Electronic Supplementary Material for

Mara, M., Appel, M., & Gnambs, T. (2022). Human-Like Robots and the Uncanny Valley: A Meta-Analysis of User Responses Based on the Godspeed Scales. *Zeitschrift für Psychologie, 230*(1). https://doi.org/10.1027/2151-2604/a000486

Supplement Material for

Human-Like Robots and the Uncanny Valley:

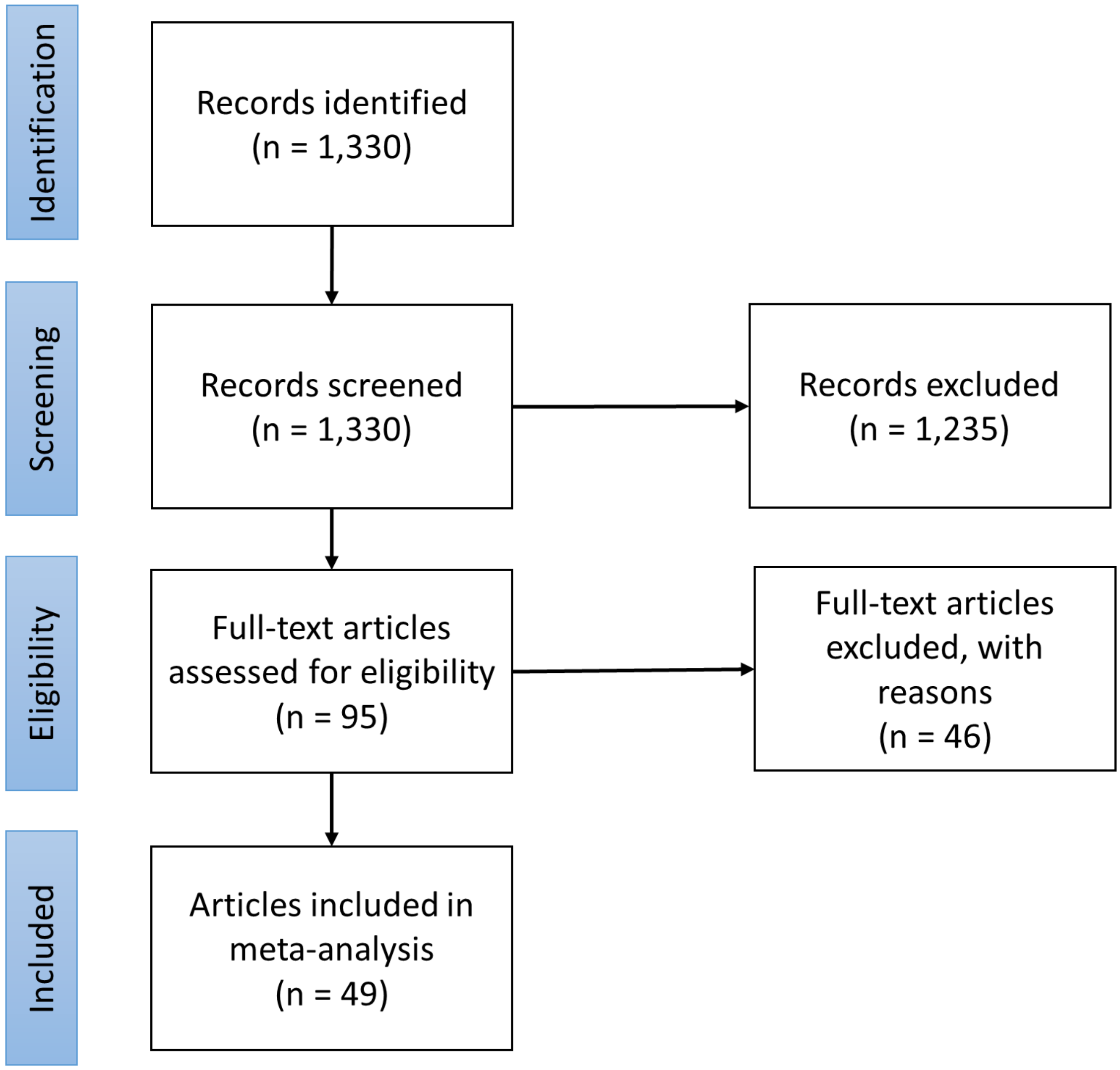
A Meta-Analysis of User Responses Based on the Godspeed Scales

Coded Variables

Variable	Description	Value	Example
study	Study ID: last name of first author + publication year	open text	schmidt2012
pubyear	Publication year	value range: [2009, 2020]	2012
sno	Unique ID for each sample	value range: [1,[	1
mno	Unique number of measure within sample	value range: [1,[	1
cntry	Country of origin of participants as ISO code	open text	DE
lang	Language of instrument	1 = English 2 = German 3 = Japanese 4 = other	1
lang2	Other language	open text	Klingon
pubtype	Publication type	1 = journal, 2 = presentation / proceedings, 3 = thesis (master/phd) 4 = book chapter 5 = other	2
robot	Description of the robot	open text	R2D2
n	Sample size	value range: [1,[	30
sample	Description of sample	open text	Undergraduates
samptype	Type of sample	0 = primarily students / university personnel 1 = general public 2 = children 3 = other	0
female	Percentage of women in sample (%)	value range: [0,100]	50
age	Mean age (in years) of participants	value range: [18,[	20
items	Number of items in anthropomorphism scale	value range: [1,[	5
m1	Mean of anthropomorphism scale	value range: [0,[	3
sd1	Standard deviation of anthropomorphism scale	value range: [0,[	1
se1	Standard error of anthropomorphism scale	value range: [0,[	1
alpha1	Cronbach's alpha for anthropomorphism scale	value range: [0,1]	0,8
items2	Number of items in likeability scale	value range: [0,[	5
m2	Mean of likeability scale	value range: [0,[	3
sd2	Standard deviation of likeability scale	value range: [0,[	1
se2	Standard error of likeability scale	value range: [0,[	1

Variable	Description	Value	Example
alpha2	Cronbach's alpha for likeability scale	value range: [0,1]	0,8
plot	Were statistics reported or derived from plots?	0 = reported 1 = from plots	0
page	Page of publication that the statistics are reported on	open text	p11
scale	Number of response scales of the administered items	value range: [2,[	5
mode	How was the robot presented?	0 = Physical presentation 1 = Photo 2 = Video 3 = other	0
mode2	How was the robot presented? other	open text	virtual environment
move	Did the robot move?	0 = not moving 1 = moving	0
talk	Did the robot talk?	0 = not talking 1 = talking	0
note	General comments	open text	

PRISMA Flow Diagram



Coded Data

Study	Year	Country	Robot	*N*	M_A	SE_A	M_L	SE_L	Real	Moving	Talking	Plot	Bias
Avelino et al. (2018)	2018	PT	Vizzy	21	3,29	0,09	4,39	0,1	yes	yes	no	yes	5
	2018	PT	Vizzy	22	2,94	0,07	4,06	0,13	yes	yes	no	yes	5
Barlas (2019)	2019	DE	NAO	30	2,62	0,11	4,04	0,01	yes	yes	yes	yes	5
	2019	DE	NAO	30	2,49	0,08	3,89	0,02	yes	yes	yes	yes	5
	2019	DE	NAO	24	2,61	0,16	4,04	0,02	yes	yes	yes	yes	5
	2019	DE	NAO	24	2,43	0,14	4,2	0,02	yes	yes	yes	yes	5
Busch et al. (2019)	2019	UK	ARMAR-6	6	2,5	0,37	3,3	0,12	no	yes	no	no	2
	2019	UK	ARMAR-6	7	2,09	0,43	3,23	0,18	no	yes	no	no	2
Churamani et al. (2017)	2017	DE	NICO	13	2,56	0,15	3,71	0,13	yes	yes	yes	yes	4
	2017	DE	NICO	14	2,58	0,09	4,11	0,09	yes	yes	yes	yes	4
Cuijpers et al. (2011)	2011	NL	NAO	14	3,44	0,21	4,13	0,22	yes	yes	no	yes	3
	2011	NL	NAO	14	3,48	0,19	3,92	0,24	yes	yes	no	yes	3
	2011	NL	NAO	14	3,63	0,17	4,16	0,22	yes	yes	no	yes	3
	2011	NL	NAO	14	3,63	0,22	4,26	0,17	yes	yes	no	yes	3
	2011	NL	NAO	14	3,46	0,2	4,14	0,22	yes	yes	no	yes	3
	2011	NL	NAO	14	3,57	0,21	4,2	0,2	yes	yes	no	yes	3
	2011	NL	NAO	14	3,33	0,17	3,9	0,2	yes	yes	no	yes	3
	2011	NL	NAO	14	3,35	0,22	3,92	0,24	yes	yes	no	yes	3
	2011	NL	NAO	14	3,58	0,19	4,04	0,24	yes	yes	no	yes	3
Foster et al. (2012)	2012		JAMES	31	2,39	0,13	3,73	0,17	yes	yes	yes	no	4
Fu et al. (2020)	2020	JP	CommU	12	2,81	0,2	3,57	0,28	yes	no	yes	no	1
	2020	JP	CommU	12	2,05	0,17	2,66	0,23	yes	no	yes	no	1
	2020	JP	CommU	12	3,28	0,21	4,07	0,15	yes	no	yes	no	1
	2020	JP	CommU	12	2,23	0,16	3,7	0,17	yes	no	yes	no	1
Ghiglino et al. (2020)	2020	IT	iCub	40	3,08	0,31	4,37	0,22	yes	yes	no	yes	3
	2020	IT	iCub	40	2,98	0,31	4,35	0,27	yes	yes	no	yes	3
	2020	IT	iCub	39	3,3	0,25	3,96	0,19	yes	yes	no	yes	3
	2020	IT	iCub	39	2,68	0,22	3,77	0,18	yes	yes	no	yes	3

Study	Year	Country	Robot	*N*	M_A	SE_A	M_L	SE_L	Real	Moving	Talking	Plot	Bias
Giuliani et al (2013)	2013		JAMES	14	1.99	0.16	2.63	0.30	yes	yes	yes	no	3
	2013		JAMES	26	1.72	0.11	3.44	0.17	yes	yes	yes	no	3
Ham et al. (2015)	2015	SG	NAO	16	2.17	0.12	3.52	0.09	yes	yes	no	yes	3
	2015	SG	NAO	16	2.44	0.14	3.49	0.16	yes	yes	no	yes	3
	2015	SG	NAO	16	2.46	0.16	3.54	0.12	yes	yes	no	yes	3
	2015	SG	NAO	16	2.30	0.16	3.68	0.23	yes	yes	no	yes	3
Haring et al. (2015)	2015	JP	Robi	20	2.54	0.13	3.8	0.19	yes	yes	no	no	2
	2015	JP	Robi	20	3.10	0.11	4.36	0.15	yes	yes	yes	no	2
	2015	JP	Robi	20	3.20	0.17	4.54	0.11	yes	yes	yes	no	2
	2015	AU	Robi	22	2.71	0.13	4.07	0.13	yes	yes	no	no	2
	2015	AU	Robi	22	2.64	0.19	4.24	0.16	yes	yes	yes	no	2
	2015	AU	Robi	22	2.96	0.22	4.09	0.25	yes	yes	yes	no	2
Haring et al. (2016)	2016	JP / AU	Geminoid-F	121	3.13	0.14	3.05	0.11	yes	yes	yes	no	4
	2016	JP / AU	Robi	64	2.31	0.10	4.11	0.11	yes	yes	yes	no	4
	2016	JP / AU	My Keepon	62	2.54	0.10	3.88	0.10	yes	no	no	no	4
Hoegen (2013)	2013	NL	Magabot	10	2.47	0.15	4.00	0.15	yes	yes	yes	no	3
	2013	NL	Magabot	11	1.97	0.19	3.51	0.14	yes	yes	yes	no	3
Iwashita & Katagami (2020)	2020	JP	Pepper	16	3.69	0.09	3.89	0.11	yes	yes	yes	yes	3
	2020	JP	Pepper	16	3.64	0.09	3.98	0.09	yes	yes	yes	yes	3
	2020	JP	Pepper	16	2.40	0.16	3.11	0.13	yes	yes	yes	yes	3
Johansson et al. (2020)	2020	NZ	EveR-4	46	2.19	0.10	3.62	0.13	yes	yes	yes	no	4
	2020	NZ	EveR-4	46	2.66	0.13	4.06	0.11	yes	yes	yes	no	4
	2020	NZ	EveR-4	45	2.14	0.10	3.78	0.12	yes	yes	yes	no	4
	2020	NZ	EveR-4	45	2.61	0.12	3.91	0.11	yes	yes	yes	no	4
Keizer et al. (2014)	2014	DE	JAMES	24	2.07	0.22	3.53	0.20	yes	yes	yes	no	4
Kerzel et al. (2020)	2020	DE	NICO	12	2.62	0.29	4.19	0.25	yes	yes	yes	no	3
	2020	DE	NICO	12	2.55	0.32	4.23	0.21	yes	yes	no	no	3
Kühnlenz (2013)	2013	DE	EDDIE	21	3.13	0.17	3.90	0.13	yes	yes	yes	no	3
	2013	DE	EDDIE	22	3.07	0.15	3.93	0.12	yes	yes	yes	no	3
	2013	DE	EDDIE	21	2.73	0.17	3.81	0.17	yes	yes	yes	no	3

Study	Year	Country	Robot	*N*	M_A	SE_A	M_L	SE_L	Real	Moving	Talking	Plot	Bias
	2013	DE	EDDIE	20	2.36	0.15	3.83	0.18	yes	yes	yes	no	3
Kühnlenz et al. (2013)	2013	DE	EDDIE	13	2.60	0.17	3.5	0.31	yes	yes	yes	no	3
	2013	DE	EDDIE	25	2.80	0.10	4.10	0.10	yes	yes	yes	no	3
	2013	DE	EDDIE	17	2.80	0.17	4.10	0.17	yes	yes	yes	no	3
Lehmann et al. (2016)	2016		iCub	14	2.63	0.06	4.11	0.05	no	no	yes	no	1
	2016		iCub	14	2.71	0.11	4.16	0.04	no	yes	yes	no	1
	2016		iCub	14	2.65	0.12	4.08	0.03	no	yes	yes	no	1
Lehmann et al. (2020)	2020	CZ	NAO	40	2.93	0.12	3.57	0.16	yes	yes	no	no	3
	2020	CZ	NAO	40	3.02	0.16	3.59	0.15	yes	yes	no	no	3
	2020	CZ	NAO	40	2.90	0.14	3.47	0.15	yes	yes	no	no	3
Löffler et al. (2019)	2019	DE	BlessU2	41	2.04	0.01	4.11	0.06	yes	yes	yes	yes	3
	2019	DE	QT	41	2.11	0.02	3.98	0.06	yes	yes	yes	yes	3
Lohse et al. (2013)	2013	NL	Magabot	40	2.54	0.11	3.54	0.09	yes	yes	no	no	3
Lugrin et al. (2018)	2018	DE	Robopec Reeti	20	2.31	0.21	3.74	0.23	yes	yes	yes	no	3
	2018	DE	Robopec Reeti	20	2.55	0.19	3.97	0.17	yes	yes	yes	no	3
Mazzola et al. (2020)	2020	IT	iCub	25	2.90	0.19	4.16	0.21	yes	yes	yes	no	5
	2020	IT	iCub	25	2.06	0.16	3.27	0.24	yes	yes	no	no	5
Meghdari et al. (2018)	2018	IR	Arash	14	4.14	0.21	4.90	0.08	yes	yes	yes	no	2
Mirnig et al. (2017a)	2017	AT	NAO	21	1.97	0.14	4.30	0.11	yes	yes	yes	no	4
	2017	AT	NAO	24	2.33	0.16	3.93	0.14	yes	yes	yes	no	4
Mirnig et al. (2017b)	2017	AT	NAO	113	2.1	0.12	3.6	0.13	no	yes	yes	no	3
Moon et al. (2013)	2013	UK	Barrett WAM robot	24	2.48	0.18	2.68	0.17	yes	yes	no	yes	4
	2013	UK	Barrett WAM robot	24	2.91	0.18	2.98	0.20	yes	yes	no	yes	4
	2013	UK	Barrett WAM robot	24	3.03	0.14	3.67	0.16	yes	yes	no	yes	4
	2013	UK	Barrett WAM robot	24	3.28	0.18	3.79	0.15	yes	yes	no	yes	4
	2013	UK	Barrett WAM robot	24	2.90	0.16	3.41	0.14	yes	yes	no	yes	4
	2013	UK	Barrett WAM robot	24	3.09	0.18	3.72	0.16	yes	yes	no	yes	4
Müller et al. (2017)	2017	DE	Virtual reality robot	76	2.06	0.08	3.19	0.07	no	yes	no	no	2
	2017	DE	Virtual reality robot	76	2.38	0.12	3.78	0.08	no	yes	no	no	2
Paetzel et al. (2020)	2020	SE	Furhead	16	3.65	0.17	3.50	0.17	yes	yes	yes	no	5

Study	Year	Country	Robot	*N*	M_A	SE_A	M_L	SE_L	Real	Moving	Talking	Plot	Bias
	2020	SE	Furhead	16	2.97	0.17	2.87	0.16	yes	yes	yes	no	5
Petrak et al. (2019)	2019	DE	Virtual reality robot	16	3.13	0.22	4.24	0.16	no	yes	no	no	3
	2019	DE	Virtual reality robot	16	2.36	0.26	2.82	0.27	no	yes	no	no	3
Rhim et al. (2019)	2019		Pepper	40	3.25	0.14	4.38	0.10	yes	yes	yes	no	3
	2019		Pepper	38	3.45	0.12	4.32	0.11	yes	yes	yes	no	3
Rosenberg-Kima et al. (2020)	2020	IS	NAO	36	2.51	0.11	3.64	0.12	yes	yes	yes	no	3
Rosenthal-von der Pütten et al. (2017)	2017	DE	NAO	20	2.01	0.22	4.06	0.17	yes	yes	yes	no	4
	2017	DE	NAO	20	2.26	0.22	4.44	0.16	yes	yes	yes	no	4
	2017	DE	NAO	20	2.01	0.17	4.98	0.18	yes	yes	yes	no	4
	2017	DE	NAO	20	2.33	0.23	4.31	0.15	yes	yes	yes	no	4
	2017	DE	NAO	20	2.4	0.19	4.14	0.15	yes	yes	yes	no	4
	2017	DE	NAO	20	2.69	0.19	4.48	0.09	yes	yes	yes	no	4
Rosenthal-von der Pütten et al. (2018)	2018	DE	NAO	20	2.10	0.20	3.90	0.20	yes	no	yes	no	4
	2018	DE	NAO	20	2.30	0.16	4.10	0.16	yes	no	yes	no	4
	2018	DE	NAO	20	2.50	0.25	4.20	0.18	yes	yes	yes	no	4
	2018	DE	NAO	20	2.50	0.22	4.20	0.13	yes	yes	yes	no	4
Ruitjen & Cuijpers (2018)	2018		Drone	64	2.13	0.13	2.99	0.13	no	no	no	no	3
	2018		Drone	64	2.48	0.14	3.49	0.14	no	yes	no	no	3
	2018		Drone	58	2.63	0.12	3.72	0.11	no	no	no	no	3
	2018		Drone	58	2.86	0.13	3.83	0.11	no	yes	no	no	3
Schneider (2019)	2019		NAO	20	2.62	0.19	4.73	0.08	yes	yes	yes	no	6
	2019		NAO	20	2.35	0.14	4.52	0.13	yes	yes	yes	no	6
Shariati et al. (2018)	2018	IR	Arash	20	3.99	0.19	4.84	0.08	yes	yes	yes	no	3
	2018	IR	Arash	20	3.82	0.18	4.81	0.07	no	yes	yes	no	3
Straßmann et al. (2020)	2020	DE	Pepper	22	2.27	0.19	3.80	0.16	yes	no	yes	no	4
	2020	DE	Pepper	22	2.53	0.12	4.11	0.14	yes	no	yes	no	4
	2020	DE	Pepper	22	2.13	0.14	3.72	0.14	no	no	yes	no	4
	2020	DE	Pepper	22	1.96	0.13	3.66	0.13	no	no	yes	no	4
Syrdal et al. (2013)	2013	UK	Sunflower housing robot	8	3.20	0.38	4.38	0.06	yes	yes	no	no	2
	2013	UK	Sunflower housing robot	8	2.88	0.35	3.93	0.07	yes	no	no	no	2

Study	Year	Country	Robot	*N*	M_A	SE_A	M_L	SE_L	Real	Moving	Talking	Plot	Bias
Trovato et al. (2015a)	2015	BR	KOBIAN	40	2.59	0.15	3.56	0.12	no	no	yes	no	6
Trovato et al. (2015b)	2015	BR	KOBIAN	20	1.20	0.20	4.65	0.22	no	no	yes	no	6
Ueno et al. (2020)	2020	CZ	Robot hand	23	2.93	0.19	3.30	0.17	yes	no	no	yes	3
Van der Hout (2017)	2017	NL	NAO	67	2.21	0.11	3.67	0.10	yes	yes	yes	no	4
	2017	NL	NAO	67	2.43	0.11	3.70	0.11	yes	yes	yes	no	4
Wieser et al. (2016)	2016		IRMA	20	2.95	0.15	4.28	0.13	yes	yes	no	no	4
Willemse & Wykowska (2019)	2019	IT	iCub	25	3.34	0.14	4.10	0.14	yes	yes	yes	no	4
	2019	IT	iCub	25	3.24	0.16	3.74	0.15	yes	yes	yes	no	4
Zanatto et al. (2019)	2019	UK	NAO	48	2.39	0.11	4.27	0.15	yes	yes	yes	no	5
	2019	UK	NAO	48	2.55	0.13	3.99	0.12	yes	no	no	no	5
Zanatto et al. (2020)	2020	UK	NAO	30	2.08	0.10	3.57	0.07	yes	yes	no	no	3
	2020	UK	NAO	29	2.71	0.10	3.66	0.11	yes	yes	no	no	3
	2020	UK	NAO	30	2.02	0.10	4.14	0.08	yes	yes	no	no	3

Note. $M_{A/L}$ = Mean anthropomorphism (A) or likability (L) score. $SE_{A/L}$ = Standard error for $M_{A/L}$. Real = Participants interacted with a real robot as compared to a photo or video. Plot = Statistics were reproduced from plots. Bias = Risk of bias using the ROBUST (Nudelman & Otto, 2020) codings.

Reliability Generalizations

The coefficient alpha reliabilities were pooled across samples with a random-effects meta-analysis using restricted maximum likelihood estimation. Because raw coefficient alphas are not normally distributed, we used the transformation and large sample variances suggested by Hakistan and Whalen (1976). To account for different test lengths (i.e., samples administering short versions), the coefficient alphas were corrected to a length of 5 items (i.e., as in the original scales) using the Spearman-Brown prophecy formula. Moreover, the average score variances were included as moderators in the meta-analytic models to adjust for range restriction (cf. Rodriguez & Maeda, 2006). The results of the two reliability generalizations in Table E1 show that both scales were generally reliable with pooled coefficient alphas of .85 und .88 for anthropomorphism and likability, respectively. For anthropomorphism, there was little variation between samples as indicated by the non-significant random component and the small value of I^2. Although the respective effect was slightly larger for likability (I^2 = 34%), unaccounted differences between samples can be considered moderate. Overall, these analyses highlight that, on average, both Godspeed scales exhibited satisfactory reliabilities in the studied samples.

Table E1

Reliability Generalizations of the Godspeed Anthropomorphism and Likability Scales

	Anthropomorphism	Likability
Number of samples	34	34
Pooled coefficient alpha	.850	.883
95% Confidence interval	[.830, .869]	[.867, .899]
95% Credibility interval	[.796, .894]	[.819, .930]
I^2	17.23%	34.03%
Test of residual heterogeneity	$Q(df = 32) = 35.817, p = .294$	$Q(df = 27) = 45.019, p = .063$
Test of moderator effects	$Q_m(df = 1) = 3.125, p = .077$	$Q_m(df = 1) = 0.328, p = .567$

Meta-Analyses of Godspeed Scale Scores by Robot

Differences in anthropomorphism and likability ratings between different robot models were examined by meta-analytically pooling the coded mean scores and using the robot model as a predictor in a meta-regression. We distinguished six robots for which ratings from at least three independent samples were available: the bartender robot JAMES (e.g., Foster et al., 2012; Giuliani et al., 2013), the iCub robot by the Italian Institute of Technology (e.g., Mazzola et al., 2020), the Magabot robot (e.g., Lohse et al., 2013), the NAO robot by SoftBank Robotics (e.g., Cuiijpers et al., 2011), the neuro-inspired companion robot NICO by the Knowledge Technology group at the University of Hamburg (e.g., Kerzel et al., 2020), and the Pepper robot by SoftBank Robotics (e.g., Iwashita and Katagami, 2020). To correct for potential setting effects, the presentation mode (real versus other) and whether the robot moved or communicated were included as covariates. The covariates were dummy coded, while the robot model was effect-coded to determine the difference of a specific model from the overall mean rating. For each scale, results of two meta-analytic models are presented (see Table E2): (a) a model that included only the covariates (Model 1) and (b) a model that additionally accounted for differences between the six robot models (Model 2).

The pooled anthropomorphism score across all robot models was $\mu = 2.64$, 95% CI [2.52, 2.76]. In line with Mori's hypothesis (Mori et al., 2012), moving robots were evaluated more human-like as compared to static robots. Moreover, robots seemed to be attributed more human-like characteristics when respondents interacted with a real robot as compared to simply viewing photos or videos of a robot. However, these effects were only significant after accounting for differences between robot types (Model 2). We also observed significant differences in anthropomorphism ratings between robot models. While the bartender robot JAMES was evaluated significantly less human-like as compared to the average evaluation, the iCub robot and Pepper were evaluated significantly more human-like (see Table E2). The robot model accounted for about 20% in the random variance of anthropomorphism ratings between samples.

The pooled likability score across all robot models was μ = 4.04, 95% CI [3.93, 4.14]. Robots that communicated with the respondents (e.g., talked) were evaluated significantly ($p < .05$) more likeable as compared to mute robots. Again, we also observed significant differences in likability ratings between robot models. While the bartender robot JAMES was evaluated significantly less likeable as compared to the average evaluation, the NAO robot was evaluated significantly more likeable (see Table E2). The robot model explained about 4% in the random variance of likability ratings between samples.

Table E2

Meta-Analyses of Godspeed Scale Scores by Robot Model

	Anthropomorphism		*Likability*	
	Model 1	Model 2	Model 1	Model 2
Intercept	2.64^{***} (0.06)	2.61^{***} (0.07)	4.04^{***} (0.05)	3.95^{***} (0.07)
Bartender robot [c]		-0.57^{**} (0.20)		-0.57^{**} (0.20)
iCub robot [c]		0.39^{**} (0.15)		0.25^{+} (0.14)
Magabot robot [c]		-0.28 (0.22)		-0.18 (0.21)
NAO robot [c]		-0.14 (0.09)		0.17^{*} (0.09)
NICO robot [c]		-0.04 (0.20)		0.16 (0.19)
Pepper robot [c]		0.40^{*} (0.16)		0.10 (0.15)
Presentation mode [a]	-0.17 (0.13)	-0.36^{**} (0.13)	-0.10 (0.11)	-0.12 (0.11)
Moving [b]	-0.22^{+} (0.12)	-0.32^{**} (0.12)	-0.15 (0.11)	-0.16 (0.11)
Communicating [b]	0.03 (0.10)	0.03 (0.03)	-0.26^{**} (0.08)	-0.29^{***} (0.08)
Random effect (τ^2)	0.36 / 0.27	0.31 / 0.25	0.30 / 0.25	0.29 / 0.25
I^2	95%	93%	96%	95%
R^2	3%	23%	9%	13%

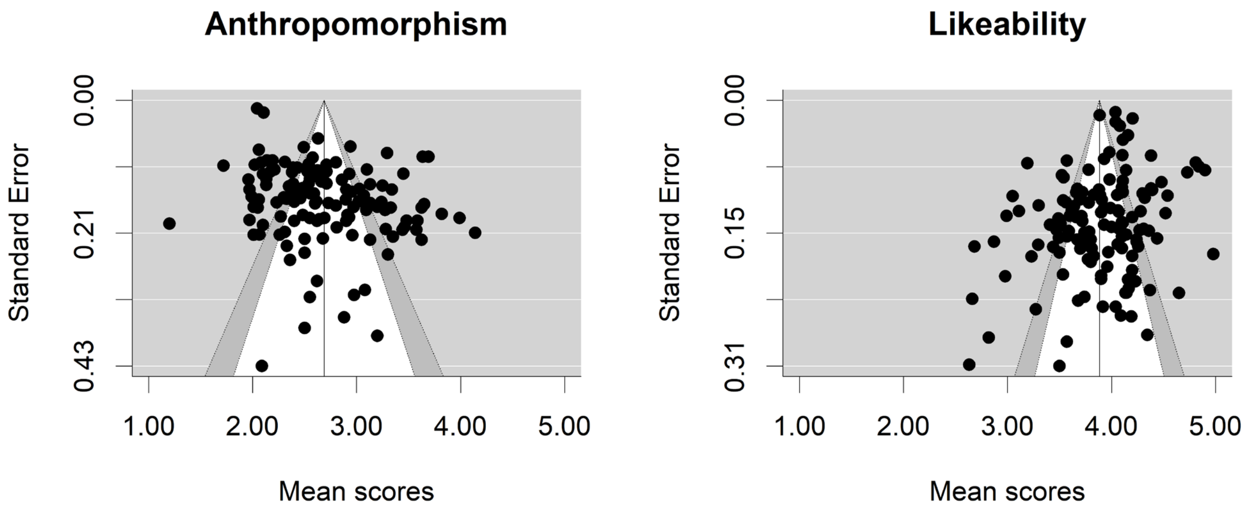
Note. Presented are meta-regression coefficients with standard errors in parentheses.
[a] 0 = physical, 1 = other ; [b] 0 = yes, 1 = no; [c] Effect-coded with other robots as reference category.
$^{***} p < .001$, $^{**} p < .01$, $^{*} p < .05$, $^{+} p < .10$

Analysis of Publication Bias

The presence and consequence of a potential publication bias were examined separately for the two Godspeed scales. The funnel plots in Figure E1 indicated a slightly asymmetric shape for the likability scores. However, this might be a consequence of a ceiling effect because many scores clustered in the upper region at the border of the scale limit. For anthropomorphism scores, a visual inspection of the funnel plot did not indicate a pronounced asymmetry.

Figure E1

Funnel Plots for Average Anthropomorphism and Likability Scores



The shapes of the funnel plots were tested for asymmetry using a regression test (Egger, Smith, Schneider, & Minder, 1997; Stanley, 2008) that predicted the mean scores from their standard errors. A significant effect would indicate an asymmetric shape of the funnel plot and potentially selective reporting. For anthropomorphism, the regression test suggested a skewed funnel plot (see Table E3). The pooled effect corrected for selective reporting ($\mu = 2.10$) was slightly smaller than the uncorrected effect ($\mu = 2.36$), indicating that some studies with low anthropomorphism ratings might be missing from the meta-analytic database. In contrast for likability, the test for funnel plot asymmetry was not significant (*p* = .058). Moreover, the corrected (μ = 4.06) and uncorrected effect (μ = 4.01) were rather similar which does not suggest pronounced reporting bias. Taken together, these analyses suggest that publication bias might have slightly distorted the publicly available research findings regarding anthropomorphism but did not give evidence of distortions for likability ratings.

Table E3

Regression Tests for Funnel Plot Asymmetry of the Godspeed Scale Scores

	Anthropomorphism		*Likability*	
	Model 1	Model 2	Model 1	Model 2
Intercept	2.36^{***} (0.04)	2.10^{***} (0.06)	4.01^{***} (0.03)	4.06^{***} (0.04)
Standard error		3.22^{***} (0.55)		-1.04^{+} (0.54)

$^{***}\ p < .001$, $^{**}\ p < .01$, $^{*}\ p < .05$, $^{+}\ p < .10$

Identification of Nonlinearity

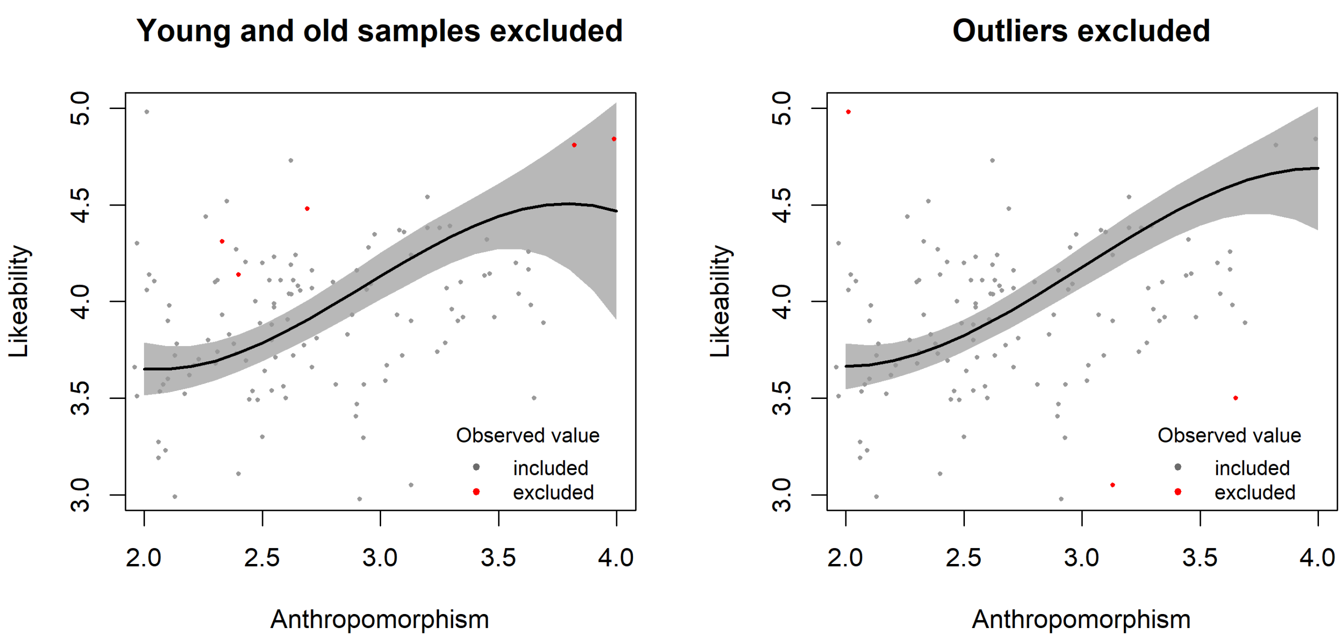
The optimal number of higher-order polynomials predicting likability from anthropomorphism was identified by comparing increasingly complex models. Models including polynomials of degree 1 to degree 6 resulted in Bayesian information criteria (BIC; Schwarz, 1978) of 130, 130, 129, 133, 138, and 142, respectively. The lowest BIC was observed for a model including polynomials of degree 3. The results of respective meta-regression analyses are summarized in Table 2.

Sensitivity Analyses

Following Voracek and colleagues (2019), we tried to determine the generalizability of the results with regard to various methodological choices. First, we repeated the meta-analyses excluding samples with children (Meghdari et al., 2018; Shariati et al., 2018) or older respondents (Rosenthal-von der Pütten et al., 2017). Because most studies relied on student samples that were rather homogenous regarding mean age, children and seniors might distort the effect estimates. However, the predicted effect with and without these samples was highly similar and replicated the curvilinear association between anthropomorphism and likability (see left panel in Figure E2). Then, we identified outliers using studentized residuals (cf. Viechtbauer & Cheung, 2010) and repeated the analyses excluding the three identified extreme values (Haring et al., 2016; Paetzel et al., 2020; Rosenthal-von der Pütten et al., 2017). Again, the resulted predicted effects between anthropomorphism and likability closely replicated the overall analyses (see right panel in Figure E2).

Figure E2

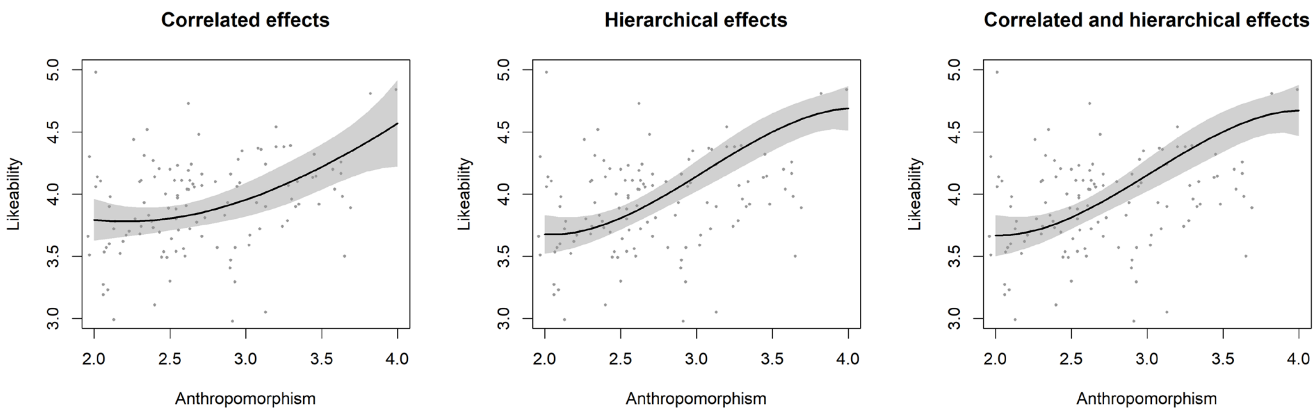
Predicted Effects Excluding Young and Old Samples or Outliers



Finally, we estimated meta-analytic models with cluster-robust standard errors (cf. Hedges et al., 2010). This involves two steps: First, preliminary standard errors are estimated using a working model that specifies a hypothesized dependency structure between observed effects. Then, the estimated standard errors are corrected for remaining unmodeled (unknown) dependencies using a sandwich estimator. Following Pustejovsky and Tipton (2021), we adopted three different working models that either assumed correlated effects, a hierarchical effect structure, or both. The predicted associations between likability and anthropomorphism for these analyses estimated with the *clubSandwich* package version 0.5.3 (Pustejovsky, 2021) are presented in Figure E3. Generally, the different modeling strategies lead to similar results; albeit ignoring a hierarchical effect structure seemed to exhibit a somewhat flatter increase. Thus, the choice of the analysis model does not substantially impact the observed results.

Figure E3

Predicted Effects Using Robust Meta-Analyses with Different Working Models



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 3
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 6
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 7
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 8
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 7
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 7
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	-
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 9
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Supplement
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Supplement
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Page 9
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Page 9
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	-
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	-
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	-
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 9

Section and Topic	Item #	Checklist item	Location where item is reported
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Page 10
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Page 10
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Supplement
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Page 10
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Page 11
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	-
Study characteristics	17	Cite each included study and present its characteristics.	Supplement
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Supplement
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Supplement
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Supplement
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Page 12-14 Supplement
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Supplement
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Supplement
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Supplement
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Supplement
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 14/15
	23b	Discuss any limitations of the evidence included in the review.	Page 16/17
	23c	Discuss any limitations of the review processes used.	Page 16/17
	23d	Discuss implications of the results for practice, policy, and future research.	Page 17
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Page 1
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 1

Section and Topic	Item #	Checklist item	Location where item is reported
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	-
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	-
Competing interests	26	Declare any competing interests of review authors.	Page 1
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Page 1/10

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

Additional References

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Hakstian, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika, 41*, 219-231. https://doi.org/10.1007/BF02291840

Pustejovsky, J. (2021). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections* (version 0.5.3). [R package]. https://CRAN.R-project.org/package=clubSandwich

Pustejovsky, J., & Tipton (2021). Meta-analysis with robust variance estimation: Expanding the range of working models. *MetaArxiv Preprints*. https://doi.org/10.31222/osf.io/vyfcj

Rodriguez, M. C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods, 11*(3), 306-322. https://doi.org/10.1037/1082-989x.11.3.306

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *Oxford Bulletin of Economics and Statistics, 70*, 103-127. https://doi.org/10.1111/j.1468-0084.2007.00487.x