

Association for Information Systems

AIS Electronic Library (AISeL)

Wirtschaftsinformatik 2021 Proceedings

Track 11: Human Computer Interaction

Let us work Together' – Insights from an Experiment with Conversational Agents on the Relation of Anthropomorphic Design, Dialog Support, and Performance

Sascha Lichtenberg
Universität Göttingen

Johannes Bürke
Universität Göttingen

Alfred Benedikt Brendel
Technische Universität Dresden

Simon Trang
Universität Göttingen

Stephan Diederich
Universität Göttingen

See next page for additional authors

Follow this and additional works at: <https://aisel.aisnet.org/wi2021>

Lichtenberg, Sascha; Bürke, Johannes; Brendel, Alfred Benedikt; Trang, Simon; Diederich, Stephan; and Morana, Stefan, "Let us work Together' – Insights from an Experiment with Conversational Agents on the Relation of Anthropomorphic Design, Dialog Support, and Performance" (2021). *Wirtschaftsinformatik 2021 Proceedings*. 4.

<https://aisel.aisnet.org/wi2021/PHuman/Track11/4>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik 2021 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Presenter Information

Sascha Lichtenberg, Johannes Bürke, Alfred Benedikt Brendel, Simon Trang, Stephan Diederich, and Stefan Morana

‘Let us work Together’ – Insights from an Experiment with Conversational Agents on the Relation of Anthropomorphic Design, Dialog Support, and Performance

Sascha Lichtenberg¹, Johannes Bührke¹, Alfred Benedikt Brendel², Simon Trang¹,
Stephan Diederich¹, Stefan Morana³

¹ Georg-August-Universität Göttingen, Chair of Information Management, Goettingen,
Germany

{Sascha.Lichtenberg, Strang}@uni-goettingen.de
{johannes.buehrke, stephan.diederich}@stud.uni-goettingen.de

² Technische Universität Dresden, Chair of Business Informatics, esp. Intelligent Systems and
Services, Dresden, Germany

{Alfred_Benedikt.Brendel}@tu-dresden.de

³ Universität des Saarlandes, Chair of Digital Transformation and Information System,
Saarbrücken, Germany

{stefan.morana}@uni-saarland.de

Abstract. In the human interaction with CAs, research has shown that elements of persuasive system design, such as praise, are perceived differently when compared to traditional graphical interfaces.

In this experimental study, we will extend our knowledge regarding the relation of persuasiveness (namely dialog support), anthropomorphically designed CAs, and task performance. Within a three-conditions-between-subjects design, two instances of the CA are applied within an online experiment with 120 participants. Our results show that anthropomorphically designed CAs increase perceived dialog support and performance but adding persuasive design elements can be counterproductive. Thus, the results are embedded in the discourse of CA design for task support.

Keywords: Conversational Agents, Persuasive System Design, Task Performance, Dialog Support, Chatbot, human computer interaction

1 Introduction

Information Systems (IS) can be designed to attain various goals. Following Benbasat [1], one of the goals is to increase the effectiveness and efficiency of users in the completion of a task, such as finding and purchasing a product online. However, IS also exhibits substantial potential to influence individual beliefs and behavior [2], for instance, regarding environmental sustainability [3] or health [4]. Studies in the context of persuasive systems and their design have received increasing attention recently, which is reflected in calls for more research [5].

While the vast majority of studies in the area of persuasive system design focuses on software with graphical user interfaces [6], we follow the notion that conversational agents (CAs) offer the opportunity to design even more persuasive IS. CAs, defined as software with which users interact through natural language (i.e. written or spoken word) [7], have been shown to trigger mindless social responses (i.e. users treat computers like it is a human being [8]) as formulated in the paradigm of computers-are-social-actors (CASA) [8], [9]. Due to the social nature of human interaction with CAs, we argue that elements of persuasive system design, such as praise or social roles [10], can be leveraged to influence individual behavior.

Initial work in the area of persuasive and anthropomorphic CAs underlines this potential. For example, Diederich, Lichtenberg, et al. [11] investigated how persuasive messages of a CA can influence an individual's environmental sustainability beliefs, finding that an anthropomorphic design of a CA increase the perceived persuasiveness. Similarly, Gnewuch et al. [12] argue that CAs can be a useful means to enable more sustainable energy consumption behavior of consumers, due to their feedback provided for the user. However, we still lack an understanding of whether persuasive CAs can extend beyond the scope of emotion and cognition, influencing actual user behavior (e.g., task performance).

In this experimental study, we address this research gap regarding the relation of persuasive, anthropomorphic CAs, and actual behavior in the form of performance. The performance of an individual can be measured by the number of completed tasks (e.g., in the context of gamification, by completed rounds [13], or the number of steps per day [14]). We conducted an experiment with three different treatment groups (no CA, anthropomorphic CA and anthropomorphic CA extended with persuasive features) in a task completion setting. Specifically, participants had to complete a certain number of tasks, with the option to voluntarily complete more of them. Against this background, this study aims to answer the following research question:

RQ: How can persuasive and anthropomorphic design of conversational agents positively influence performance?

2 Research Background

The following section contains the relevant background information for understanding this work: (1) persuasive system design and performance and (2) anthropomorphic conversational agents and social response theory.

2.1 Persuasive System Design and Performance

The observation that technology can influence human cognition, emotion, and behavior has been made around two decades ago. On this basis, the paradigm of CASA [9], [11], [16] has been formulated. The paradigm of CASA posits that individuals mindlessly apply social rules and expectations to computers once they receive cues associated with human traits or behavior [17]. Against this background, research in the domain of persuasive design investigates the social responses people show to computers [9], [10].

Research in this context entails the development and application of design elements intended to shape user perception and promote desired behavior. An example of this is the display of anthropomorphic communication features, such as humor, empathy, and praise, to trigger social dynamics, such as competition or cooperation [10].

These persuasive design elements can be distinguished into five types of social cues Fogg [2]: physical (e.g., touch, facial expressions, movement), psychological (e.g., empathy, humor), language (e.g., written or spoken language, turn-taking in a conversation), social dynamics (e.g., praise, judgment), and social roles (e.g., guide, partner). In sum, designers are provided with a wide selection of design elements and cues that can be used to persuade individuals in a variety of application domains, such as environmental sustainability, work, or education [18]. Regarding the effects of these social cues, four different categories can be distinguished [10]: (1) primary task support (e.g., individual tailoring of information), (2) dialog support (e.g., providing praise), (3) credibility support (e.g., displaying trustworthiness), and (4) social support (e.g., referring to social facilitation).

In the domain of work and performance, persuasive design offers the opportunity to incline individuals to perform their primary task [10]. In the context of performance, for instance, this can mean enabling an individual to measure their primary task progress via self-monitoring [6] (e.g., displaying heart rate while exercising to ensure progress and commitment [19]). Similar examples can be found in the context of the academic performance of students [20], promoting physical activity at the workplace [4] and provoke “work-like” performance in experimental contexts [21], [22]. Dialog support has shown that users are encouraged to use the enhanced IS and consecutively motivated to perform their primary task [23]. One example is praise in the form of images, symbols, or words [6] to support a person in achieving his or her goals (e.g., increase the number of steps per day [14]).

2.2 Anthropomorphic Conversational Agents and Social Response Theory

Through technological progress regarding natural language processing and machine learning, CA-related technology has become widely available [24]. Consequently, CAs are currently attracting strong interest from research and practice [7], [24], [25]. Users can interact with CAs using written (e.g., chatbots) or spoken language (e.g., personal assistants like Siri or Microsoft Cortana). Furthermore, CAs can be disembodied, have a virtual embodiment [26], or a physical embodiment, e.g. service robots [27]. Through various means, CAs can display human characteristics, such as having a human name and participating in a dialogue with turn-taking [28]. These anthropomorphic characteristics trigger mindless social responses by users [28], [29], as postulated in the social response theory [17], [30].

The intensity of these social responses varies according to the degree of perceived anthropomorphism (i.e., human-likeness) of a CA [31]. Current studies on CA design found that a higher degree of anthropomorphism can lead to various positive effects, such as an increase in service satisfaction [32], trustworthiness [33], and persuasiveness [11]. In order to better understand the relation of anthropomorphic CA design, perceived anthropomorphism, and related benefits, CAs are studied in various

application areas, such as customer service (e.g., marketing and sales [34]), and healthcare [35]). Synthesizing current research on anthropomorphic CA design, Seeger et al. [15] developed a conceptual framework that comprises three dimensions: (1) human identity, (2) verbal cues, and (3) non-verbal cues. The dimension of *human identity* includes cues regarding the representation of the agent, for example, having an avatar [31]. The second dimension of *verbal cues* comprises the language used by a CA, for instance, using self-references (“I think that...” [36]), expressing artificial thoughts and emotions (“In my experience...” [37]), or variability in syntax and word choice [15]. The third dimension of *non-verbal cues* includes conveying information on attitudes or emotional state [38], such as indicating thinking through dynamic response times depending on message length and complexity [32] or using emoticons to express emotions [39].

3 Research Model and Hypotheses

Our research will contribute to a better understanding of the relation between CA design, its perception, and user performance. Our research model is depicted in Figure 1. Specifically, we hypothesize that CAs equipped with social cues as part of an anthropomorphic design [15] persuade users to complete a higher number of tasks when combined with persuasive design elements, such as dialog support [40].

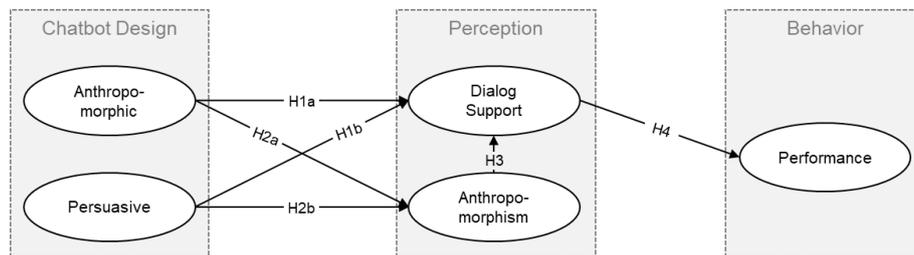


Figure 1. Research Model

Based on the paradigm of CASA [17], [30], technology influences individual beliefs and behavior [2]. CAs equipped with anthropomorphic characteristics, such as a human name and participating in a dialogue with turn-taking [28], trigger social responses by users [28], [29]. The human appearance leads individuals to perceive the CA as more persuasive, giving it the potential to influence the beliefs and behavior of individuals. Specifically, CAs provide users with the option to interact with the system via written dialog, providing dialog support [23]. Thus, we formulate the following hypothesis:
H1a: An anthropomorphically designed chatbot yields a higher level of perceived dialog support than no chatbot.

In the context of this study, we focus on CAs that are praising the user for their performance and award points for certain achievements, thereby providing dialog support [23]. Kamali et al. [41] were able to show that praise was expected (i.e., for specific behavior) when elderly people interact with a CA. Similarly, receiving points

for certain behavior increases participation [42]. Therefore, we formulate our next hypothesis as follows:

H1b: A persuasively and anthropomorphically designed chatbot yields a higher level of perceived dialog support than an anthropomorphically designed chatbot.

Furthermore, CAs offer various possibilities for anthropomorphic design. An agent equipped with a name, gender, and avatar [31], displaying emotions through verbal cues [8], and applying nonverbal cues, such as dynamic response delays to indicate thinking or typing [32], can contribute to the perception of the agent as more anthropomorphic, even when users are aware of the artificial nature of it. Thus, we propose the following hypothesis:

H2a: An anthropomorphically designed chatbot yields a higher level of perceived anthropomorphism than no chatbot.

Furthermore, CAs additionally displaying persuasive cues, such as praising their user, add further to the anthropomorphic perception [10]. For instance, the study of Xu and Lombard [43] have shown that even a small cue (e.g., the name of the CA) can change the perception of the CA. Therefore, we hypothesize that such cues contribute to users anthropomorphizing the agent:

H2b: A persuasively and anthropomorphically designed chatbot yields a higher level of perceived anthropomorphism than an anthropomorphically designed chatbot.

Recent studies, which explore the interaction of anthropomorphic design of CAs and their persuasiveness, suggest that perceived anthropomorphism can increase the persuasiveness of the agent. For instance, Harjunen et al. [44] found that virtual offers are more likely to be accepted when the agent shows typical human behavior, such as smiling or touching (with a haptic glove). Similarly, Adler et al. [45] showed that a CA displaying positive emotions leads to a higher degree of perceived persuasiveness compared to a CA without emotionally loaded language. Against this background, we hypothesize:

H3: Perceived anthropomorphism positively impacts perceived dialog support.

Following Lehto et al. [23], persuasive design elements have the potential to reinforce, change, or shape the behavior of individuals by increasing the overall persuasiveness of information systems. Superficially, dialog support has shown to encourage users to perform their primary task, such as increasing the amount of physical exercise [14]. Thus, we propose the following hypothesis:

H4: Perceived dialog support positively impacts performance.

4 Research Design

To test our hypotheses, we conducted an online experiment with three-conditions (no design, anthropomorphic design, and persuasive design) in a between-subjects design, avoiding carryover effects [46]. We conducted an a priori power analysis using GPower [47] to estimate the required sample size. We assume a large effect and estimated a minimum amount of 102 participants, given an effect size $f = 0.4$, $\alpha = .05$ and power (beta) = 0.95). We collected data from the 2nd to the 15th of October 2019 until we had at least forty observations per treatment, resulting in a total of 120 participants. Overall,

the sample consisted of 37% of females (5% of the participants preferred not to specify their gender). The age of the participants ranges from 18 to 83 (mean 33), and all participants are currently residing in Germany.

4.1 Data Collection Procedure and Sample

The experiment consisted of four steps: (1) Explanation of the experiment, (2) chat with the chatbot, (3) perform the task, and (4) fill out the questionnaire. In the first step, the participants received a briefing screen, which explained the context [48] and the structure of the experiment (completing five of 15 slider tasks with a subsequent survey) and described the tasks. Every participant received the same explanations to make sure that all participants have the same information [49]. Following the instructions, participants got two attempts to answer three comprehension questions. Those who failed both attempts were excluded from the experiment. This procedure ensures that no participant completed more tasks because the rules related to the number of completed tasks were not understood properly. After this step, all participants were randomly assigned to one of the three treatments and proceeded to step 2. The second step is divided into two sub-steps: (2a) chat with chatbot and (2b) perform the task. In step 2a, the participants had to chat with a chatbot. Via the chatbot, participants were able to start a task and end the experiment (see Control and Treatment Configuration section for details). If the participant was not in a chatbot treatment, the start of a task, and the end of the experiment could be triggered by a button. In step 2b, users had to perform slider tasks [48]. For the slider task, the participants had to set five sliders from 0 to 50 by using the mouse pointer. After completing each task, the participants returned to step 2a. When five tasks were completed, participants had the option to proceed to the questionnaire or complete up to ten more tasks. In step (3), participants had to fill out a questionnaire (see Measures section for details).

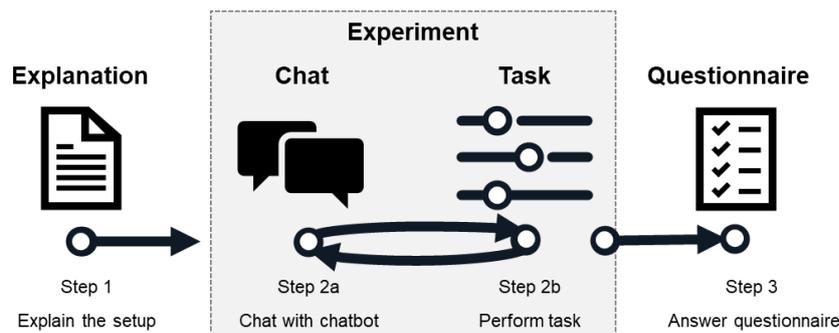


Figure 2. Procedure of the Experiment

4.2 Control and Treatment Configurations

Our experiment had three conditions: (1) no chatbot (control treatment), (2) anthropomorphic chatbot, and (3) persuasive chatbot. Every participant was randomly

assigned to one experimental condition (between-subjects design). For condition (1), users did not have the option to communicate with a chatbot. For conditions (2) and (3), two chatbots were developed via the natural language processing platform Dialogflow by Google. Both chatbots received the same training phrases (i.e., exemplary statements that users might make during the interaction) to train them to understand a user's intent and provide the correct reply. The chatbots were able to process different variations of sentences with the same meaning and could extract parameters, such as the intention to proceed to the next task or to exit the experiment and react appropriately. We further implemented a custom-built web interface to provide convenient access to the chatbots, ensure device independence, and minimize distraction.



Figure 3. Slider Task

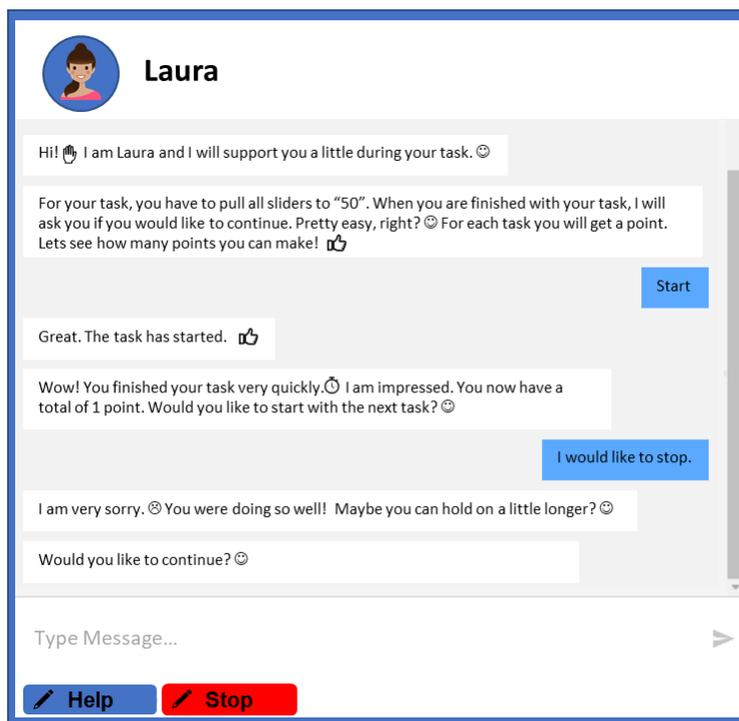


Figure 4. Persuasive Chatbot

Both chatbots were equipped with various cues for anthropomorphic CA design according to the three dimensions (human identity, verbal, non-verbal) as suggested by Seeger et al. [19] to establish a baseline for perceived anthropomorphism. Regarding the human identity, we equipped the chatbot with the name “Laura,” a female gender, and a human pictogram representing a female individual. Concerning verbal communication, the CA was designed to use self-references, turn-taking, and a personal introduction (“Hi! I am Laura and I will...”), including a greeting in the form of a welcome message. Regarding the non-verbal anthropomorphic CA design dimension, we implemented blinking dots in combination with dynamic response delays depending on the length of the previous message to simulate thinking and typing of replies by the CAs [32].

Overall, both chatbot instances were identical except for the addition of persuasive messages for condition (3). The chatbot provides dialog support by using praise, suggestions, and rewards [10]. The persuasive chatbot praises users after every task completed (“Wow! You finished your task very quickly.”), whereas the anthropomorphic chatbot renounces praise. Furthermore, in case users want to end the experiment and proceed to the questionnaire, the chatbots suggests continuing and completing more tasks (“Maybe you can hold on a little longer? Would you like to continue?”). Lastly, the chatbot introduces a point system, rewarding the user with one point for every completed task (“You now have a total of X points”).

4.3 Measures and Descriptive Statistics

Our research variables included experimentally manipulated variables, questionnaire-based variables (i.e., dialogue support and control variables), and the task outcome variable.

Table 1. Questionnaire Items (*Note that the items are translated from German to English.*)

Constructs and Items	FL	REF
Perceived Dialogue Support ($\alpha = .911$)		
I believe that the tool has supported me with appropriate feedback.	.873	
I believe that the tool has encouraged me to continue working on the task.	.909	[23]
I believe that the tool motivated me to complete the task by praise.	.889	
Perceived Anthropomorphism ($\alpha = .934$)		
I believe that the tool has a mind.	.759	
I believe that the tool has a purpose.	.305	
I believe that the tool has free will.	.909	[15]
I believe that the tool has a consciousness.	.926	
I believe that the tool desires something.	.857	
I believe that the tool has beliefs.	.912	
I believe that the tool has the ability to experience emotions.	.602	
Perceived Persuasiveness (Single Scale)		
I believe that the tool convinced me to perform the task.	-	[23]

FL = factor loadings, REF = reference, α = Cronbach's alpha;

First, the effect of the experimentally manipulated variables for the different types of chatbots. As the three treatments build on one another, we detangled the different

effects and coded variables that capture commonalities and differences between the treatments. Second, dialog support, anthropomorphism, and control variables in terms of age, gender, education, and experience with chatbots were captured using a questionnaire. All items were measured on a scale from 1 (strongly disagree) to 7 (strongly agree). For the design of the survey, only established constructs from previous studies were considered. Additionally, we included attention checks by asking two questions that prompt the participant to select a specific number on a scale. If the participant failed to answer the questions correctly, the data was not considered for the analysis. Perceived dialog support was measured using a 7-Point Likert scale adapted from [23]. Perceived anthropomorphism is based on a 7-Point Likert scale adapted from [15]. Additionally, we measured perceived persuasiveness [23] as a single-scale item to conduct a manipulation check. The items are displayed in Table 1. Third, the outcome variable of the task was measured in terms of the number of completed tasks, where the number of completed tasks equals the times a participant positioned all sliders correctly.

5 Results

In the following two sub-sections, we will present our results regarding the descriptive statistics and structural model.

5.1 Descriptive Statistics

The group averages of the performance show that the anthropomorphic chatbot ($M=7.375$, $SD=5.309$) and anthropomorphic chatbot with persuasive elements ($M=4.3$, $SD=2.893$) differ from the control group, which yields a lower number of tasks performed ($M=3.150$, $SD=3.519$). Similarly, we observed that the perceived dialog support is lower for the control group ($M=2.45$, $SD=1.693$) when compared to the anthropomorphic chatbot ($M=5.15$, $SD=1.743$) and anthropomorphic chatbot with persuasive elements ($M=2.858$, $SD=1.571$). As for anthropomorphism, the system is perceived lower in the control group ($M=2.107$, $SD=1.318$) when compared to the treatments anthropomorphic chatbot ($M=3.279$, $SD=1.734$) and anthropomorphic chatbot with persuasive elements ($M=2.504$, $SD=1.045$) (see Table 2).

To test whether our manipulation of the interface designs for the three different treatments was successful, we assessed users' perceived social persuasiveness. A test for variances homogeneity was not significant ($F(2, 117) = 13.467$; $p = .597$). Based on this result, we conducted a one-way ANOVA. The ANOVA was significant with $F(2,117) 13.467$; $p < .001$. The result of a Tukey HSD post hoc comparison revealed following significant differences between for control ($M=2.7$; $SD=1.951$) – anthropomorphic chatbot ($M=4.88$; $SD=1.977$) ($p < .001$), and anthropomorphic chatbot - anthropomorphic chatbot with persuasive elements ($M=3.08$; $SD=1.789$) ($p < .001$). We applied PLS (partial least squares) to evaluate the measurement model and estimate the structural model. As our analysis includes dialog support as a latent variable, we applied a structural equation approach. We used partial least squares (PLS) path modeling and employed SmartPLS 3.2.9. In the following paragraph, we first

inspect the measurement models and will then estimate and interpret the structural model.

Table 2. Descriptive Statistics

Dependent Variables	Treatments				
		All	Control	AC	ACwPE
Performance	Mean	4.942	3.150	7.375	4.300
	SD	4.387	3.519	5.309	2.893
Perceived Dialogue Support	Mean	3.486	2.450	5.150	2.858
	SD	2.042	1.693	1.743	1.571
Perceived Anthropomorphism	Mean	2.629	2.107	3.279	2.504
	SD	1.467	1.318	1.734	1.045
Perceived Persuasiveness (Manipulation Check)	Mean	3.55	2.7	4.88	3.08
	SD	1.903	1.951	1.977	1.789

SD = Standard deviation, AC = Anthropomorphic, ACwPE = AC with Persuasive Elements

5.2 Measurement Model and Structural Model

The measurement model includes manifest variables in terms of the experimentally manipulated variables, the number of completed tasks, and reflective constructs. From the experimental treatments, we derived four variables (see Table 3). The no chatbot variable (control treatment) was not included (reference group).

Table 3. Inter-Construct Correlations, CR, and AVE

(Latent) Variable	CR	AVE	1	2	3	4	5
1. Number of Completed Tasks	-	-	-				
2. Dialogue Support	.95	.86	.43	.93			
3. Anthropomorphism	.94	.69	.33	.53	.83		
4. Anthropomorphic Chatbot Design	-	-	.17	.14	.08	-	
5. Persuasive Chatbot Design	-	-	-.11	-.25	-.11	.58	-

CR = composite reliability, AVE = average variance extracted

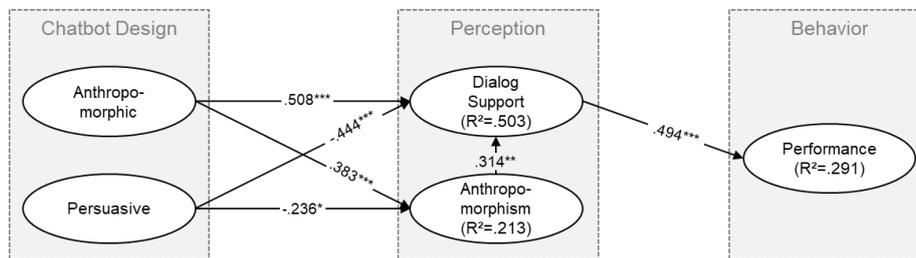


Figure 5. PLS Structural Model *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$

We then assessed the reflective measurement model of anthropomorphism and dialogue support for individual item reliability, convergent validity, and discriminant

validity. The model displays good measurement properties: all factor loadings are meaningful and significant, the composite reliability is above .7, the average variance extracted is above .5, and the Fornell–Larker criterion is satisfied. We then applied a bootstrap resampling procedure (with 4999 samples) to test the relationships. We favor the SEM for our research design with latent variables because it takes into account measurement errors or multidimensional structures of theoretical constructs [50]. The PLS estimator has advantages with respect to restrictive assumptions and is therefore widely used in experimental research [51], [52]. The different experimental conditions (no chatbot, anthropomorphically designed chatbot, persuasively and anthropomorphically designed chatbot) were dummy coded for our structural model, to compare the manipulations with a baseline condition (no chatbot). The structural model explains variances in Anthropomorphism ($R^2 = .213$, $f^2 = .156$), Dialog Support ($R^2 = .503$, $f^2 = .312$) and Performance (measured as number of completed tasks) ($R^2 = .291$). The results of the PLS estimation are illustrated in Figure 5.

Table 4. Results for Hypotheses

	Hypothesis	B	t	
1	a) An anthropomorphically designed chatbot yields a higher level of perceived dialog support than no chatbot.	.51***	6.11	s
	b) A persuasively and anthropomorphically designed chatbot yields a higher level of perceived dialog support than an anthropomorphically designed chatbot.	-.44***	5.25	<i>c</i>
2	a) An anthropomorphically designed chatbot yields a higher level of perceived anthropomorphism than no chatbot.	.38***	4.02	s
	b) A persuasively and anthropomorphically designed chatbot yields a higher level of perceived anthropomorphism than an anthropomorphically designed chatbot.	-.24*	2.49	<i>c</i>
3	Perceived anthropomorphism positively impacts perceived dialog support.	.31**	3.96	s
4	Perceived dialog support positively impacts the number of completed tasks.	.49***	6.41	s

s = supported, c = contradicted, ns = non-supported, B = path coefficient

In summary, we find support for hypotheses H1a, H2a, H3, and H4. We find contradicting results for H1b and H2b, namely the role of the persuasive design (see Table 4). Concerning, our control variables, we find significant effects for prior experience with chatbots on Anthropomorphism ($\beta = -.239$, $p < .05$). Moreover, we find a significant effect on Gender on Number of Completed Tasks ($\beta = -.181$, $p < .05$), with male participants completing fewer tasks.

6 Discussion

Our experiment aimed to explore the relationship between the persuasive and anthropomorphic design of conversational agents and performance. The results have implications for CA and persuasive system design. In this regard, we provide empirical

evidence for the relation of anthropomorphism and persuasive design in CAs. We found contradicting evidence for our hypotheses that persuasive cues (explicitly praise, suggestion, and rewards) lead to higher perceived anthropomorphism and dialogue support. These results can be explained from different perspectives.

6.1 Implications for Research

First, when looking at CA literature, Seeger et al. [15] state that simply applying more social cues and anthropomorphic design elements will not automatically lead to a higher level of perceived anthropomorphism. Selecting and combining them should be done with caution. In this context, Clark et al. [53] see the expectations of a user as decisive. Users are experienced with the interaction with humans and know the mistakes they make in an interaction. However, computers make errors that can rarely be found in humans. Therefore, these errors are unexpected. Regarding our CA design, the anthropomorphic chatbot was well perceived, leading to higher perceived anthropomorphism and dialog support. However, by adding the intended-to-be-persuasive elements to the design, the perception of the chatbot is vastly different from the other one. This observation indicates that users did not expect the added social cues.

Second, it could also be hypothesized that the persuasive chatbot appears to be disingenuous. A slider task does not require specific skills, qualifications, or knowledge [48]. Furthermore, unlike tasks in crowdsourcing, such as labeling pictures, performing a slider task has no trigger for enjoyment (task enjoyment leading to increased performance [54]), has no deeper meaning (perceived meaning is linked with satisfaction and performance [55]), and does not enable a user to contribute to a greater good (like voluntary work where the reward is intrinsic to the act of volunteering [56]). Hence, we would suggest that individuals perceive the high level of praise, combined with suggestions to keep going and receiving arbitrary point rewards, as disingenuous and not fitting the task.

Lastly, the negative perception of the persuasive chatbot might be explained by the cognitive fit theory [57]. The theory proposes that the fit between task and presentation of supporting information shapes the task performance. Our results indicate that an anthropomorphic CA provides a better information presentation in terms of dialog support, fitting the task at hand. This fit leads to higher performance. Thus, through the lens of the cognitive fit theory, the addition of persuasive elements appears to reduce the fit between task and task support.

In summary, our results can be embedded in the current discourse of CA design for task support. However, the significant negative change in the CA's perception by adding persuasive elements was unexpected. Thus, our results highlight a research opportunity to investigate the design of CAs for task support. Specifically, the framing and nature of a task appear to interact with the perception of a CA. CAs should meet expectations, appear genuine, and be adapted to the nature of the task. However, understanding how to design such a CA has yet to be addressed.

6.2 Implications for Practice

For practice, our result indicates that using a CA to frame and support tasks can be beneficial. To be specific, we would relate our results to the context of crowdworking. In crowdworking, crowd workers perform multiple tasks [58], which fits the experimental setup of this study. Our participants were inclined to complete more tasks than necessary. This indicates that adding the option to perform more tasks, accompanied by an anthropomorphic CA, can lead crowd workers to do more tasks. Furthermore, our study provides a blueprint regarding the design of such an anthropomorphic CA. Specifically, we would advise against adding persuasive messages or other design elements to an anthropomorphic CA that is intended to provide dialog support. Therefore, our results can be used to better design chatbots in the context of (crowdworking) tasks.

6.3 Limitations and Future Research

Our study is not free of limitations and offers different opportunities for future research. We conducted the online experiment in a rather controlled setting, with a set of specific tasks that every participant was asked to complete, and a single interaction with the conversational agent. Moreover, we did not compare the provided CA's with a CA without any social cues. Thus, we benefitted from control yet lacked realism in our research design [49]. Similarly, our results are limited by the selection and reimbursement of participants. In a real-world work environment, individuals are under the constant influence of expecting and receiving payment for work. For instance, crowd workers primarily perform tasks to be paid [58]. In our setting, participants did not receive a comparable form of pay. They were allowed to participate in a raffle for 10€ online shopping vouchers. Thus, it is safe to assume that participants were motivated by other factors, such as curiosity or escaping boredom.

7 Conclusion

In this study, we set out to explore the relation of persuasive and anthropomorphic CA design and performance (measured as the number of completed tasks). By means of a three-condition online experiment with two chatbots and 120 participants, we find empirical evidence for the positive influence an anthropomorphic CA has on an individual's perceived dialog support, mediated by the perceived anthropomorphism. However, a CA that displays the same anthropomorphic features and additionally provides persuasive messages, intended to provide further dialog support, is negatively perceived. This observation supports the proposition of Seeger et al. [15] that merely adding social cues and anthropomorphic characteristics to a CA is not always beneficial. In this context, our results indicate that a chatbot that provides dialog support (in our case praise, suggestions, and rewards) for simple tasks appears to be disingenuous. Therefore, our results indicate a potential for future research regarding the interaction of task and persuasive CA design. Our study makes three main contributions: First, we empirically demonstrate how the application of

anthropomorphic characteristics and persuasive messages can influence performance. Thereby, we add to the body of knowledge regarding the perception and influence anthropomorphic IS has on users. Second, we present CAs as a new type of persuasive IS that triggers social responses by users and offers new opportunities for interface and task design. Third, we bridge the gap between knowledge on persuasions and anthropomorphism of IS and the design of CA for dialog support.

8 Acknowledgements

We would like to thank Jonas Gehrke and Jessica Lühnen for their support during this research project.

9 References

- [1] I. Benbasat, "HCI Research: Future Challenges and Directions," *AIS Trans. Human-Computer Interact.*, vol. 2, no. 2, pp. 16–21, 2010.
- [2] B. J. Fogg, "Computers as persuasive social actors," in *Persuasive Technology*, San Francisco, USA: Morgan Kaufmann Publishers, 2003, pp. 89–120.
- [3] C.-M. Looock, T. Staake, and F. Thiesse, "Motivating Energy-Efficient Behavior with Green IS: An Investigation of Goal Setting and the Role of Defaults," *Manag. Inf. Syst. Q.*, vol. 37, no. 4, pp. 1313–1332, 2013.
- [4] M. S. Haque, M. Isomursu, M. Kangas, and T. Jämsä, "Measuring the influence of a persuasive application to promote physical activity," *CEUR Workshop Proc.*, vol. 2089, pp. 43–57, 2018.
- [5] P. Slattey, R. Vidgen, and P. Finnegan, "Persuasion: An analysis and common frame of reference for is research," *Commun. Assoc. Inf. Syst.*, vol. 46, pp. 30–69, 2020.
- [6] H. Oinas-Kukkonen and M. Harjumaa, "A Systematic Framework for Designing and Evaluating Persuasive Systems," *LNCS*, vol. 5033, pp. 164–176, 2008.
- [7] M. McTear, Z. Callejas, and D. Griol, *The Conversational Interface: Talking to Smart Devices*. Basel, Switzerland: Springer Publishing Company, 2016.
- [8] N. Wang, W. L. Johnson, R. E. Mayer, P. Rizzo, E. Shaw, and H. Collins, "The politeness effect: Pedagogical agents and learning outcomes," *Int. J. Hum. Comput. Stud.*, vol. 66, no. 2, pp. 98–112, 2008.
- [9] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. The Center for the Study of Language and Information Publications, 1996.
- [10] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive Systems Design: Key Issues, Process Model, and System Features," *Commun. Assoc. Inf. Syst.*, vol. 24, no. 1, p. 96, 2009.
- [11] S. Diederich, S. Lichtenberg, A. B. Brendel, and S. Trang, "Promoting sustainable mobility beliefs with persuasive and anthropomorphic design: Insights from an experiment with a conversational agent," Nov. 2019.

- [12] U. Gnewuch, S. Morana, C. Heckmann, and A. Maedche, "Designing Conversational Agents for Energy Feedback," in *Lecture Notes in Computer Science*, vol. 10844 LNCS, Springer Verlag, 2018, pp. 18–33.
- [13] M. J. Koeder, E. Tanaka, and H. Mitomo, "'Lootboxes' in digital games - A gamble with consumers in need of regulation? An evaluation based on learnings from Japan," *22nd Bienn. Conf. Int. Telecommun. Soc. "Beyond boundaries Challenges business, policy Soc.*, 2018.
- [14] T. Toscos, A. Faber, S. An, and M. P. Gandhi, "Chick Clique: Persuasive technology to motivate teenage girls to exercise," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1873–1878, 2006.
- [15] A. M. Seeger, J. Pfeiffer, and A. Heinzl, "Designing anthropomorphic conversational agents: Development and empirical evaluation of a design framework," in *ICIS*, 2018, pp. 1–17.
- [16] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," in *ACM CHI*, 1994, p. 204.
- [17] C. Nass and Y. Moon, "Machines and mindlessness: Social responses to computers," *J. Soc. Issues*, vol. 56, no. 1, pp. 81–103, 2000.
- [18] S. Langrial, T. Lehto, H. Oinas-Kukkonen, M. Harjumaa, and P. Karppinen, "Native mobile applications for personal wellbeing: A persuasive systems design evaluation," in *PACIS*, 2012, pp. 1–16.
- [19] S. Consolvo, K. Everitt, I. Smith, and J. A. Landay, "Design requirements for technologies that encourage physical activity," in *Conference on Human Factors in Computing Systems - Proceedings*, 2006, vol. 1, pp. 457–466.
- [20] J. Filippou, C. Cheong, and F. Cheong, "Modelling the impact of study behaviours on academic performance to inform the design of a persuasive system," *Inf. Manag.*, vol. 53, no. 7, pp. 892–903, 2016.
- [21] S. Lichtenberg and A. B. Brendel, "Arrr you a Pirate? Towards the Gamification Element 'Lootbox,'" *AMCIS (Forthcoming)*, 2020.
- [22] S. Lichtenberg, T. Lembcke, M. Brenig, A. B. Brendel, and S. Trang, "Can Gamification lead to Increase Paid Crowdworkers Output?," in *15. Internationale Tagung Wirtschaftsinformatik*, 2020, no. December 2019.
- [23] T. Lehto, H. Oinas-Kukkonen, and F. Drozd, "Factors affecting perceived persuasiveness of a behavior change support system," in *ICIS*, 2012, vol. 3, pp. 1926–1939.
- [24] S. Diederich, A. B. Brendel, and L. M. Kolbe, "On Conversational Agents in Information Systems Research: Analyzing the Past to Guide Future Work," *Proc. Int. Conf. Wirtschaftsinformatik*, pp. 1550–1564, 2019.
- [25] Oracle, "Can Virtual Experiences Replace Reality? The future role for humans in delivering customer experience," p. 19, 2016.
- [26] N. V. Wunderlich and S. Paluch, "A Nice and Friendly Chat With a Bot: User Perceptions of AI-based Service Agents," in *ICIS*, 2017, no. 1, pp. 1–11.
- [27] R. M. Stock and M. Merkle, "Can Humanoid Service Robots Perform Better Than Service Employees? A Comparison of Innovative Behavior Cues," 2018.
- [28] J. Feine, U. Gnewuch, S. Morana, and A. Maedche, "A Taxonomy of Social Cues for Conversational Agents," *Int. J. Hum. Comput. Stud.*, vol. 132, pp.

- 138–161, Dec. 2019.
- [29] T. Verhagen, J. van Nes, F. Feldberg, and W. van Dolen, “Virtual customer service agents: Using social presence and personalization to shape online service encounters,” *J. Comput. Commun.*, vol. 19, no. 3, pp. 529–545, 2014.
 - [30] B. J. Fogg and C. Nass, “How users reciprocate to computers,” in *ACM CHI*, 1997, no. March, p. 331.
 - [31] L. Gong, “How social is social responses to computers? The function of the degree of anthropomorphism in computer representations,” *Comput. Human Behav.*, vol. 24, no. 4, pp. 1494–1509, 2008.
 - [32] U. Gnewuch, S. Morana, M. T. P. Adam, and A. Maedche, “Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction,” in *ECIS*, 2018, pp. 1–17.
 - [33] T. Araujo, “Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions,” *Comput. Human Behav.*, vol. 85, pp. 183–189, 2018.
 - [34] M. D. Hanus and J. Fox, “Persuasive avatars: The effects of customizing a virtual salespersons appearance on brand liking and purchase intentions,” *Int. J. Hum. Comput. Stud.*, vol. 84, pp. 33–40, 2015.
 - [35] J. Sebastian and D. Richards, “Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents,” *Comput. Human Behav.*, vol. 73, pp. 479–488, 2017.
 - [36] Y. J. Sah and W. Peng, “Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website,” *Comput. Human Behav.*, vol. 45, pp. 392–401, 2015.
 - [37] R. M. Schuetzler, J. S. Giboney, G. M. Grimes, and J. F. Nunamaker, “The Influence of Conversational Agents on Socially Desirable Responding,” in *HICSS*, 2018, vol. 9, pp. 283–292.
 - [38] P. Ekman and W. V. Friesen, “The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding,” *Semiotica*, vol. 1, no. 1, pp. 49–98, 1969.
 - [39] R. E. Mayer, W. L. Johnson, E. Shaw, and S. Sandhu, “Constructing computer-based tutors that are socially sensitive: Politeness in educational software,” *Int. J. Hum. Comput. Stud.*, vol. 64, no. 1, pp. 36–42, 2006.
 - [40] N. Shevchuk and H. Oinas-Kukkonen, “Exploring Green Information Systems and Technologies as Persuasive Systems: A Systematic Review of Applications in Published Research,” in *ICIS*, 2016, pp. 1–11.
 - [41] M. El Kamali, L. Angelini, M. Caon, G. Andreoni, O. A. Khaled, and E. Mugellini, “Towards the Nestore e-Coach: A tangible and embodied conversational agent for older adults,” in *UbiComp/ISWC 2018*, Oct. 2018, pp. 1656–1663.
 - [42] J. Hamari, “Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service,” *Electron. Commer. Res. Appl.*, vol. 12, no. 4, pp. 236–245, 2013.
 - [43] K. Xu and M. Lombard, “Persuasive computing: Feeling peer pressure from multiple computer agents,” *Comput. Human Behav.*, vol. 74, pp. 152–162, Sep.

- 2017.
- [44] V. J. Harjunen, M. Spapé, I. Ahmed, G. Jacucci, and N. Ravaja, “Persuaded by the machine: The effect of virtual nonverbal cues and individual differences on compliance in economic bargaining,” *Comput. Human Behav.*, vol. 87, pp. 384–394, Oct. 2018.
 - [45] R. F. Adler, F. Iacobelli, and Y. Gutstein, “Are you convinced? A Wizard of Oz study to test emotional vs. Rational persuasion strategies in dialogues,” *Comput. Human Behav.*, vol. 57, pp. 75–81, Apr. 2016.
 - [46] M. C. Boudreau, D. Gefen, and D. W. Straub, “Validation in information systems research: A state-of-the-art assessment,” *MIS Q. Manag. Inf. Syst.*, vol. 25, no. 1, pp. 1–16, 2001.
 - [47] E. Erdfelder, F. FAul, A. Buchner, and A. G. Lang, “Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses,” *Behav. Res. Methods*, vol. 41, no. 4, pp. 1149–1160, 2009.
 - [48] E. Lezzi, P. Fleming, and D. J. Zizzo, “Does it Matter Which Effort Task You Use? A Comparison of Four Effort Tasks When Agents Compete for a Prize,” *SSRN Electron. J.*, 2015.
 - [49] A. R. Dennis and J. S. Valacich, “Conducting Experimental Research in Information Systems,” *Commun. Assoc. Inf. Syst.*, vol. 7, no. 5, pp. 1–41, 2001.
 - [50] R. P. Bagozzi and Y. Yi, “On the Use of Structural Equation Models in Experimental Designs,” *J. Mark. Res.*, vol. 26, no. 3, p. 271, Aug. 1989.
 - [51] P. W. Fombelle, S. A. Bone, and K. N. Lemon, “Responding to the 98%: face-enhancing strategies for dealing with rejected customer ideas,” *J. Acad. Mark. Sci.*, vol. 44, no. 6, pp. 685–706, Nov. 2016.
 - [52] M. Trenz, D. Veit, and C.-W. Tan, “Disentangling the Impact of Omnichannel Integration on Consumer Behavior in Integrated Sales Channels,” *Manag. Inf. Syst. Q.*, vol. 44, no. 3, Sep. 2020.
 - [53] L. Clark *et al.*, “What makes a good conversation? Challenges in designing truly conversational agents,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2019, vol. 12.
 - [54] R. M. Puca and H. D. Schmalt, “Task enjoyment: A mediator between achievement motives and performance,” *Motiv. Emot.*, vol. 23, no. 1, pp. 15–29, 1999.
 - [55] A. Wrzesniewski, J. E. Dutton, and G. Debebe, “Interpersonal Sensemaking.pdf,” *Research in Organizational Behavior*, vol. 25. pp. 93–135, 2003.
 - [56] H. Bussell and D. Forbes, “Understanding the volunteer market: the what, where, who and why of volunteering,” *Int. J. Nonprofit Volunt. Sect. Mark.*, vol. 7, no. 3, pp. 244–257, 2002.
 - [57] R. Agarwal, A. P. Sinha, and M. Tanniru, “Cognitive Fit in Requirements Modeling: A Study of Object and Process Methodologies,” *J. Manag. Inf. Syst.*, vol. 13, no. 2, pp. 137–162, 1996.
 - [58] D. Durward, I. Blohm, and J. M. Leimeister, “The Nature of Crowd Work and its Effects on Individuals’ Work Perception,” *J. Manag. Inf. Syst.*, vol. 37, no. 1, pp. 66–95, 2020.