affiliation; clusters 1 and 3 are composed of mainly participants with high levels of negative attitude toward emotions in interaction with robots. Cluster 2 is composed of high
and low levels with a slight tendency to a low level. The mosaic plot in Figure 15 illustrates the association between clusters’ affiliation and personality variables.

4.4.5 **Summary.** We have found differences in participants’ preferences regarding the design of SARs for home use. Some were related to gender, and some to participants’ negative attitudes toward the social influence of robots. Participants assembled their robotic model by choosing its body
structure, outline, and color scheme (out of 30 possible options); they also selected their preferred GUI. Exploring the chosen design outcomes revealed that 24 out of the 30 possible combinations (80%) were selected by at least two respondents. Two varieties share first place as the most popular robot design with 11 selections (10% of the respondents). The two differ in all three VQs; the first is built of a chamfered-edge rectangle using a white and blue color scheme, and the second is a mainly white rounded V shape. Although it may seem that these results show no tendency toward a specific VQ, it was found that most participants choose VQs that are perceived as friendlier over those that are perceived as more elegant or innovative. This finding can be insightful for designers on what to incorporate in their future design work. The results demonstrate the importance of involving the users in the design process, as noted by the HCA.

5 DISCUSSION

Most research in the field of SARs design focuses either on evaluating existing off-the-shelf SARs [4, 35, 76–78] or on using participatory design processes to develop a new solution [79, 80]. Among the studies that were using purposely designed SARs, the majority focused on anthropomorphism [10]. Very few studies dealt with isolated visual features using designated SARs [25].

To begin our research, we first conducted a targeted market survey to explore and analyze SARs’ VQs; we then deconstructed the design models into small building boxes for our studies. Our aim was to evaluate the effect of isolated VQs on user perception and explore users’ preferences. We selected three basic VQs—body structure, outline, and color scheme—in the context of personal assistant robots in the home domain. These three VQs were selected as they are related to users’ first impressions of the SAR. Based on the literature review, we hypothesized that these VQs would significantly affect users’ perception, as found in other design fields. We assessed users’ design preferences and perceptions of these three VQs utilizing our own building boxes and the ability to design the robots’ features from scratch. Based on two studies with a total of 270 participants (study 1, N = 160; study 2, N = 110), we demonstrate how VQs such as structure, outline, and color significantly affect users’ perception of the SAR’s characteristics and support hypotheses [H1a], [H1b], and [H1c]. These findings correlate with previous studies assessing the effect of structure, outline, and colors in different products [38, 42].

In addition, we found that preference differences relate to users’ gender and NARS score; these findings support hypothesis [H2]. Hypothesis [H1d] assumed that participants would choose VQs that share similar characteristics, and hypothesis [H3] assumed that positive perceptions would correlate with design preferences. However, the results did not support these two assumptions. We found that out of 30 possible combinations for the SAR’s design (excluding the GUI), 24 models (80%) were selected at least twice (no combination was selected only once). The two most popular varieties received 11 selections each (10% of the respondents). This implies there is no consensus among users regarding the appropriate appearance for SARs in the domestic environment, which may be caused by a lack of familiarity and experience with SARs.

We added the option of selecting the GUI to give the users an option to make the design more human-like if they wished to do so since the anthropomorphism level is one of the most frequent categorizations of robots and was found to affect users’ behaviors and attitudes in various contexts [81]. Our literature review and the market survey revealed that manufacturers often use the screen as the robot’s head to increase its anthropomorphism [56]. However, users tend to prefer less anthropomorphized robots in different contexts [33–37]. Hence, we found it interesting to explore users’ preferences when designing a new robot. The results were consistent with previous studies; most participants preferred to use the icons screen (64%, n = 70) rather than the face display. The participants’ preferences regarding the robot’s GUI were mostly affected by their NARS score;
all participants with a low negative attitude toward the social influence of robots (S2) preferred the face display, and most of the participants with a high S2 preferred the icons display. These insights may suggest that using a screen that can perform as a face in some cases (but not always) is important to gain acceptance and reduce anxiety in some populations.

Analyzing the results of study 2 (see Figure 13) brought us to the definition of three distinct clusters; each is suitable for a particular user group. We then cross-checked the results of the two studies to find the perceived characteristics for each cluster. Table 8 summarizes our three clusters of robots and their potential users. Note that each of the three clusters has at least one character perceived as friendly in the first study, implying that humans seek this specific character when

| Cluster | Cluster | Cluster |
|---------|---------|---------|
| 1       | 2       | 3       |
| Body Shape | Outline | Color Scheme | GUI |
| A-shape   | V-Shape  | Mainly white | Icons |
| Rounded   | Rounded  | White and blue | Face |
| Rectangle | Chamfered | White and blue | Icons |

Table 8. Three Clusters of Robots and Their Potential Users

Users' perception (According to study 1)
- Friendly
- Innovative
- Massive
- Old fashioned
- Massive
- Old fashioned
- Friendly

Gender
- Female/Male
- Female/Male
- Female

Age
- 56 and above
- 36-55
- 18 and above

S1
- Low
- Low
- Low

S2
- High
- High
- High

S3
- High
- Low
- High

Note: All images are a synthetic representation of the same features in various ages/genders. We used our first author’s facial image to generate all images.
designing SARs for the home environment. This correlates with previous studies that investigated SARs’ personalities and users’ preferences; in a systematic review by Broadbent et al. [82], it was found that participants seek more sociable robots. The perception of a SAR’s sociability may be caused by its behavior [83], its appearance, or its voice [84].

6 CONCLUSION, LIMITATIONS, AND FUTURE WORK

This article presents an evaluation of the effect of three basic VQs—body structure, outline, and color—on users’ perception of the SAR’s characteristics. Based on a targeted market survey of SARs and three preliminary studies, we created our own self-designed 30 SARs models using CAD software. These designs share most of their visual features (the same height, screen dimension and orientation, and wheels dimensions and location) and are divided by their body structure, outline, and color scheme. This allowed us to evaluate the effect of each isolated VQ using online questionnaires.

The article also offers empirical findings on how isolated VQs impact people’s perception of a robot’s characteristics: friendly, childish, innovative, threatening, old-fashioned, massive, elegant, medical, and gender. The results imply that to achieve user acceptance, designers must select VQs appropriate to the robot’s role and its desired characteristics; moreover, users’ preferences may vary by different personal factors. Hence, allowing the user to further apply minor design adjustments may increase their satisfaction. These conclusions, of course, can be relevant to more types of robots and products than SARs, particularly the ones we built here, but we find them extremely important, especially in the field of SAR design, as SARs aim to create social interactions.

We plan to continue deconstructing SARs’ VQs to expand our knowledge and contribution to set up design guidelines to assist designers in the design process of a new SAR. In addition, the investigated VQs should be further evaluated using different scales and figures of SARs, as these may affect the results.

A key limitation of our study is the fact that participants watched the images of the different SARs online. We believe that the effect of the robot appearance should also be tested with real interaction rather than online questionnaires; therefore, our following studies will evaluate the effect of SARs’ VQs on user behaviors involving live interaction and a real task. The nature of the study may also be the cause for the users’ preferences diversity; although we defined the context of a personal assistant robot for home use, without a real interaction, participants may have interpreted this description differently according to their needs, desires, and presumptions. Another limitation may be due to the language of the study; studies were conducted in Israel among Hebrew-speaking participants and via snowball recruitment. To make sure that the perception describing words’ translation to English has the same meaning, we followed a two-way translation procedure. However, chances exist that our results are sensitive to language or cultural differences.

To align these findings with our proposed relationships models [15], our subsequent studies will focus on evaluating user expectations in each context of use. What would be the relevant and desired characteristics of different SARs (by domains, e.g., healthcare, companion, educational)? These studies will explore different stakeholders’ perceptions as well as the effect of culture. These findings are expected to support the design process of new SARs depending on their context of use, their intended role, and their users, as well as form design guidelines for future SARs.
APPENDIX

Fig. A.1. Three clusters dendrogram of user preferences derived from the nomclust HCA.

REFERENCES

[1] D. Feil-Seifer and M. J. Matarić. 2005. Defining socially assistive robotics. In Proceedings of the 2005 9th International Conference on Rehabilitation Robotics (ICORR’05). IEEE, Los Alamitos, CA, 465–468.

[2] M. J. Matarić and B. Scassellati. 2016. Socially assistive robotics. In Springer Handbook of Robotics. Springer, 1973–1994.

[3] M. Heerink. 2010. Assessing Acceptance of Assistive Social Robots by Aging Adults. Ph.D. dissertation. Universiteit van Amsterdam.

[4] A. Lazar, H. J. Thompson, A. M. Piper, and G. Demiris. 2016. Rethinking the design of robotic pets for older adults. In Proceedings of the 2016 ACM Conference on Designing Interactive Systems. 1034–1046.

[5] International Federation of Robotics. 2020. Executive Summary World Robotics 2020 Service Robots. Retrieved May 5, 2021 from https://ifr.org/img/worldrobotics/Executive_Summary_WR_2020_Service_Robots.pdf.

[6] M. Tavakoli, J. Carriere, and A. Torabi. 2020. Robotics, smart wearable technologies, and autonomous intelligent systems for healthcare during the COVID-19 pandemic: An analysis of the state of the art and future vision. Advanced Intelligent Systems 2, 7 (2020), 2000071.

[7] S. S. Honig, T. Oron-Gilad, H. Zaichyk, V. Sarne-Fleischmann, S. Olatunji, and Y. Edan. 2018. Toward socially aware person-following robots. IEEE Transactions on Cognitive and Developmental Systems 10, 4 (2018), 936–954.

[8] T. W. Schulz, J. Herstad, and J. Terresen. 2018. Moving with style: Classifying human and robot movement at home. In Proceedings of the International Conferences on Advances in Computer-Human Interactions (ACHI’18). 188–193.

[9] J. M. Beer, A. Prakash, T. L. Mitzner, and W. A. Rogers. 2011. Understanding Robot Acceptance. Georgia Institute of Technology.

[10] J. Goetz, S. Kiesler, and A. Powers. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In Proceedings of the 2003 12th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN’03). IEEE, Los Alamitos, CA, 55–60.

[11] N. Crilly, J. Moultrie, and P. J. Clarkson. 2004. Seeing things: Consumer response to the visual domain in product design. Design Studies 256 (2004), 547–577.

[12] A. Kristoffersson, K. Severinson Eklundh, and A. Louifi. 2013. Measuring the quality of interaction in mobile robotic telepresence: A pilot’s perspective. International Journal of Social Robotics 5 (2013), 89–101.

[13] G. Hoffman and W. Ju. 2014. Designing robots with movement in mind. Journal of Human-Robot Interaction 31 (2014), 91–122.

[14] S. Chatterjee, Y. Parmet, and T. Oron-Gilad. 2022. Body language for personal robot arm assistant. International Journal of Social Robotics 14 (2022), 15–37.

[15] E. Liberman-Pincu, E. D. Van Grondelle, and T. Oron-Gilad. 2021. Designing robots with relationships in mind: Suggesting two models of human-socially assistive robot (SAR) relationship. In Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI’21). 555–558.
[16] L. Onnasch and E. Roesler. 2021. A taxonomy to structure and analyze human–robot interaction. *International Journal of Social Robotics* 13 (2021), 833–849.

[17] E. Broadbent, V. Kumar, X. Li, J. Sollers 3rd, R. Q. Stafford, B. A. MacDonald, and D. M. Wegner. 2013. Robots with display screens: A robot with a more human-like face display is perceived to have more mind and a better personality. *PLoS One* 8, 8 (2013), e72589.

[18] A. Kulegina, G. Schroeder, A. Allchin, K. Berlin, and M. Cakmak. 2018. Characterizing the design space of rendered robot faces. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 96–104.

[19] J. Zlotowski, D. Proudfoot, K.YGoeswaran, and C. Bartneck. 2015. Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics* 7, 3 (2015), 347–360.

[20] F. W. Tung. 2011. Influence of gender and age on the attitudes of children towards humanoid robots. In *Proceedings of the International Conference on Human–Computer Interaction*. 657–646.

[21] K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. Assif, and M. Blow. 2009. KASPAR—A minimally expressive humanoid robot for human–robot interaction research. *Applied Bionics and Biomechanics* 6, 3–4 (2009), 369–397.

[22] C. Bartneck and J. Forlizzi. 2004. A design-centred framework for social human–robot interaction. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN’04)*. IEEE, Los Alamitos, CA, 591–594.

[23] M. Blaurock, M. Čaič, M. Okan, and A. P. A. Henkel. 2022. A transdisciplinary review and framework of consumer interactions with embodied social robots: Design, delegate, and deploy. *International Journal of Consumer Studies* 46, 5 (2022), 1877–1899.

[24] C. F. DiSalvo, F. Gumperle, J. Forlizzi, and S. 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*. 321–326.

[25] L. Björklund. 2018. *Knock on Wood: Does Material Choice Change the Social Perception of Robots?* Master’s thesis. KTH School of Electrical Engineering and Computer Science, Stockholm, Sweden.

[26] A. N. Tuch, E. E. Presslauer, M. Stöcklin, K. Opwis, and J. A. Bargas-Avila. 2012. The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human–Computer Studies* 70, 11 (2012), 794–811.

[27] C. M. Conway, J. E. Pelet, P. Papadopoulos, and M. Limayem. 2010. Coloring in the lines: Using color to change the perception of quality in e-commerce sites. In *Proceedings of the International Conference on Information Systems (ICIS’10)*, 224.

[28] C. J. Auster and C. S. Mansbach. 2012. The gender marketing of toys: An analysis of color and type of toy on the Disney store website. *Sex Roles* 67, 7–8 (2012), 375–388.

[29] J. Benedek and T. Minner. 2002. Measuring desirability: New methods for evaluating desirability in a usability lab setting. *Proceedings of Usability Professionals Association* 57 (2003), 8–12.

[30] M. Lohse, F. Hegel, A. Swadzba, K. Rohlfing, S. Wachsmuth, and B. Wrede. 2007. What can I do for you? Appearance and application of robots. *Proceedings of AISB 7 (2007)*, 121–126.

[31] D. Li, P. L. Rau, and Y. Li. 2010. A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics* 2, 2 (2010), 175–186.

[32] G. Trovato, F. Cuellar, and M. Nishimura. 2016. Introducing ‘thomorphic robots.’ In *Proceedings of the 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. IEEE, Los Alamitos, CA, 1245–1250.

[33] N. Piçarra, J. C. Giger, G. Pochwatko, and J. Možaryn. 2016. Designing social robots for interaction at work: Socio-cognitive factors underlying intention to work with social robots. *Journal of Automation Mobile Robotics and Intelligent Systems* 10, 4 (2016), 17–26.

[34] J. Forlizzi, C. DiSalvo, and F. Gumperle. 2004. Assistive robotics and an ecology of elders living independently in their homes. *Human–Computer Interaction* 19, 1–2 (2004), 25–59.

[35] Y. H. Wu, C. Fassett, and A. S. Rigaud. 2012. Designing robots for the elderly: Appearance issue and beyond. *Archives of Gerontology and Geriatrics* 54, 1 (2012), 121–126.

[36] P. Dario, E. Guglielmini, C. Laschi, and G. Teti. 1999. MOVAID: A personal robot in everyday life of disabled and elderly people. *Technology and Disability* 10, 2 (1999), 77–93.

[37] K. O. Arras and D. Cerqui. 2005. *Do We Want to Share Our Lives and Bodies with Robots? A 2000 People Survey*. Technical Report. LSA.

[38] O. Demirbilek and B. Sener. 2003. Product design, semantics and emotional response. *Ergonomics* 46, 13–14 (2003), 1346–1360.

[39] R. Mugge, P. C. Govers, and J. P. Schoormans. 2009. The development and testing of a product personality scale. *Design Studies* 30, 3 (2009), 287–302.
[40] J. Khalaj and O. Pedgley. 2019. A semantic discontinuity detection (SDD) method for comparing designers’ product expressions with users’ product impressions. Design Studies 62 (2019), 36–67.

[41] L. E. Janlert and E. Stolterman. 1997. The character of things. Design Studies 18, 3 (1997), 297–314.

[42] Marta Perez Mata, Saeema Ahmed-Kristensen, Per Brunn Brockhoff, and Hideyoshi Yanagisawa. 2017. Investigating the influence of product perception and geometric features. Research in Engineering Design 28, 3 (2017), 357–379.

[43] Jaha Otterbacher and Michael Talias. 2017. S/he’s too warm/agentic! The influence of gender on uncanny reactions to robots. In Proceedings of the 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI’17). IEEE, Los Alamitos, CA, 214–223.

[44] J. Bernotat, F. Eyssel, and J. Sachse. 2017. Shape it—The influence of robot body shape on gender perception in robots. In Proceedings of the International Conference on Social Robotics. 75–84.

[45] D. Tama and Z. Ondoglan. 2014. Fitting evaluation of pattern making systems according to female body shapes. Fibres and Textiles in Eastern Europe 22, 4 (2014), 107–111.

[46] S. J. Stroessner, J. Benitez, M. A. Perez, A. B. Wyman, C. M. Carpinella, and K. L. Johnson. 2020. What’s in a shape? Evidence of gender category associations with basic forms. Journal of Experimental Social Psychology 87 (2020), 103915.

[47] E. Fogelström. 2013. Investigation of Shapes and Colours as Elements of Character Design? Upssala Universitet.

[48] C. L. Larson, J. Aronoff, and E. L. Steuer. 2012. Simple geometric shapes are implicitly associated with affective value. Motivation and Emotion 36, 3 (2012), 404–413.

[49] A. Toet and S. Tak. 2013. Look out, there is a triangle behind you! The effect of primitive geometric shapes on perceived facial dominance. i-Perception 4, 1 (2013), 53–56.

[50] Alejandro Salgado-Montejo, Isabell Tapia Leon, Andrew J. Elliot, Carlos José Salgado, and Charles Spence. 2015. Smiles over frowns: When curved lines influence product preference. Psychology & Marketing 32, 7 (2015), 771–781. https://doi.org/10.1002/mar.20817

[51] M. Bertamini, L. Palumbo, T. N. Gheorghes, and M. Galatsidas. 2016. Do observers like curvature or do they dislike angularity? British Journal of Psychology 107, 1 (2016), 154–178.

[52] M. Bar and M. Neta. 2006. Humans prefer curved visual objects. Psychological Science 17, 8 (2006), 645–648.

[53] M. Seckler, O. Opwis, and A. N. Tuch. 2015. Linking objective design factors with subjective aesthetics: An experimental study on how structure and color of websites affect the facets of users’ visual aesthetic perception. Computers in Human Behavior 49 (2015), 375–389.

[54] C. Taft. 1997. Color meaning and context: Comparisons of semantic ratings of colors on samples and objects. Color: Research and Application 22, 1 (1997), 40–50.

[55] C. Taylor, A. Clifford, and A. Franklin. 2013. Color preferences are not universal. Journal of Experimental Psychology: General 142, 4 (2013), 1015.

[56] E. Kulegina, G. Schroeder, A. Allchin, K. Berlin, and M. Cakmak. 2018. Characterizing the design space of rendered robot faces. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. 96–104.

[57] A. Powers and S. 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. 218–225.

[58] C. Diana and A. L. Thomaz. 2011. The shape of Simon: Creative design of a humanoid robot shell. In Extended Abstracts on Human Factors in Computing Systems (CHI EA’11). 283–298.

[59] Y. Ferstl, E. Kokkinara, and R. McDonnell. 2016. Do I trust you, abstract creature? A study on personality perception of abstract virtual faces. In Proceedings of the ACM Symposium on Applied Perception. 39–43.

[60] A. Powers and S. 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. 218–225.

[61] C. Diana and A. L. Thomaz. 2011. The shape of Simon: Creative design of a humanoid robot shell. In Extended Abstracts on Human Factors in Computing Systems (CHI EA’11). 283–298.

[62] Y. Ferstl, E. Kokkinara, and R. McDonnell. 2016. Do I trust you, abstract creature? A study on personality perception of abstract virtual faces. In Proceedings of the ACM Symposium on Applied Perception. 39–43.

[63] T. Nomura, T. Suzuki, T. Kanda, and K. Kato. 2006. Measurement of negative attitudes toward robots. Interaction Studies 7, 3 (2006), 437–454.

[64] M. Ratchford and M. Barnhart. 2012. Development and validation of the technology adoption propensity (TAP) index. Journal of Business Research 65, 8 (2012), 1209–1215.

[65] C. Bartneck, D. Kulić, E. Croft, and S. Zoghi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics 1, 1 (2009), 71–81.

[66] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner. 2017. The Robotic Social Attributes Scale (RoSAS): Development and validation. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. 254–262.
[67] E. Liberman-Pincu and T. Oron-Gilad. 2021. Impacting the perception of socially assistive robots—evaluating the effect of visual qualities among children. In Proceedings of the 2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN’21). IEEE, Los Alamitos, CA, 612–618.

[68] E. Liberman-Pincu. 2021. Audrey—Flower-like social assistive robot: Taking care of older adults in times of social isolation during the COVID-19 pandemic. In Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction. 613–614.

[69] K. A. Hsiao and L. L. Chen. 2006. Fundamental dimensions of affective responses to product shapes. International Journal of Industrial Ergonomics 36, 6 (2006), 553–564.

[70] E. Liberman-Pincu and Y. Bitan. 2021. FULE—Functionality, usability, look-and-feel and evaluation novel user-centered product design methodology—illustrated in the case of an autonomous medical device. Applied Sciences 11, 3 (2021), 985.

[71] C. Putnam, B. Kolko, and S. Wood. 2012. Communicating about users in ICTD: Leveraging HCI personas. In Proceedings of the 5th International Conference on Information and Communication Technologies and Development. 338–349.

[72] M. L. Schrum, M. Johnson, M. Ghuy, and M. C. Gombolay. 2020. Four years in review: Statistical practices of Likert scales in human-robot interaction studies. In Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. 43–52.

[73] J. W. Castro, S. T. Acuña, and N. Juristo. 2008. Integrating the personas technique into the requirements analysis activity. In Proceedings of the 2008 Mexican International Conference on Computer Science. IEEE, Los Alamitos, CA, 104–112.

[74] H. Cramer, N. Kemper, A. Amin, B. Wielinga, and V. Evers. 2009. 'Give me a hug': The effects of touch and autonomy on people’s responses to embodied social agents. Computer Animation and Virtual Worlds 20, 2–3 (2009), 437–445.

[75] Z. Šulc and H. Řezanková. 2019. Comparison of similarity measures for categorical data in hierarchical clustering. Journal of Classification 36, 1 (2019), 58–72.

[76] M. Pino, M. Boulay, F. Jouen, and A. S. Rigaud. 2015. “Are we ready for robots that care for us?” Attitudes and opinions of older adults toward socially assistive robots. Frontiers in Aging Neuroscience 7 (2015), 141.

[77] A. von der Pütten and N. Krämer. 2012. A survey on robot appearances. In Proceedings of the 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI’12). IEEE, Los Alamitos, CA, 267–268.

[78] B. Reeves, J. Hancock, and X. Liu. 2020. Social robots are like real people: First impressions, attributes, and stereotyping of social robots. Technology, Mind, and Behavior 1, 1 (2020), 1–14.

[79] C. McGinn, E. Bourke, A. Murtagh, C. Donovan, P. Lynch, M. F. Cullinan, and K. Kelly. 2020. Meet Stevie: A socially assistive robot developed through application of a ‘design-thinking’ approach. Journal of Intelligent & Robotic Systems 98, 1 (2020), 39–58.

[80] F. Hegel, F. Eysel, and B. Wrede. 2010. The social robot ‘Flobi’: Key concepts of industrial design. In Proceedings of the 19th International Symposium in Robot and Human Interactive Communication. IEEE, Los Alamitos, CA, 107–112.

[81] K. Akdim, D. Belanche, and M. Flavián. 2021. Attitudes toward service robots: Analyses of explicit and implicit attitudes based on anthropomorphism and construal level theory. International Journal of Contemporary Hospitality Management E-public ahead of print, October 13, 2021.

[82] E. Broadbent, R. Stafford, and B. MacDonald. 2009. Acceptance of healthcare robots for the older population: Review and future directions. International Journal of Social Robotics 1, 4 (2009), 319.

[83] M. L. Walters, M. Lohse, M. Hanheide, B. Wrede, D. S. Syrdal, K. L. Koay, A. Green et al. 2011. Evaluating the robot personality and verbal behavior of domestic robots using video-based studies. Advanced Robotics 25, 18 (2011), 2233–2254.

[84] A. Powers and S. 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. 218–225.

Received 12 June 2021; revised 4 September 2022; accepted 20 October 2022