ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect

Computers in Human Behavior 24 (2008) 2456–2474

Computers in Human Behavior

www.elsevier.com/locate/comphumbeh

Sensitivity to the proportions of faces that vary in human likeness

Robert D. Green, Karl F. MacDorman *, Chin-Chang Ho, Sandosh Vasudevan

Indiana University School of Informatics, 535 West Michigan Street, IT 487, Indianapolis, IN 46202, USA

Available online 16 April 2008

Abstract

Despite the often quoted adage "beauty is in the eye of the beholder," studies indicate people perceive certain facial and bodily proportions as attractive regardless of their culture. This preference, which is present even in infants, may be more hardwired than learned. Designers of computer games, animation, virtual reality, and robots must make choices about how to depict humanlike forms. An understanding of human perception and preferences can lead to design principles for successful interaction. This study measured human responses to varying facial proportions in people, androids, mechanical-looking robots, and two- and three-dimensional characters. Participants showed greater agreement on the best proportions of faces they considered more humanlike and more attractive and less tolerance for deviation from these proportions in more attractive faces.

Keywords: Anthropomorphism; Attractiveness perception; Facial acceptability; Inter-rater agreement; Uncanny valley

1. Introduction

Research in evolutionary aesthetics is challenging the traditional view that personal tastes and cultural attitudes determine what is beautiful (Etcoff, 1999). Studies seeking evaluations of attractiveness across diverse cultures demonstrate both differences in

doi:10.1016/j.chb.2008.02.019

cultural preferences for certain features (e.g., nose size) and the universal appeal of other features such as eye height and width, cheekbone prominence, and chin length (Cunningham, Roberts, Barbee, Druen, & Wu, 1995; Jones, 1995).

Judgments of attractiveness begin early in life. Infants gaze longer at photographs that adults rate as attractive (Langlois et al., 1987), and children prefer to play with their more attractive peers (Salvia, Sheare, & Algozzine, 1975). First impressions are important. Participants who rated faces displayed for 13 ms agreed significantly ($t(9) = 4.90$, $p < .01$) with pretest assessments (Olson & Marshuetz, 2005). Not only do people assess attractiveness quickly, but those judgments remain strong after details are forgotten (Goldstein & Papageorge, 1980). This indicates people remember whether someone was attractive, but not why.

Universal aspects of attractiveness indicate good genes, developmental and hormonal health, and a strong immune system. Males tend to find features indicating fertility attractive in potential mates. Because female child-bearing years are limited, a mix of traits indicating sufficient maturity to give birth and sufficient youthfulness to bear many children constitutes the most attractive set of features. Male fertility is less limited, leading females to seek mates who will be good providers (Alam & Dover, 2001; Drury, 2000; Grammer & Thornhill, 1994; Scheib, Gangstead, & Thornhill, 1999). Females near ovulation (that is, when intercourse would most likely result in pregnancy) selected more "masculine" faces (Johnson, Hagel, Franklin, Fink, & Grammer, 2001). Males selected different faces when asked which of 16 females they would prefer for a dinner date, sexual intercourse, or raising children (Cunningham, 1986). Facial features varied among the groups.

Galton (1879) may have been the first to observe the attractiveness of the "averaged" face. In an attempt to isolate cues of criminality, he created composites by projecting overlaid negatives of convicts. The resulting composites were found to be more attractive than villainous. Langlois and Roggman (1990) used morphing software to create composites of 2, 4, 8, 16, and 32 faces. The attractiveness ratings of the composite faces were higher than the average of the faces that contributed to the composites. Composites made using more faces were rated as more attractive.

Averaged faces may be more attractive, but they are not the most attractive faces. Indeed, some faces are rated as more attractive than composites (Langlois & Roggman, 1990). Perrett, May, and Yoshikawa (1994) demonstrated that creating composites of faces deemed more attractive produced faces that were judged more attractive than an average of the entire group.

There are many possible explanations for the attractiveness of composite photographs. Their creation smooths out skin blemishes and irregularities. Jones, Little, and Perrett (2004) found a strong correlation between the perceived health of facial skin and the attractiveness ratings of male faces. Maintenance of clear facial skin can be an indicator of health (Grammer & Thornhill, 1994). While some models may have distinctive facial features, such as Cindy Crawford's mole, they exhibit otherwise clear skin.

Another result of averaging faces is the increased symmetry of the resultant face. Symmetry has long been an artistic ideal (Alam & Dover, 2001; Drury, 2000). Faces exhibiting symmetry were judged as more attractive (Rhodes, Proffitt, Grady, & Sumich, 1998). Maintaining facial symmetry through development indicates strong hormonal health (Grammer & Thornhill, 1994; Scheib et al., 1999), though developmental factors like sleep patterns may influence symmetry (Mealey, Bridgstock, & Townsend, 1999). However, people's overt judgments of symmetry are not as good as one might think. Scheib et al. (1999)

found ratings of facial symmetry were weakly correlated to measured symmetry. In the same study, participants rated faces that possessed symmetry as more attractive than asymmetric faces, even when half of the face was blocked and participants could not observe the symmetry.

There are many facial features that could potentially influence attractiveness apart from averageness, symmetry, and lack of blemishes. Cunningham (1986) measured 25 facial features of 50 women. From these measurements he compared ratings of attractiveness against 21 facial proportions, finding significant correlations in 12 of the 21 proportions. A regression analysis indicated eye height, nose area, cheek width, and smile width accounted for more than 50% of the variance in attractiveness ratings. A study using a larger number of facial landmarks (135 for male and 130 for female stimuli) and proportions (156 for male and 155 for female stimuli) found significant correlations between attractiveness and about 20% of male facial proportions, but only about 10% of female facial proportions (Farkas, 1994).

Using a more limited set of facial proportions, Grammer and Thornhill (1994) found different facial proportions contributed to the perception of traits such as attractive, dominant, sexy, and healthy. Prominent eyes and cheekbones contributed most to males' evaluations of females, while jaw width and lower-face proportions contributed most to females' evaluations of males. In a study of profiles of faces, participants viewed video of changing facial proportions and indicated when the profile was "acceptable." A 1-mm change could render an "acceptable" face "unacceptable" (Giddon, Sconzo, Kinchen, & Evans, 1996).

If our perceptions of other human beings are at least partly rooted in natural selection, how do we perceive synthetic characters? Goetz, Kiesler, and Powers (2003) found participants believed humanlike robots were best suited for interactive tasks, while mechanical-looking robots were best suited for routine jobs. How do we determine whether a robot is humanlike? A study of 48 commercial, research, and fictional robots indicates that to be considered humanlike a robot should have a distinctively human head shape and a facial area dominated by human features. Most significantly, a nose, eyelids, and mouth suggest humanness (DiSalvo, Gemperle, Forlizzi, & Kiesler, 2002).

DiSalvo et al. (2002) recommended exaggerated features and an encasement to hide the mechanics in the head so the robot seems not only humanlike but also product-like. This product focus is intended to keep the robot from falling into the uncanny valley. Masahiro Mori proposed that as robots become more humanlike they appear more familiar up to a point. They risk becoming eerie when they are nearly human but not quite, especially when they are discovered to be mechanical through touch or by other means (Mori, 1970). Mori cited corpses and zombies as residents of the uncanny valley. The uncanny valley may be caused by a mismatch between experience and expectation. Another example cited by Mori is a prosthetic hand that looks realistic but feels artificial. This hypothesis was supported by a study of robot videos that found fear, disgust, anxiety, dislike, and shock were the emotions that best predicted eeriness (Ho, MacDorman, & Pramono, 2008).

Avoidance of the uncanny valley became a rubric, so roboticists designed robots with features that only approximate human likeness. MacDorman and Ishiguro (2006) plotted the uncanny valley by having participants rate images that morphed from a mechanical-looking humanoid robot to an android to the human model for the android. In a similar experiment, Hanson (2006) demonstrated the uncanny valley could be avoided, not by shunning human likeness, but by careful design, modifying the uncanny faces to emphasize features identified with friendliness and youthfulness (neoteny). But the uncanny valley is not just an artifact of viewing still images; MacDorman (2006) also found participants responded to the eerieness of certain robotic motions.

1.1. Hypotheses

Facial proportions affect how humans perceive one another, but there is little understanding of how much leeway designers of synthetic characters have in representing faces. We are sensitive to differing facial proportions in other humans (Cunningham, 1986; Cunningham et al., 1995; Grammer & Thornhill, 1994) and prefer those who look like us (Golby, Gabrieli, Chiao, & Eberhardt, 2001; Jones, 1995). Will we extend this preference to synthetic characters? To examine the relation between human likeness and facial proportions, this study explored the following hypotheses:

H1: Figures that are rated more humanlike have a narrower range of acceptable facial proportions than those rated less humanlike.

H2: There is greater inter-rater agreement on what facial proportions are best in more humanlike figures than in less humanlike figures.

Attractiveness has many benefits. Attractive people are perceived as more intelligent, sociable, healthy, and trustworthy. While attractive people are presumed to be more sexually experienced, they are also presumed to be less likely to have affairs. People are more willing to perform acts of altruism, such as helping to move home or donating blood, for an attractive person than an unattractive person (Cunningham, 1986). Nevertheless, there are specific facial proportions that are considered more attractive, which leads to the following hypothesis:

H3: Figures that are rated more attractive have a narrower range of acceptable facial proportions than those rated less attractive.

Studies have indicated that there is agreement in which faces are attractive as well as unattractive. If possessing certain facial proportions makes a face attractive, best proportions must be best regardless of how attractive or unattractive the face being evaluated is. The final hypothesis for this study is

H4: Inter-rater agreement on which facial proportions are best in more attractive figures is not significantly different from inter-rater agreement on the best facial proportions in less attractive figures.

2. Methodology

2.1. Participants

Two hundred and eight participants completed the study. 61% ($n = 126$) were male and 39% were female ($n = 81$). The mean age of participants was 31.9 ($SD = 10.27$), ranging from 17 to 79. The largest group of participants were those born in the United States (31.3%, $n = 65$), followed by Indonesia (30.3%, $n = 63$) and the United Kingdom

(12.0%, $n = 25$). The majority of participants lived in the United States (39.9%, $n = 83$) followed again by Indonesia (30.3%, $n = 63$) and the United Kingdom (12.0%, $n = 25$). Twenty-six countries were recorded as country of birth, and 20 as country of residence. Nearly a fifth of participants (18.6%, $n = 29$) did not live in their country of birth. Participants had an average of 16.1 years of education (representing the completion of a bachelor's degree in the United States) with a standard deviation of 4.61. Participants were recruited through e-mail, newsgroup postings, and direct contact.

2.2. Stimuli

Eleven stimuli were prepared for the study. Stimuli included two photographs of humans (one male and one female), 3 three-dimensional computer graphics characters (one male human, one female human, and one robot), three computer drawings (one male human, one female human, and one robot), and three photographs of robots. Four videos were created for each stimulus. Each video warped one facial proportion between the extremes of ±10%. The proportions altered were cheek width, eye separation, face height, and jaw width. These dimensions were selected because cheek width, eye separation, and face height had the strongest correlations to attractiveness for female faces (Cunningham, 1986; Grammer & Thornhill, 1994), while face height and jaw width had the strongest correlations to attractiveness for male faces (Grammer & Thornhill, 1994).

2.2.1. Still images

Original photographs and artwork were converted to 400 × 500 pixel images framing the face on a 50% gray background. Fig. 1 depicts the converted images.

Five facial dimensions, depicted in Fig. 2, were used to calculate the proportions. The dimensions are defined as follows:

E2–E1: Distance between the center of the pupils.
C2–C1: Distance between the outer edge of the cheekbones at the most prominent point.
J2–J1: Width of the face at the level of the middle of the smile.
F3–F2: Distance between the mid-point of the pupils and the bottom of the chin.
F3–F1: Distance between the top of the head and the bottom of the chin.

Four facial proportions were calculated using these dimensions.

Cheek width: The width of the face at the cheek bones divided by the overall height of the head (C2–C1)/(F3–F1).
Eye separation: The distance between the pupils divided by the width of the face at the cheek bones (E2–E1)/(C2–C1).
Face height: The height of the face between the eyes divided by the overall height of the head (F3–F2)/(F3–F1).
Jaw width: The width of the face at the mouth divided by the distance between the pupils (J2–J1)/(E2–E1).

The numerator was increased or decreased from 1 to 10% to change the proportions along a dimension. Table 1 shows the ratio and dimension that were altered for each of the four proportions for all 11 stimuli. All measurements are in pixels.

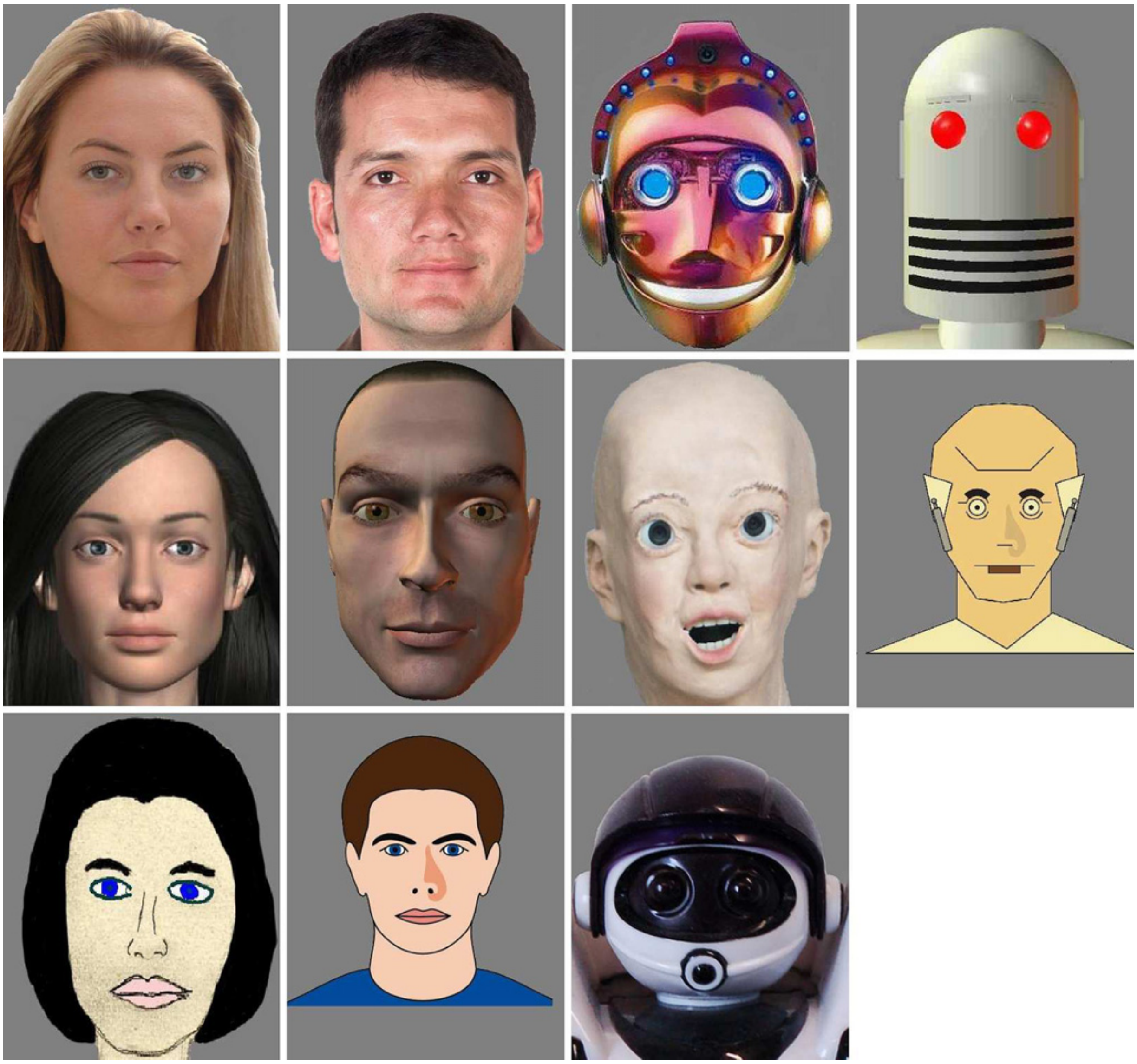

Fig. 1. Original stimuli normalized to 400 × 500 pixels. Top row: Female and male photographs, Anthrobot, 3D robot. Middle row: 3D female, male, Barthoc, Jr., and 2D robot. Bottom row: 2D female, male, and Robosapien.

2.2.2. Video

Four videos, one for each proportion, were created for each stimulus in Fig. 1. Each video warped one of the above proportions between the extremes of ±10% and was composed of 21 frames. Hundreds of points were set on each face to facilitate warping.

The point under consideration was moved in or out by 1 to 10% of the measurement listed in Table 1, and neighboring points were moved to provide the smooth transitions. Fig. 3 demonstrates the extreme frames for eye separation for the 3D female character. For the lower facial length dimension the point identified as F2 in Fig. 2 was moved up or down, but the chin (F3) remained fixed to maintain the overall facial length.

2.3. Procedures

The website Exploring the Uncanny Valley (http://experiment.informatics.iupui.edu) hosted this study. The initial page instructed participants to choose a language. Available

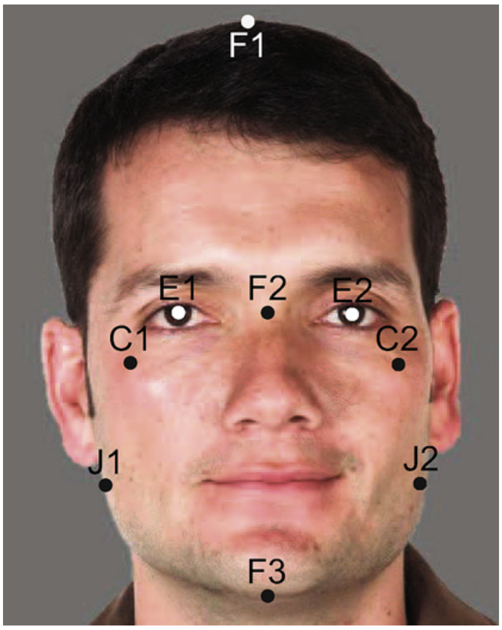

Fig. 2. Points used in measuring facial dimensions.

Table 1
Measurement of facial dimensions

Stimulus	Cheek width[a]	C2–C1[e]	Eye separation[b]	E2–E1[e]	Face height[c]	F3–F2[e]	Jaw width[d]	J2–J1[e]
Female photo	0.62	253	0.48	122	0.66	269	1.69	122
Male photo	0.45	206	0.67	138	0.58	265	1.90	262
3D female	0.51	221	0.56	123	0.49	214	1.87	230
3D male	0.63	289	0.52	151	0.75	342	1.71	258
2D female	0.33	161	0.71	115	0.59	252	1.70	196
2D male	0.74	175	0.47	82	0.73	174	1.76	144
Anthroboot	0.62	246	0.57	139	0.70	248	1.57	218
Barthoc, Jr.	0.47	230	0.60	139	0.51	237	1.91	266
Robosapien	0.73	249	0.46	114	0.62	212	1.96	233
3D robot	0.59	251	0.49	124	0.78	333	2.02	251
2D robot	0.51	173	0.45	78	0.57	198	1.79	140

[a] Width of the face at the cheek bones/overall height of head.
[b] Distance between pupils/width of face at the cheek bones.
[c] Height of the face between the eyes/overall height of head.
[d] Width of the face at the mouth/distance between pupils.
[e] Distance in pixels.

languages were English, Japanese, Traditional Chinese, Simplified Chinese, and *Bahasa Indonesia*. The website invited visitors to provide their demographic data. There were four tasks in the study. Participants were required to complete the tasks in sequence.

2.3.1. Task 1

Video stimuli were presented one at a time in random order. A frame was selected at random as a starting point. This frame was displayed with positioning buttons that enabled participants to adjust the face. Participants were instructed to adjust it until it looked best. Then the next face was presented, and the process was repeated for all faces.

2.3.2. Tasks 2 and 3

Tasks 2 and 3 mirrored one another. Frame sequences stimuli were presented one at a time in random order. The video frame selected as best in Task 1 was the single frame presented. One direction was locked at random for the task (frames could only be selected to the left or right of the starting point). The participant was instructed to find the last point where the face looked acceptable. Then the next face was presented, and the process repeated for all faces. For Task 3, the direction was reversed after the participant selected an acceptable point.

2.3.3. Task 4

Still images of the base faces were presented one at a time in random order. Participants were asked to indicate their agreement with a series of statements on a seven-point Likert scale. The order of the questions was randomized for each participant. The survey asked for the participant's level of agreement with the following statements:

- This figure looks female.
- This figure looks creepy.
- This figure looks sexy.
- This figure looks ugly.
- This figure looks alive.
- This figure looks humanlike.

Responses on the scale were *strongly disagree*, *moderately disagree*, *slightly disagree*, *neutral*, *slightly agree*, *moderately agree*, and *strongly agree*.

3. Results

This section looks first at participants' assessments of the best point, acceptable range, and ratings for the various attributes. Next, we examine the relations between attributes

Fig. 3. Extremes in placements of the eyes of the 3D female character.

and sensitivity to the best point, followed by the relations between attributes and tolerance for acceptable range. Figures are grouped by type (Human or Robot), and compared for significant differences. The relations between attributes and participant-selected best points are evaluated. Finally, the effect of gender on responses is compared.

3.1. Sensitivity and tolerance to facial proportions

The first task required participants to select the best position on 44 adjustable images (11 characters by 4 facial proportions). Each adjustable image was a 21-frame Flash movie with the 11th frame representing the original image. A change of one frame represents a 1% change in the proportion. The first task required participants to select the frame in the sequence that represented the best proportion. Tasks 2 and 3 required participants to indicate the last point at which the figure looked acceptable as the proportion either increased or decreased. The acceptable range of facial proportions was computed as the difference between these two points.

Participant selections were recorded in frames and converted to proportions. The standard deviation of the best point and the mean of the range for each figure are detailed in Table 2.

Sensitivity was defined as the standard deviation of the best point. Tolerance of acceptable proportions was defined as the mean of the acceptable range. For cheek width, participants showed the greatest sensitivity ($SD = 0.016$) and the least tolerance ($M = 0.026$) to the 2D female. The 2D male had the greatest sensitivity ($SD = 0.021$) for eye separation, while the female photograph had the least tolerance ($M = 0.029$) in the same proportion. For face height participants showed the greatest sensitivity to the 3D female ($SD = 0.017$) and the least tolerance to the male photograph ($M = 0.027$). The 3D female had the greatest sensitivity ($SD = 0.055$) and least tolerance ($M = 0.047$) for jaw width. Participants

Table 2
Sensitivity and tolerance by figure and facial proportion

Figure	Sensitivity[a]				Tolerance[b]			
	Cheek[c]	Eyes[d]	Face[d]	Jaw[d]	Cheek[c]	Eyes[d]	Face[d]	Jaw[d]
Female photo	0.017	0.022	0.019	0.059	0.028	0.029	0.028	0.115
Male photo	0.028	0.030	0.022	0.070	0.045	0.037	0.027	0.132
3D female	0.029	0.026	0.017	0.055	0.048	0.036	0.027	0.095
3D male	0.023	0.025	0.028	0.057	0.030	0.030	0.041	0.100
2D female	0.016	0.037	0.026	0.065	0.026	0.053	0.038	0.111
2D male	0.027	0.021	0.030	0.079	0.045	0.034	0.048	0.142
Anthrobot	0.033	0.032	0.038	0.070	0.044	0.047	0.053	0.114
Barthoc, Jr.	0.027	0.028	0.028	0.102	0.039	0.035	0.032	0.147
Robosapien	0.033	0.026	0.038	0.092	0.058	0.037	0.057	0.168
3D robot	0.023	0.028	0.037	0.106	0.026	0.045	0.059	0.123
2D robot	0.031	0.026	0.027	0.111	0.048	0.039	0.041	0.153

[a] Sensitivity measure as standard deviation from the best proportion. Lower standard deviation indicates greater sensitivity.
[b] Tolerance measured as the acceptable range of a proportion.
[c] $n = 194$.
[d] $n = 208$.

showed greater sensitivity and less tolerance for deviation in human than robotic figures and for more realistic humans (photographs and 3D) than for 2D drawings.

3.2. Figure attributes

Task 4 required participants to rate each figure on six adjectives, or attributes, on a seven-point Likert scale from *strongly disagree* (−3) to *strongly agree* (3). The two human figures were rated the most alive (male $M = 2.53$ and female $M = 2.51$) followed by the 3D human figures, the three robots, the 2D human figures, and finally the 3D and 2D robots ($M = -2.13$). The female figure was rated most sexy ($M = 1.57$), female ($M = 2.79$), and humanlike ($M = 2.87$) and least creepy ($M = -2.39$) and ugly ($M = -2.38$). By contrast, the robot Barthoc, Jr. was rated least sexy ($M = -2.85$) and most creepy ($M = 2.43$) and ugly ($M = 2.30$). The results of Task 4 are summarized in Table 3.

3.3. Comparison of attributes and proportions

To compare sensitivity, each participant's selection of the best point was converted to the difference from the mean best proportion ($X - \overline{X}$). A one-way ANOVA by figure was performed on each of the six attributes and difference and range for each of the four facial proportions. All dependent variables varied significantly ($p < .001$). The effect size (ω) was large for all variables except the ranges, which had ω values from .31 (face) to .20 (jaw).

A two-tailed Pearson's correlation was performed using the difference in each facial proportion (sensitivity), the mean range (tolerance), and each attribute. Table 4 presents the correlations between participants' assessments of the six attributes and their sensitivity to best proportions and tolerance for acceptable proportions. The test found the strongest correlations between the attribute humanlike and sensitivity to face height ($r = -.64$) and jaw width ($r = -.54$), both $p < .01$. All correlations between attributes and sensitivity were significant at a level of $p < .01$ except the correlation between alive and eye separation ($p < .05$) and between face height and creepy and ugly, both of which failed to reach

Table 3
Participant ratings of figure attributes

Figure	Humanlike[a]	Alive[b]	Female[b]	Sexy[b]	Creepy[b]	Ugly[b]
Female photo	2.87	2.51	2.79	1.57	−2.39	−2.38
Male photo	2.80	2.53	−2.90	−0.42	−1.78	−1.26
3D female	1.81	0.32	2.63	1.19	−1.91	−2.21
3D male	1.62	0.15	−2.73	−0.65	−0.90	−1.03
2D female	0.96	−1.90	1.90	−2.10	−0.50	0.15
2D male	0.57	−1.80	−2.56	−1.87	−1.32	−0.81
Anthrobot	−2.14	−1.47	−1.63	−2.08	−0.97	−0.95
Barthoc, Jr.	0.62	−1.10	−0.97	−2.85	2.43	2.30
Robosapien	−2.35	−1.40	−1.94	−2.07	−1.82	−1.49
3D robot	−2.68	−2.09	−2.13	−2.37	−1.25	−0.91
2D robot	−1.46	−2.13	−2.62	−2.31	−0.90	−0.47

[a] $n = 142$.
[b] $n = 208$.

Table 4
Correlation between selected attributes and facial proportions

Attribute	Sensitivity[a]				Tolerance[b]			
	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[c]	−.45**	.11**	−.64**	−.54**	.015	−.003	−.17**	−.012
Alive[d]	−.20**	−.044*	−.39**	−.39**	−.087**	−.12**	−.23**	−.14**
Female[d]	−.42**	.21**	−.51**	−.43**	−.05*	.053*	−.11**	−.07**
Sexy[d]	−.15**	−.16**	−.43**	−.50**	−.089**	−.13**	−.19**	−.15**
Creepy[d]	−.13**	.18**	−.014	.25**	.090**	.091**	.070**	.15**
Ugly[d]	−.14**	.23**	.018	.29**	.10**	.10**	.080**	.17**

[a] Sensitivity is participants' proximity to the mean.
[b] Tolerance is participants' acceptable range.
[c] $n = 1408$ for cheek width and 1562 for other proportions.
[d] $n = 2134$ for cheek width and 2288 for other proportions.
* $p < .05$.
** $p < .01$.

a significance level of $p < .05$. Most correlations between attributes and tolerance were significant at either $p < .01$ or $p < .05$. These correlations were generally weaker than those related to sensitivity. The strongest correlation was between alive and face height ($r = -.23,\ p < .01$).

Participants showed increased sensitivity towards the best point in all proportions as ratings for the attribute humanlike increased, except eye separation, for which they showed slightly decreased sensitivity (all *p*s $< .01$). Sensitivity increased in all four proportions as ratings for alive, female, and sexy increased (all *p*s $< .01$). Correlations between sensitivity and the attributes creepy and ugly were mixed; participants showed decreased sensitivity to the best point for eye separation and jaw width but greater sensitivity to cheek width as ratings increased (all *p*s $< .01$).

The relations between tolerance and the attribute humanlike were mixed; tolerance decreased in all proportions except cheek width as ratings for humanlike increased. The relation between face height and humanlike was the only one to reach significance ($p < .01$). Participants indicated a narrow acceptable range (i.e., less tolerance to variation), in all four proportions as ratings increased for attributes alive, female, and sexy (all *p*s $< .01$). Tolerance increased in all proportions as ratings of attributes creepy and ugly increased (all *p*s $< .01$).

Figs. 4 and 5 show the strongest relations for human likeness. These graphs combine the statistics for each figure as reported in Tables 2 and 3.

3.4. Comparison of human and robot figures

Reviewing Table 2 indicates participants were generally more sensitive and less tolerant of variation in the facial proportions of human characters than robots. The dataset was divided between those figures known to be human (2D, 3D, and photos of female and male figures) and robot (2D and 3D robots, Anthrobot, Barthoc, Jr., and Robosapien). A one-way ANOVA was performed on the six attributes and difference and range of the four facial proportions. Significant differences ($p < .001$) were detected in all six attributes and in difference from the best point. Acceptable range had significant differences at $p < .001$, except cheek width ($p < .01$).

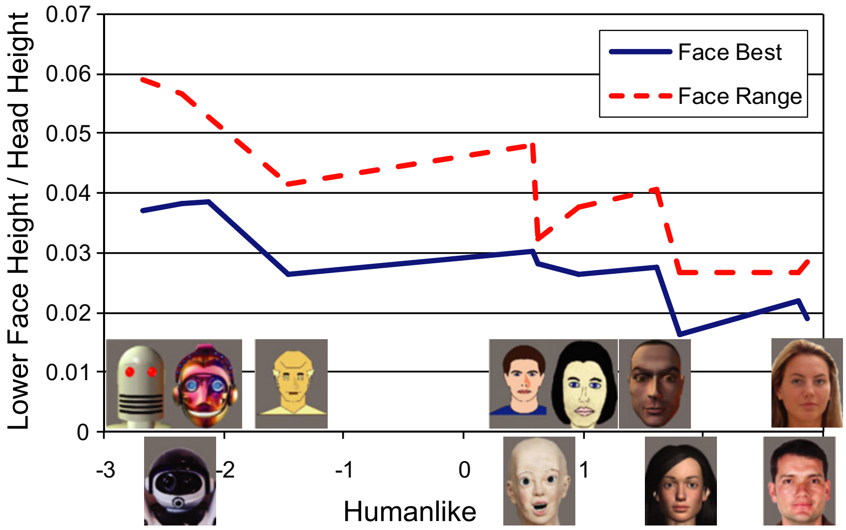

Fig. 4. Sensitivity and tolerance in face height sorted by human likeness.

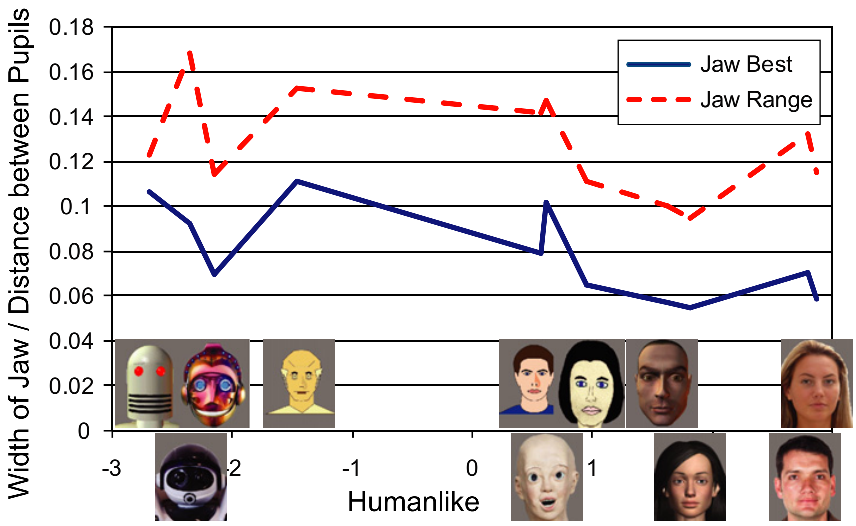

Fig. 5. Sensitivity and tolerance in jaw width sorted by human likeness.

Two-tailed bivariate correlation tests were performed on the Human and Robot groups separately (see Tables 5 and 6). The Human group had significant increases in sensitivity to the best point, in all three of four proportions, with attributes humanlike, alive, female, and sexy. It also had significantly decreased sensitivity in three of four proportions with increased ratings of the attributes creepy and ugly. All *p*s were less than .01 except alive-cheek width with $p < .05$ and humanlike-eye separation, sexy-cheek width and creepy-jaw width, all $p > .05$. The Robot group produced less predictable results for sensitivity. Relations between attributes humanlike, creepy, and ugly were significant at $p < .01$ for all four proportions, but the direction of those relations was mixed, showing increased sensitivity to cheek width and face height, and decreased sensitivity for eye separation and jaw width for each of these three attributes.

The relation between attributes and sensitivity to the best point is predicted by the standard deviation of the attribute for that group. For example, the Human group was

Table 5
Correlation between attributes and facial proportions of human figures

Attribute	Sensitivity[a]				Tolerance[b]			
	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[c]	−.14**	−.053	−.35**	−.18**	.13**	.085*	.003	.10**
Alive[d]	−.064*	−.18**	−.39**	−.20**	−.051	−.14**	−.28**	−.07*
Female[d]	−.44**	.15**	−.64**	−.53**	−.069*	.086**	−.12**	−.076**
Sexy[d]	.051	−.30**	−.48**	−.40**	−.039	−.13**	−.18**	−.12**
Creepy[d]	−.12**	.21**	.22**	.052	.11**	.16**	.21**	.14**
Ugly[d]	−.17**	.31**	.27**	.20**	.094**	.15**	.18**	.16**

[a] Sensitivity is participant's proximity to the mean.
[b] Tolerance is participant's acceptable range.
[c] $n = 768$ for cheek width and 852 for other proportions.
[d] $n = 1164$ for cheek width and 1248 for other proportions.
* $p < .05$.
** $p < .01$.

Table 6
Correlation between attributes and facial proportions of robotic figures

Attribute	Sensitivity[a]				Tolerance[b]			
	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[c]	−.33**	.24**	−.55**	.14**	.054	.014	.12**	.032
Alive[d]	.061	.17**	−.017	−.084**	−.065*	−.059	−.13**	−.15**
Female[d]	−.042	.32**	−.040	−.11**	.073*	.097**	.054	.043
Sexy[d]	.20**	−.050	.16**	−.12**	−.080*	−.091**	−.069*	−.13**
Creepy[d]	−.43**	.33**	−.47**	.10**	.040	.008	−.091**	.11**
Ugly[d]	−.42**	.26**	−.47**	.16**	.070*	.026	−.068*	.14**

[a] Sensitivity is participant's proximity to the mean.
[b] Tolerance is participant's acceptable range.
[c] $n = 640$ for cheek width and 710 for other proportions.
[d] $n = 970$ for cheek width and 1040 for other proportions.
* $p < .05$.
** $p < .01$.

considered more humanlike with greater agreement ($M = 1.77$, $SD = 1.447$) than the Robot group ($M = -1.60$, $SD = 1.728$). The resulting absolute r values were greater for each proportion in the Robot group than the Human group.

Tolerance for the acceptable range increased in the Human group for all four proportions as ratings of humanlike, creepy, and ugly increased, and as ratings of alive, sexy, and female decreased except female-eye separation. All ps were less than .01, except humanlike-eye separation, alive-jaw width, and female-cheek width $p < .05$, and humanlike-face height, alive-cheek width, and sexy-cheek width $p > .05$. Correlations for tolerance in the Robot group were weaker, and fewer correlations reached a significance level of $p < .05$.

3.5. Relation between attributes and best points

All results reported up to this point have been in terms of sensitivity to the best point (standard deviation) or tolerance (acceptable range). Table 7 lists the mean of the best

Table 7
Best proportions by figure and facial proportion

Figure	Cheek[a]	Eyes[b]	Face[b]	Jaw[b]
Female photo	0.60	0.48	0.66	1.65
Male photo	0.44	0.65	0.59	1.80
3D female	0.54	0.56	0.49	1.76
3D male	0.61	0.51	0.75	1.64
2D female	0.33	0.72	0.59	1.69
2D male	0.73	0.47	0.74	1.74
Anthrobot	0.61	0.56	0.72	1.55
Barthoc, Jr.	0.48	0.59	0.51	1.91
Robosapien	0.73	0.47	0.63	1.92
3D robot	0.59	0.48	0.75	1.99
2D robot	0.50	0.45	0.58	1.85

[a] $n = 194$.
[b] $n = 208$.

point recorded by face and proportion. The variation between figures is because of the variation between original proportions as listed in Table 1.

A two-tailed bivariate correlation between the six attributes and best points for the four facial proportions indicated that a narrower cheek and jaw, shorter face height, and wider eyes were related positively with humanlike and alive (all ps $< .01$). The attribute sexy only had significant correlations with face height and jaw width (both ps $< .01$), aligning with humanlike and alive. Creepy and ugly also had positive correlations with a narrower cheek, shorter face height, and wider eye separation, but jaw width had a negative effect on both attributes. Table 8 details the correlations between attributes and best points.

3.6. *Difference in response by gender*

An independent-samples t-test was performed, grouped by participant gender, on the six attributes, four differences from best point, and four acceptable ranges. Female participants showed less sensitivity and more tolerance in selecting facial proportions than male participants. The difference in tolerance was significant ($p < .01$) in all four proportions, though the effect sizes were small with tolerance for cheek and jaw width the only differences with an $r > .10$. To understand further the effect of participant gender, the stimuli were separated into Female, Male, and Robot groups.

Table 8
Correlation between attributes and best facial proportions

Attribute	Cheek	Eyes	Face	Jaw
Humanlike[a]	−.28**	.37**	−.23**	−.27**
Alive[b]	−.077**	.12**	−.090**	−.18**
Female[b]	−.27**	.27**	−.37**	−.23**
Sexy[b]	.037	−.008	−.094**	−.25**
Creepy[b]	−.23**	.17**	−.19**	.13**
Ugly[b]	−.25**	.20**	−.17**	.16**

[a] $n = 1408$ for cheek width and 1562 for other proportions.
[b] $n = 2134$ for cheek width and 2288 for other proportions.
* $p < .05$.
** $p < .01$.

In these groupings female participants continued to show less sensitivity and more tolerance than male participants. Male participants rated female figures as sexier ($M = .39$, $SE = .110$) than female participants ($M = -.02$, $SE = .13$; $t(619) = -2.64$, $p < .05$, $r = .095$). Female participants rated robots as more humanlike ($M = -1.38$, $SE = .11$) than male participants ($M = -1.75$, $SE = .083$). The difference was significant with a small effect size ($t(557) = 2.73$, $p < .01$, $r = .11$). Male participants rated the Robot group slightly more sexy ($M = -2.26$, $SE = 0.048$) than female participants ($M = -2.45$, $SE = 0.053$). The difference in ratings of sexy was significant though the effect size was very small ($t(1033) = -2.65$, $p < .01$, $r = .082$).

3.7. Result summary

Relations between attributes and sensitivity and between attributes and tolerance are generally stronger the more human and realistic a stimulus appears. In most instances participants showed increased sensitivity and decreased tolerance as ratings of humanlike, alive, female, and sexy increased, and as ratings of creepy and ugly decreased. However, this pattern had some variance between figure groupings and proportions. Participant gender led to different results, particularly in assessing sexiness and the human likeness of robots. Female participants also seemed to be more tolerant of varying facial proportions.

4. Discussion

4.1. Sensitivity and tolerance relative to human likeness

Hypothesis H1 predicted that the acceptable range narrows (tolerance decreases) as ratings of human likeness increase. This hypothesis cannot be supported by this study. Tolerance for the range of acceptable facial proportions is not strongly correlated with participants' ratings of human likeness when viewing results for all 11 figures. Only tolerance to face height decreased as human likeness increased, while tolerance of cheek width slightly increased as human likeness increased. This may be owing to lack of experimental control because the representations were all of different figures; greater control could have been obtained by varying the human likeness of the depiction of the same figure.

This study provides good support for H2: there is heightened sensitivity, that is, greater inter-rater agreement concerning the best point, as human likeness increases. (Heightened sensitivity is measured by smaller differences from the best point.) As ratings of humanlike increased, sensitivity increased in all proportions except eye separation. Face height and jaw width had large effect sizes while viewing all figures.

4.2. Sensitivity and tolerance relative to attractiveness

H3 and H4 are both related to ratings of attractiveness. This study asked participants to rate ugly and sexy. While these terms may not be exact antonyms (attractive–ugly) or synonyms (attractive–sexy), they will be used to evaluate attractiveness.

The results of this study support H3, which predicts decreased tolerance (a narrower acceptable range) as ratings of attractiveness increase. Tolerance decreased significantly for the relation between sexy and eye separation, face height, and jaw width. The relation between sexy and cheek width showed decreased tolerance; however the effect size was very small, and the relation did not reach significance. The relations between ugly and each facial proportion showed increased tolerance as ugliness increased, and all relations were significant.

Hypothesis H4, claiming sensitivity does not vary based on assessments of attractiveness, is not supported. Participants had greater sensitivity to the best proportions in sexy figures and less sensitivity in ugly figures than those at the other end of the scale.

There were significant correlations between attributes sexy and ugly and all four facial proportions. Sensitivity increased as sexiness increased in eye separation, face height, and jaw width. Likewise, sensitivity increased as ugliness decreased in eye separation, face height, and jaw width. Sensitivity to the best cheek width moved in the opposite direction of the other facial proportions, increasing as ugliness increased and decreasing as sexiness increased. This effect was observed in both the Human and Robot groups, though the relation between sensitivity and facial height flipped for sexy in the Robot group.

4.3. What proportions are best?

Thus far, all discussion has been on sensitivity and tolerance. Do the selections of best points reveal anything interesting? Costa and Corassa (2006) found artists exaggerate the size of the eyes and lips and the length of the face. Table 9 lists the frame selected as the best point for each figure and facial proportion and shows the extent of change from the figure's original proportion. Participants preferred narrower cheeks and jaws and longer faces than the original images. Preferences for eye separation differed between human and robotic figures. Participants preferred narrower-set eyes in human figures and wider-set eyes in robotic figures.

4.4. The creepiness of the unknown human

A two-tailed bivariate correlation among attributes showed strong, positive correlations between humanlike and alive, sexy, and female. Further analysis revealed a region

Table 9
Best frames by figure and facial proportion

Figure	Best frame				Difference from original frame[c]			
	Cheek[a]	Eyes[b]	Face[b]	Jaw[b]	Cheek[a]	Eyes[b]	Face[b]	Jaw[b]
Female photo	8.10	11.88	11.42	8.36	−2.90	0.88	0.42	−2.64
Male photo	9.15	8.18	9.10	5.55	−1.85	−2.82	1.90	−5.45
3D female	12.73	11.10	11.63	5.00	1.73	0.10	0.63	−6.00
3D male	8.24	8.33	11.17	7.18	−2.76	−2.67	0.17	−3.82
2D female	11.53	9.89	11.02	10.35	0.53	1.11	0.02	−0.65
2D male	12.27	11.05	12.05	10.07	−1.27	−0.05	1.05	−0.93
Anthrobot	9.04	9.84	13.16	9.63	−1.96	−1.16	2.16	−1.37
Barthoc, Jr.	9.76	13.44	10.14	10.88	1.24	−2.44	0.86	0.12
Robosapien	10.95	7.82	13.23	9.17	−0.05	3.18	2.23	−1.83
3D robot	11.09	13.28	14.44	9.43	0.09	2.28	−3.44	−1.57
2D robot	8.31	9.94	12.18	7.43	−2.69	1.06	1.18	3.57

[a] $n = 194$.
[b] $n = 208$.
[c] Difference in frames (or percentage) corrected for direction of change in Flash movie.

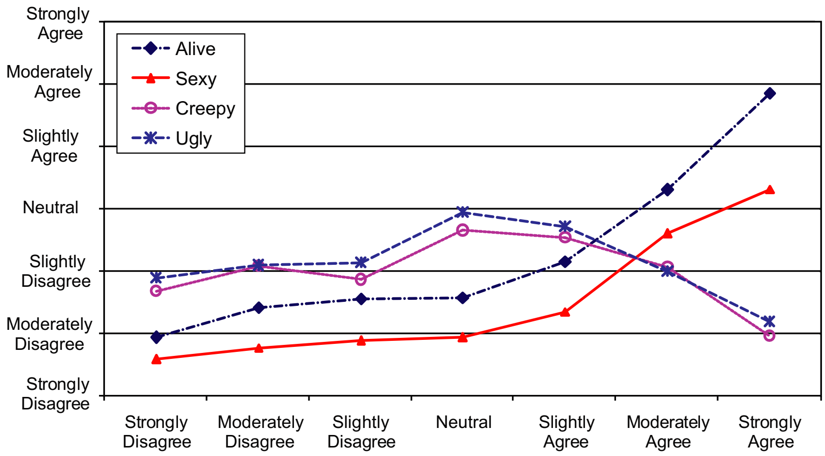

Fig. 6. Ratings of humanlike versus the mean of other attributes. The rating for the attribute creepy is at its highest when human likeness is indeterminate.

of heightened creepiness associated with the uncanny valley. Fig. 6 plots the mean value for the attributes alive, sexy, creepy, and ugly when participants rated humanlike according to the values on the *x*-axis. Mori (1970) predicted creepiness when a figure is nearly human. This study indicates participants sensed heightened creepiness when they rated humanlike as neutral.

It would be easy to ascribe this phenomenon to the effect of the robot Barthoc, Jr. This figure had the highest rating for both creepy and ugly, while the rating for humanlike was very close to neutral. But Barthoc, Jr. was considered least creepy when its human likeness was rated as neutral.

4.5. Gender differences

We were concerned that male participants would be more reluctant to rate male figures as sexy than female participants would be for female figures. Indeed, ratings of male figures for the attribute sexy were higher from female participants than from male participants; however, this difference was not statistically significant. What did achieve statistical significance was the difference between ratings of sexy for female figures; male participants provided higher ratings than their female counterparts.

Female participants showed greater tolerance in the acceptable range of facial proportions for all stimulus types (Female, Male, and Robot). Female participants also considered the robotic characters to be more humanlike than did male participants. These results may be related. Possible causes include traditional feminine nurturing roles causing increased acceptance, and males' generally higher familiarity with robots and other mechanical objects as tools and diminished need to anthropomorphize them. These explanations are purely conjectural.

4.6. Summary

This study examined the relations among sensitivity to best facial proportions, tolerance for acceptable facial proportions, and several characters attributes. It demonstrated significant correlations between the selection of best proportions and ratings of human likeness and attractiveness. It also demonstrated that while there is a significant correlation between acceptable ranges for facial proportions and ratings of attractiveness, such a relation cannot be established between acceptable range and human likeness. An uncanny valley was found when participants were most ambivalent about the human likeness of a face.

Acknowledgments

We would like to thank Thomas Busey, Patrick Huehls, Clint Koch, Himalaya Patel, and Alan Roberts for helpful comments and suggestions. We would also like to thank Heryati Madiapuri for translating the study website into *Bahasa Indonesia*, for recruiting Indonesian participants, and for fielding their queries. Thanks are also due to Robert Doornick (International Robotics; Anthrobot), Matthias Hackel (Mabotic Robotics & Automation; Barthoc, Jr.), and Sara McGrath for providing images of robots and granting us permission to use them in this study.

References

Alam, M., & Dover, J. S. (2001). On beauty: Evolution, psychological considerations and surgical enhancement. *Archives of Dermatology, 137*(6), 795–807.

Costa, M., & Corassa, L. (2006). Aesthetic phenomena as supernormal stimuli: The case of eye, lip and lower-face size and roundness in artistic portraits. *Perception, 35*(2), 229–246.

Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of Personality and Social Psychology, 50*(5), 925–935.

Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C.-H. (1995). Their ideas of beauty are on the whole the same as ours? Consistency and variability in cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology, 68*(2), 261–279.

DiSalvo, C. F., Gemperle, F., Forlizzi, J. & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads. In *Proceedings of the conference on designing interactive systems: Processes, practices, methods, and techniques* (pp. 321–328), London.

Drury, N. E. (2000). Beauty is only skin deep. *Journal of the Royal Society of Medicine, 93*(2), 89–92.

Etcoff, N. (1999). *Survival of the prettiest: The science of beauty*. New York: Random House.

Farkas, L. G. (Ed.). (1994). *Anthropometry of the head and face*. New York: Raven Press.

Galton, F. (1879). Composite portraits, made by combining those of many different persons into a single resultant figure. *The Journal of the Anthropological Institute of Great Britain and Ireland, 8*, 132–144.

Giddon, D. B., Sconzo, R., Kinchen, J. A., & Evans, C. A. (1996). Quantitative comparison of computerized discrete and animated profile preferences. *The Angle Orthodontist, 66*(6), 441–448.

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

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience, 4*(8), 845–850.

Goldstein, A. C., & Papageorge, J. (1980). Judgments of physical attractiveness in the absence of eye movements. *Bulletin of the Psychonomic Society, 15*, 269–270.

Grammer, K., & Thornhill, R. (1994). Human (*Homo sapiens*) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology, 108*(3), 233–243.

Hanson, D. (2006). Exploring the aesthetic range of humanoid robots [electronic resource]. In *Toward social mechanisms of android science. The 28th annual conference of the cognitive science society long seminar* (pp. 16–20).

Ho, C.-C., MacDorman, K. & Pramono, Z. A. D. (2008). Human emotion and the uncanny valley: A GLM, MDS, and ISOMAP analysis of robot video ratings. In *Proceedings of the third ACM/IEEE international conference on human–robot interaction*, March 11–14, Amsterdam.

Johnson, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: Evidence for hormone-mediated adaptive design. *Evolution and Human Behavior, 22*(4), 251–267.

Jones, D. (1995). Sexual selection, physical attractiveness, and facial neoteny: Cross-cultural evidence and implications. *Current Anthropology, 36*(5), 723–748.

Jones, B. C., Little, A. C., & Perrett, D. I. (2004). When facial attractiveness is only skin deep. *Perception, 33*(5), 569–576.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average [electronic version]. *Psychological Science, 1*(2), 115–121.

Langlois, J. H., Roggman, L. A., Casey, R. J., Ritter, J. M., Rieser-Danner, L. A., & Jenkins, V. Y. (1987). Infant preferences for attractive faces: Rudiments of a stereotype [electronic version]. *Developmental Psychology, 23*(3), 363–369.

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley [electronic resource]. In ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science. Vancouver, Canada (http://www.macdorman.com/kfm/writings/pubs/MacDorman2006SubjectiveRatings.pdf).

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research [electronic version]. *Interaction Studies, 7*(3), 297–337.

Mealey, L., Bridgstock, R., & Townsend, G. C. (1999). Symmetry and perceived facial attractiveness: A monozygotic co-twin comparison. *Journal of Personality and Social Psychology, 76*(1), 151–158.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy, 7*(4), 33–35.

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion, 5*(4), 498–502.

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shapes and judgements of female attractiveness. *Nature, 368*(6468), 239–242.

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review, 5*(4), 659–669.

Salvia, J., Sheare, J. B., & Algozzine, B. (1975). Facial attractiveness and personal–social development. *Journal of Abnormal Child Psychology, 3*(3), 171–178.

Scheib, J. E., Gangstead, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London. Biological sciences, 266*, 1913–1917.