Personality Design of Social Robots: The Role of the Mental Model Attribution Process

Supplementary Materials:

1.1 Definitions of concepts and theories synthesized in the MMAP

Table SM1: Definitions of concepts and theories synthesized in the MMAP

Concept	Definition	Source
Theory of mind (ToM)	ToM is the human ability to infer mental models to oneself and others.	(Premack and Woodruff 1978)
Mind perception	"Ascribing a mind to others". (p.103)	(Gray et al. 2012)
Mentalizing	Mentalizing describes the capability of inferring a mental state or theory of mind to others.	(Frith and Frith 2005)
Mental models	"Mental models are the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states" (Rouse and Morris 1986, p. 7). They represent individually constructed frameworks to help individuals describe, explain, and predict their surroundings.	(Kiesler and Goetz 2002; Phillips et al. 2011; Rouse and Morris 1986)
Mental states	Represent an individual's reality.	(Frith and Frith 2005)
Agency	Agency is the self-control, morality, memory, and emotion recognition capabilities of an agent.	(Gray et al. 2007)
Internal experience	Internal experience describes the desires, beliefs, and emotions of an agent.	(Gray et al. 2007)
Intentional agent	An intentional agent is a rational agent that acts according to their goals, intentions, and beliefs.	(Dennett 1996)
Intentionality	Intentionality or intentional stance is "the strategy of interpreting the behavior of an entity (person, animal, artifact, whatever) by treating it as if it were a rational agent who governed its 'choice' of 'action' by a 'consideration' of its 'beliefs' and 'desires'" (p. 27).	(Dennett 1996)
Anthropomorphism	The inference and attribution of human characteristics, traits, motivations, or qualities to non-human agents (i.e., gods, animals, objects). It indicates the non-human agent's potential for social interaction.	(Airenti 2015; Breazeal 2003; Damiano and Dumouchel 2018; Epley et al. 2007; Gambino et al. 2020; Oxford University 2018; Waytz et al. 2010)
Social cues	Social cues are a source of information that stimulates a social reaction from others.	(Feine et al. 2019; Nass and Moon 2000)

1.2 Robot Manipulation

Table SM2. Manipulation of the Spoken Words in the Animated Video

The manipulation of the spoken words		
Extroverted scenario	Introverted scenario	Difference
Communicating news		
Jamie starts blinking. Jamie: "Sorry to interrupt. I just received news. According to one of the third-party vendor's newsroom, they will change the access to their Natural Language Processing services, which will most likely affect our chatbot service. It's likely that WEBMEX will not be able to deliver in time. I prognose a 10-day delay in delivery."	Jamie starts blinking. You: "Do you have any news?" Jamie: "Yes I do. I just learned that according to the third-party vendor's newsroom, they will change the access to their Natural Language Processing services, which will most likely affect our chatbot service. I prognose a 10-day delay in delivery due to this unforeseen change."	Both robots are programmed to blink to signal that they received news. This blinking manifests in form of a lightbulb appearing over their head, which blinks, in the animated videos. In the extrovert scenario, the robot initiates the conversation, in comparison to the introvert scenario, where the robot shows that it has news through blinking, but it waits to speak until it is asked about the news. Additionally, the extrovert robot chooses an intonation which reflects its commitment to the project, while the introvert robot is more monotone and focuses on the facts.
Reaction to a proposition		
Jamie: "Awesome! The customers will love this. Alex, should I add these features to the product backlog, and you can discuss them during the next product backlog refinement meeting? I will quickly calculate the additional costs and time for this."	You: "Jamie could you calculate the additional costs and time for the additional features, please?"	In the extrovert scenario, the robot again initiates the conversation and is warm, enthusiastic, and very outgoing. It also shows engagement with the task. Furthermore, the robot does not act on its own but proposes an option and requests approval for actions that affects the project also at the outsourced company. In comparison, in the introvert scenario, the robot does not react to the proposition and has to be asked to start an action.
Calculating costs and times of the proposition		

Jamie: “Based on historical data, I estimate approximately one additional month of development time for the project. This will cost us more or less additional $8,000. But it is crucial to check with the development team at WEBMEX to get their estimate.”	“Analysis complete” sign appears on the whiteboard. You: “What did you find?” Jamie: “Here you can see the historic data from prior projects from WEBMEX, where they developed a one-click purchase button, a customer profile, and an algorithm. The first table shows the average costs and time of all prior projects. The second table refers to projects that have a similar scope to ours. Therefore, I estimate an additional 1 month and 1.47 days development time for the project. This will cost us additional $7.525,00. But it is crucial to check with the development team at WEBMEX to get their estimate.”	In the extrovert scenario, the robot focuses on the most important findings and takes for granted that the members can understand the tables and graphs displayed. Additionally, the robot rounds the numbers of the findings, while the introvert robot is concrete in its explanations of the graphs and presentation of its findings.
	Scheduling an appointment	
Jamie: “I have all of our calendars in the system and the earliest time all of us are available is Friday morning. Is that okay with you? If it is, I will call the assistant at WEBMEX, Ms. Jones, and check if the Scrum master, Eva, and Tony are available at this time.”	Jamie: “Our next available time slot is on Friday morning. Do you want me to confirm with the assistant at WEBMEX, Ms. Jones, if the Scrum master, Eva, and Tony are available at this time?”	When the extrovert robot schedules the appointment, it again uses a more engaged wording, which transport enthusiasm. The introvert, on the other hand, is cool and reserved while fulfilling the requested action.
	Leaving the meeting	
Jamie: “Okay. If you need anything just call me. Have a nice day. I will see you later.” Jamie waves goodbye and moves out of the room.	Jamie leaves the room.	When the robot is told that the meeting ends the introvert robot leaves as it strives more for alone time, while the extrovert still offers its services and exchanging pleasantries.

1.3 Screenshot video

a) b)

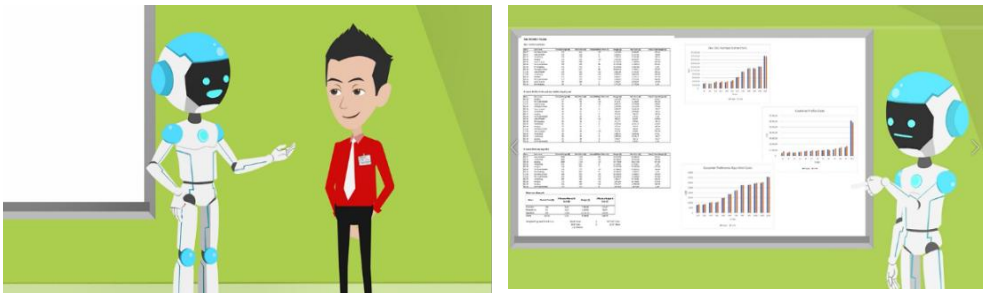

Fig. SM1 Video vignette same interaction point in time: a) extroverted robot, b) introverted robot

1.4 Items

Table SM3 Constructs and Items

Construct name *Source*	***# of Items***	***Code***	***Items***
Robot extroversion-introversion *Goldberg (1990)*	20	RExt1	Active*
		RExt2	Assertive*
		RExt3	Bold*
		RExt4	Daring*
		RExt5	Energetic*
		RExt6	Extraverted*
		RExt7	Talkative*
		RExt8	Unrestrained*
		RExt9	Verbal*
		RExt10	Vigorous*
		RInt1	Bashful (rev)
		RInt2	Inhibited (rev)
		RInt3	Introverted (rev)
		RInt4	Quiet (rev)
		RInt5	Reserved (rev)
		RInt6	Shy (rev)
		RInt7	Timid (rev)
		RInt8	Unadventurous (rev)
		RInt9	Untalkative (rev)
		RInt10	Withdrawn (rev)
Task satisfaction *Weiss et al. (1967)*	5	TSat1	I would feel fairly well satisfied with my job in the project.
		TSat2	Most days I would be enthusiastic about working in this project.
		TSat3	Each day of work in the project would seem like it will never end. (rev) *
		TSat4	I would find real enjoyment in my project work.
		TSat5	I would consider my job in the project rather unpleasant. (rev) *
Anthropomorphism *Bartneck et al. (2009)*	4	Anth1	Fake vs. natural
		Anth2	Machinelike vs. humanlike
		Anth3	Unconscious vs. conscious
		Anth4	Artificial vs. lifelike
Experience *Gray et al. (2011)*	3		How capable is Jamie of …
		Exp1	... experiencing pleasure
		Exp2	... experiencing hunger

		Exp3	... experiencing desire
Agency *Wooldridge and Jennings (1995)*	3/7		How capable is Jamie of …
		Agc1	... responding to perceived changes in the environment
		Agc2	... taking control over its own actions*
		Agc3	... pro-actively taking initiative*
		Agc4	... interacting with other human and computerized agents*
		Agc5	... learning new knowledge based on prior experiences
		Agc6	... using information from prior experiences in future interactions
		Agc7	... initiating and responding to interactions in a flexible manner*
Robot empathy *Pitt et al. (1995)*	5	Emp1	Jamie gives the team members individual attention.
		Emp2	Jamie has operating hours that are convenient to the team members.*
		Emp3	Jamie gives the team members personal attention.
		Emp4	Jamie has the team members' best interests at heart.
		Emp5	Jamie understands the team members' specific needs.
Attitude towards robots *Heerink et al. (2010) & Tay et al. (2014)*	3	Att1	I think it is a good idea to use the robot in a software development project.
		Att2	The robot would make my life more interesting.
		Att3	It is good to make use of the robot
Human extroversion *International Personality Item Pool (2018)*	10	HExt1	I am the life of the party.
		HExt2	I feel comfortable around people.
		HExt3	I start conversations.
		HExt4	I talk to a lot of different people at parties.
		HExt5	I don't mind being the center of attention.
		HExt6	I don't talk a lot. (rev)
		HExt7	I keep in the background. (rev)
		HExt8	I have little to say. (rev)
		HExt9	I don't like to draw attention to myself. (rev)
		HExt10	I am quiet around strangers. (rev)
Participant's gender	2		Please indicate the gender that describes you best: Male, Female
Attitude towards blue *Simmering et al. (2015)*	3	AttB1	I like the color blue.
		AttB2	I like blue clothes.
		AttB3	I prefer blue to other colors.

Note. (rev) Items were reverse coded. * Items were dropped from the survey.

1.5 Factor Loadings, CR, AVE, Correlations

Table SM4 Validity and Reliability Measures

	CR	AVE	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Robot Extroversion*	0.87	0.42								
(1) Robot Introversion (recoded)	0.93	0.58	**0.760**							
(2) Task Satisfaction	0.87	0.69	0.296	**0.832**						
(3) Empathy	0.83	0.55	0.362	0.399	**0.740**					
(4) Anthropomorphism	0.90	0.70	0.318	0.470	0.599	**0.838**				
(5) Attitude	0.88	0.71	-0.078	0.233	0.215	0.342	**0.844**			
(6) Human Extroversion	0.92	0.53	0.085	0.140	0.016	0.063	0.202	**0.730**		
(7) Experience	0.90	0.74	-0.426	0.007	0.215	0.434	0.195	0.091	**0.860**	
(8) Agency	0.83	0.62	0.132	0.175	0.257	0.290	0.124	0.034	0.018	**0.790**

Note. CR = Composite reliability, AVE = Average variance extracted, * Construct dropped from further analysis, Bold values equal to the square root of the AVE for testing the Fornell-Larcker criterion.

1.6 Common Method Bias

Table SM5 Common Method Bias

Model	χ^2 (df)	CFI	RMSEA (90% CI)	LR of $\Delta\chi^2$	Model comparison
CFA with marker	1160.7 (467)	0.908	.061 (.057, .065)		
Baseline	1175.5 (480)	0.908	.060 (.056, .065)		
Method-C	1173.0 (479)	0.908	.060 (.056, .065)	2.4, $df = 1$, $\rho = 0.118$	vs. Baseline
Method-U	1131.1 (450)	0.910	.062 (.057, .066)	41.9, $df = 29$, $\rho = 0.057$	vs. Method-C
Method-R	1197.4 (471)	0.904	.062 (.058, .066)	66.3, $df = 21$, $\rho < .001$	vs. Method-U

Note. CFA = confirmatory factor analysis, CFI = comparative fit index, RMSEA = root mean square error of approximation, LR = likelihood ratio test, U = unconstraint, C = common, R = restricted

1.7 Summary of the results

Table SM6 Results of hypotheses testing

Hypothesis / Path	Path coeff.	95% CI	Supported
H1 Robot extroversion → Anthropomorphism	**0.641**	[0.351, 0.924]	Y
H2a Anthropomorphism → Agency	**0.173**	[0.120, 0.222]	Y
H2b Anthropomorphism → Experience	**0.167**	[0.116, 0.222]	Y
H3a Agency → Robot empathy	**0.432**	[0.282, 0.569]	Y
H3b Experience → Robot empathy	**0.124**	[0.035, 0.211]	Y
H4 Robot extroversion → Task satisfaction (direct effect)	-0.044	[-0.184, 0.105]	
Robot extroversion → Anthropomorphism → Experience → Perc. robot empathy → Task satisfaction (indirect effect)	**0.005**	[0.001, 0.010]	Y
Robot extroversion → Anthropomorphism → Agency → Perc. robot empathy → Task satisfaction (indirect effect)	**0.017**	[0.007, 0.031]	Y

Bibliography:

Airenti G (2015) The Cognitive Bases of Anthropomorphism: From Relatedness to Empathy. Int J Social Robotics. 7(1): 117-127 https://doi.org/10.1007/s12369-014-0263-x

Bartneck C, Kulic D, Croft E, Zoghbi S (2009) Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. Int J Social Robotics. 1: 71-81 https://doi.org/10.1007/s12369-008-0001-3

Breazeal C (2003) Towards Sociable Robots. Robotics and Autonomoous Systems. 42: 167-175.

Damiano L, Dumouchel P (2018) Anthropomorphism in Human–Robot Co-evolution. Front in Psychol. 9 https://doi.org/10.3389/fpsyg.2018.00468

Dennett DC (1996) Kinds of Minds: Toward an Understanding of Consciousness. United States of America, BasicBooks A Member of the Perseus Books Group.

Epley N, Waytz A, Cacioppo JT (2007) On Seeing Human: A Three-Factor Theory of Anthropomorphism. Psychol Review. 114(4): 864-886 https://doi.org/10.1037/0033-295X.114.4.864

Feine J, Gnewuch U, Morana S, Maedche A (2019) A Taxonomy of Social Cues for Conversational Agents. Int J Hum Comp Studies. 132: 138-161 https://doi.org/10.1016/j.ijhcs.2019.07.009
Frith C, Frith U (2005) Theory of mind. Curr Biol. 15(17): 644-645 https://doi.org/10.1016/j.cub.2005.08.041
Gambino A, Fox J, Ratan RA (2020) Building a stronger CASA: Extending the computers are social actors paradigm. Hum-Machine Commun. 71-85 https://doi.org/10.3316/INFORMIT.097034846749023
Goldberg LR (1990) An Alternative "Description of Personality": The Big-Five Factor Structure. J Pers and Psychol. 59(6): 1216-1229 https://doi.org/10.1037/0022-3514.59.6.1216
Gray HM, Gray K, Wegner DM (2007) Dimensions of Mind Perception. Sci. 315(December 2016): 619-619 https://doi.org/10.1007/s11097-010-9186-7
Gray K, Knobe J, Sheskin M, Bloom P, Barrett LF (2011) More than a body: mind perception and the nature of objectification. J Pers and Soc Psychol. 101(6): 1207.
Gray K, Young L, Waytz A (2012) Mind Perception Is the Essence of Morality. Psychological Inquiry. 23(2) https://doi.org/10.1080/1047840X.2012.651387
Heerink M, Kröse B, Evers V, Wielinga B (2010) Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. Int J Social Robotics. 2: 361-375 https://doi.org/10.1007/s12369-010-0068-5
International Personality Item Pool (2018). *Big-Five Factor Markers.*
Kiesler S, Goetz J (2002) Mental Models and Cooperation with Robotic Assistants. CHI'02 extended abstracts on Human factors in computing systems. 576-577 https://doi.org/10.1145/506443.506491
Nass C, Moon Y (2000) Machines and mindlessness: Social responses to computers. J Social Issues. 56(1): 81-103 https://doi.org/10.1111/0022-4537.00153
Oxford University P (2018). *Anthropomorphism. Oxford Learner's Dictionaries.*
Phillips E, Ososky S, Grove J, Jentsch F (2011) From Tools to Teammates: the Development of Appropriate Mental Models for Intelligent Robots. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 1491-1495 https://doi.org/10.1177/1071181311551310
Pitt LF, Watson RT, Kavan CB (1995) Service Quality: A Measure of Information Systems Effectiveness. MISQ. 19(2): 173-187 https://doi.org/10.2307/249687
Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? Behav Brain Sci. 1(4): 515-526 https://doi.org/10.1017/S0140525X00076512
Rouse WB, Morris NM (1986) On Looking into the Black Box: Prospects and Limits in the Search for Mental Models. Psychological bulletin. 100(3): 1-349.
Simmering MJ, Fuller CM, Richardson HA, Ocal Y, Atinc GM (2015) Marker Variable Choice, Reporting, and Interpretation in the Detection of Common Method Variance: A Review and Demonstration. Organ Research Methods. 18(3): 473-511 https://doi.org/10.1177/1094428114560023
Tay B, Jung Y, Park T (2014) When Stereotypes Meet Robots: The Double-Edge Sword of Robot Gender and Personality in Human – Robot Interaction. Comput in Hum Behav. 38: 75-84 https://doi.org/10.1016/j.chb.2014.05.014
Waytz A, Cacioppo J, Epley N (2010) Who sees human? The stability and importance of individual differences in anthropomorphism. Perspectives on Psychological Science. 5(3): 219-232 https://doi.org/10.1177/1745691610369336
Weiss DJ, Dawis RV, England GW (1967) Manual for the Minnesota Satisfaction Questionnaire.
Wooldridge M, Jennings NR (1995) Intelligent agents: Theory and practice. Knowl Eng Rev. 10(2): 115-152.