

Design Features of Embodied Conversational Agents in eHealth: a Literature Review



Silke ter Stal^{a,b,1,*}, Lean Leonie Kramer^{c,d,1}, Monique Tabak^{a,b,2}, Harm op den Akker^{a,b,2}, Hermie Hermens^{a,b,3}

^a eHealth Group, Roessingh Research and Development, Enschede, the Netherlands

^b Biomedical Systems and Signals Group, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands

^c Strategic Communication Group, Wageningen University and Research, Wageningen, The Netherlands

^d Consumption and Healthy Lifestyles Group, Wageningen University and Research, Wageningen, The Netherlands

ARTICLE INFO

Keywords:

Embodied Conversational Agent
eHealth
design feature
review

ABSTRACT

Embodied conversational agents (ECAs) are gaining interest to elicit user engagement and stimulate actual use of eHealth applications. In this literature review, we identify the researched design features for ECAs in eHealth, the outcome variables that were used to measure the effect of these design features and what the found effects for each variable were. Searches were performed in Scopus, ACM Digital Library, PsychINFO, Pubmed and IEEE Xplore Digital Library, resulting in 1284 identified articles of which 33 articles were included. The agents speech and/or textual output and its facial and gaze expressions were the most common design features. Little research was performed on the agent's looks. The measured effect of these design features was often on the perception of the agent's and user's characteristics, relation with the agent, system usage, intention to use, usability and behaviour change. Results show that emotion and relational behaviour seem to positively affect the perception of the agents characteristics and that relational behaviour also seems to positively affect the relation with the agent, usability and intention to use. However, these design features do not necessarily lead to behaviour change. This review showed that consensus on design features of ECAs in eHealth is far from established. Follow-up research should include more research on the effects of all design features, especially research on the effects in a long-term, daily life setting, and replication of studies on the effects of design features performed in other contexts than eHealth.

1. Introduction

To relieve the burden on the healthcare sector caused by the ageing society, the use of eHealth applications is being widely investigated. These applications can be used in establishing a user's behaviour change in daily life either under the supervision of a healthcare professional, or in stand-alone use to promote self-management. Although they seem promising, many eHealth applications face the problem of actual use rapidly decreasing after several weeks (Nijland, 2011). Often, existing eHealth applications provide advice in the form of plain text or via a text-based question-answer module (Kaptein et al., 2012). Face-to-face interaction remains one of the best ways to communicate health information; it incorporates grounding – dynamically assessing

the other persons level of understanding and repeating or elaborating on information when necessary (Clark and Brennan, 1991). In addition, face-to-face interaction elicits trust, better communication and satisfaction via both verbal and non-verbal behaviour (Bickmore et al., 2009b).

Face-to-face interaction seems to be a possibility to elicit user engagement and stimulate actual use of eHealth applications. Therefore, the use of embodied conversational agents (ECAs) is gaining interest as an alternative means. ECAs are *more or less autonomous and intelligent software entities with an embodiment used to communicate with the user* (Ruttkay et al., 2004). By interacting with the user, ECAs can build trust and rapport, leading to companionship and long-term, continual use (Vardoulakis et al., 2012).

* Corresponding author.

E-mail address: s.terstal@rrd.nl (S. ter Stal).

¹ MSc

² PhD

³ Prof. dr. ir.

Table 1

Terms used for the database searches. For databases that do not allow the use of the asterisk (*), the asterisks were removed.

Term [Embodied Conversational Agent]		Term [eHealth]
"virtual agent*" OR "conversational agent*" OR "virtual *coach*" OR "digital coach*" OR "counsel* agent*" OR "virtual counsel*" OR "virtual advisor*" OR "motivational agent*" OR "virtual human*" OR "animated character*" OR "virtual character*" OR "relational agent*" OR "social agent*" OR "interface agent*" OR "interface character"	AND	e-health OR ehealth OR tele-medicine OR telemedicine OR tele-health OR telehealth OR m-health OR mhealth OR health* OR wellbeing OR e-coaching OR ecoaching OR medic*

1.1. ECAs in eHealth: a Lack of Design Guidelines

Although research indicates that incorporating ECAs into eHealth applications could elicit user engagement, little is known about how these agents should be designed in order to accomplish this engagement. Some research on the agent's design has been performed, but no design guidelines exist. A taxonomy of the different *design features* of ECAs can be essential to establish a common ground for developing design guidelines. Ruttkay et al. (Ruttkay et al., 2004) created a taxonomy of relevant design and evaluation aspects of ECAs. They distinguish the agent's embodiment (its looks, speech and/or textual output, hand and body gestures and facial and gaze expressions), mental capacities (its social role, personality, user model, natural language generator and dialogue manager) and the application interface (including background knowledge processing). In addition, Straßmann and Krämer (Straßmann and Krämer, 2017) identify design features related to the agent's appearance. They categorise the variables: embodiment vs no embodiment, species, realism, 2D vs 3D and feature specification (socio-demographic and styling).

Despite the attempts to create a taxonomy of design features, little is known about how these features should actually be designed. Some agent design guidelines exist, such as the design guidelines for pedagogical agents by Veletsianos et al. (Veletsianos et al., 2009), but these guidelines do not focus on eHealth. Many studies on agent design features with respect to eHealth explore a single design feature (such as an agent's culture background (Zhou et al., 2017) and body shape (van Vugt et al., 2006)). Therefore, we conclude that no design guidelines for ECAs in eHealth exist.

A literature review of research on design features for ECAs in eHealth can, therefore, be a valuable input for the development of these guidelines. Such a literature review could provide insight into how often particular design features have been researched and could draw general conclusions on the effects of particular design features flowing from results of multiple studies. Several literature reviews on conversational agents in eHealth have been performed. However, they either focus on conversational agents in general, not on ECAs specifically (Laranjo et al., 2018; Rist et al., 2004), are not up-to-date (Rist et al., 2004), focus on a broader context than health (Scholten et al., 2017) or focus on a subarea of health, such as clinical psychology (Kramer et al., 2019; Provoost et al., 2017; Rist et al., 2004). In addition, all of the reviews focus on technological and clinical possibilities. Although they sometimes include a description of the ECA designs used, they do not present effects of particular design features. Thus, a structured literature review of the available studies on particular design features, including a general conclusion with respect to the researched effect of the design features, is missing.

1.2. Research Objectives

The main goal of this literature review is to identify the researched design features for ECAs in eHealth. The sub goals of this research are to identify the outcome variables used to measure the effect of these design features and to identify what the found effects for each variable were.

2. Method

2.1. Search Strategy

Searches were performed in November 2018 in the electronic databases of Scopus, ACM Digital Library, PsychINFO, PubMed and IEEE Xplore Digital Library, as discussed and agreed upon by three researchers: the first, third and fourth author. The searches were restricted to queries containing terms related to (1) *embodied conversational agent* and (2) *eHealth*. The list of search terms was composed after several iterations and refinement by the first, third and fourth author. The final list of search terms can be seen in Table 1.

The searches were performed on titles and abstracts and were not restricted on publication date. For databases that allowed to, Scopus and Pubmed, the language was limited to English and Dutch. In addition, we limited the searches on Scopus to the subject areas Computer Science, Medicine, Mathematics, Social Sciences, Engineering, Psychology, Health Professions, Neuroscience, Nursing, Arts and Humanities and Decision Sciences and the document type Conference Paper, Article, Book Chapter and Book. Again, these limitations were discussed and agreed upon by three researchers: the first, third and fourth author. The final database searches were performed by one researcher (StS).

2.2. Screening Strategy

From the articles identified by the database searches, the duplicates were removed by the first author. Then, two researchers, the first and second author, performed the title, abstract and full text screening independently. The inclusion and exclusion criteria used for the screenings were discussed and agreed upon by the first, third and fourth author and can be seen in Table 2 and Table 3. The taxonomy of design features used for exclusion criterion E4 was created by combining the categories identified by Ruttkay et al. and Straßmann and Krämer (see Table 3). After each screening, the researchers discussed disagreements until they reached consensus. For the full-text screening, a third researcher, the third author, screened the texts for which the other two researchers had difficulties in reaching consensus. Finally, forward-snowballing was used to screen the references in the included articles using the same technique as used for the screening of the database searches, consisting of a title screening, abstract screening and full-text screening. Duplicates and articles that were already selected for the

Table 2

Inclusion criteria used for the article screenings.

Inclusion Criteria
I1 - The article is written in English or Dutch
I2 - The article is a journal article, conference paper or book (chapter)

Table 3
Exclusion criteria used for the article screenings.

Exclusion Criteria	Explanation
E1 – The article does not report on primary data	The article is a review article
E2 – The virtual agent is not an embodied conversational agent	Embodied conversational agents are more or less autonomous and intelligent software entities with and embodiment used to communicate with the user (Ruttkay et al., 2004)
E3 – The virtual agent is not applied in a health context	Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity (WHO, 1946)
E4 – The article does not research a design feature of the virtual agent	Design features: <ul style="list-style-type: none"> • Looks (Straßmann & Krämer, 2017): <ul style="list-style-type: none"> • Species (e.g. human, animal, robots, objects, and mystical creatures) • Realism (e.g. stylization, resolution and detailedness) • 2D/3D • Feature specification: Socio-demographic (e.g. gender, ethnicity, race) and Styling (hair, make-up) • Speech and/or textual output (Ruttkay et al., 2004) • Hand and body gestures (Ruttkay et al., 2004) • Facial and gaze expressions (Ruttkay et al., 2004)
E5 – The article does not provide any outcomes on the effect of or opinions of users on a design feature of the virtual agent	
E6 – There is no full-text available	

review through the screening of the database searches, were removed in a pre-processing stage.

2.3. Article Reviews and Synthesis

Two review tables were created. The first table, [Table 5](#), lists general information about the articles found: the goal of the application in which the ECA was implemented (either in the context of alcohol consumption, mental health, nutrition, physical activity, medical treatment or other) and characteristics of the participants in the research (the age group: adults, children or elderly; education: low, at least some college, students and university; and cultural background: Asian, African American, Caucasian and Hispanic). In addition, the evaluation of each study was classified as one of the four evaluation stages of DeChant, according to the renewed framework for the evaluation of telemedicine by [Jansen Kosterink et al. \(Jansen Kosterink et al., 2016\)](#). Evaluations were classified as either being in:

- Stage I: technical efficacy – focus on the feasibility and usability of the technology.
- Stage II: specific system objectives – gaining an initial idea about the potential added value for clinical practice and possible working mechanism.
- Stage III: system analysis – technology evaluated in the way they will be implemented in daily clinical practice.
- Stage IV: external validity – elaboration of the adoption as addressed in stage III.

Furthermore, each study was classified as either experimental (meaning the researcher allocates subjects to an intervention or exposure group), observational analytic (the researcher simply measures the exposure or treatments of the groups) or as a survey or qualitative study.

The structure of the second review table, [Table 6](#), was agreed upon by two researchers (StS and MT). For each article, the table presents the category of the design feature(s) researched, the design feature(s)

researched and the outcome variable(s). The design feature categories were the same categories as used in exclusion criterion E4 (see [Table 3](#)). The categories of the outcome variables were designed retrospectively by thematic analysis of all outcome variables found in the articles. In addition, the table displays the method and the results of the research with respect to the design feature. The articles are grouped on design feature and sorted alphabetically within this category.

3. Results

From the 1284 articles identified by the database searches, 23 articles were included in the review. In addition, 10 articles were included via the snowballing method, resulting in **33 articles included in the review**. [Figure 1](#) shows the flow diagram of the database searches and article screenings.

[Table 5](#) lists general information about the articles found. The included studies were published between 2001 and 2018. Most of the ECAs were developed in the context of *physical activity* (thirteen ECAs ([Bickmore et al., 2005a; 2009a; 2010; Bickmore and Picard, 2004; 2005; Forlizzi et al., 2007; Frost et al., 2012; Nguyen and Masthoff, 2007; Olafsson et al., 2017; Schmeil and Suggs, 2014; Wissen et al., 2016; Yin et al., 2010; Zhou et al., 2017](#))), *medical treatment* (eight ECAs ([Forlizzi et al., 2007; Parmar et al., 2018; Ring et al., 2014; Robertson et al., 2015; Silverman et al., 2001; Skalski et al., 2007; Wissen et al., 2016; Zhou et al., 2014](#))), *mental health* (six ECAs ([Alsharbi and Richards, 2017; Bickmore and Schulman, 2007; Grillon and Thalmann, 2008; Kang and Gratch, 2011; Nguyen and Masthoff, 2009; Tielman et al., 2017](#))) and *nutrition* (four ECAs ([Creed and Beale, 2012; Creed et al., 2015; Olafsson et al., 2017; Schmeil and Suggs, 2014](#))). Just a few articles describe ECAs in the context of *alcohol consumption* (3 articles ([Amini et al., 2014; 2013; Lisetti et al., 2013](#))) or *other topics* (three articles ([Bickmore and Ring, 2010; Malhotra et al., 2016; van Vugt et al., 2006](#))). The amount of participants differed from 11 to 764 (M = 91, SD = 147). Most studies included *both male and female* participants. Three studies focused on children ([Alsharbi and Richards, 2017; Frost et al., 2012; Zhou et al., 2017](#)), two on elderly ([Malhotra et al., 2016;](#)

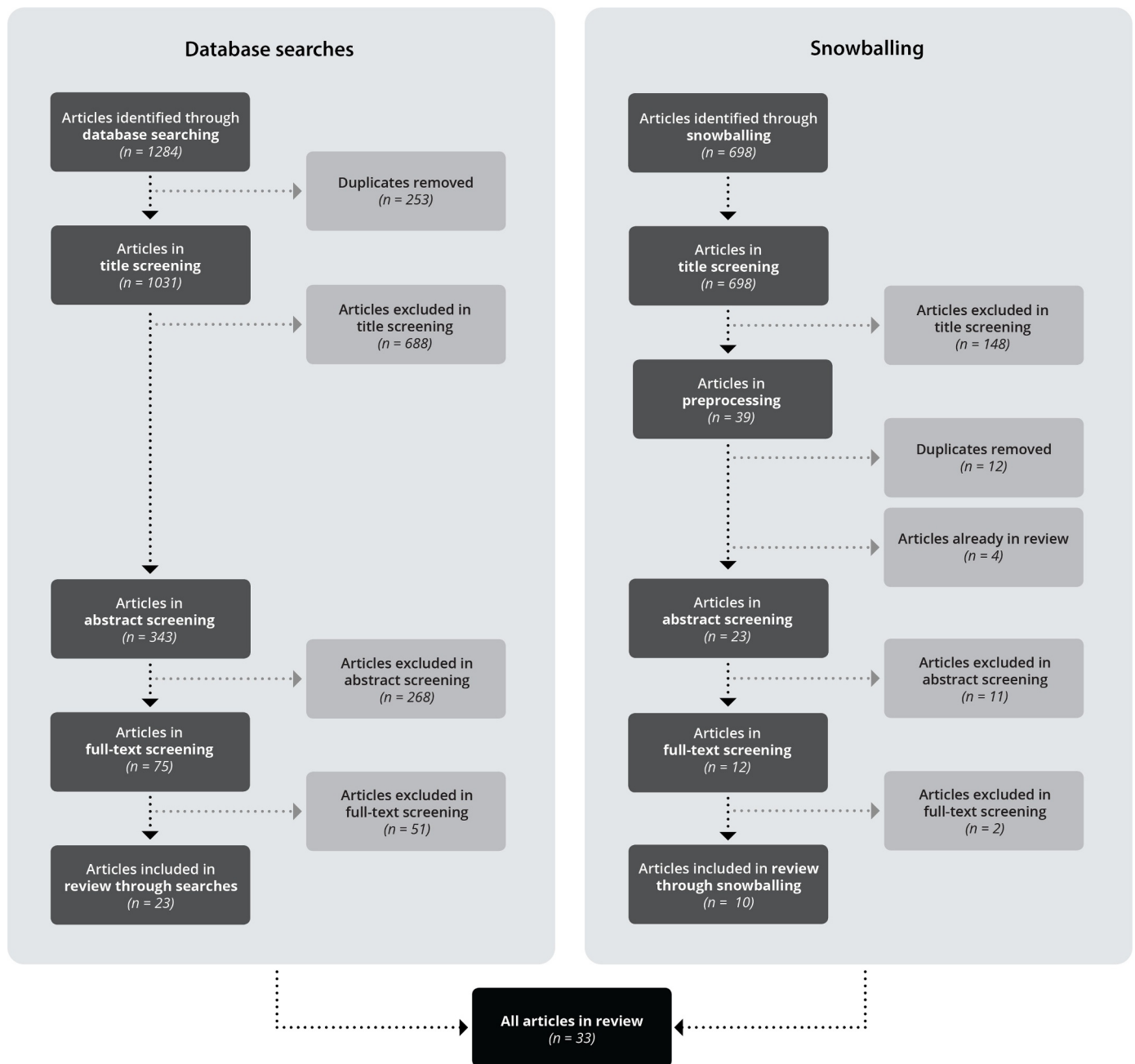


Fig. 1. Flow diagram of the database searches and article screenings.

Wissen et al., 2016) and the rest on adults. Of the articles that reported on the participants' education, most participants were students (Amini et al., 2013; Bickmore and Schulman, 2007; Bickmore and Picard, 2004; Creed et al., 2015; Lisetti et al., 2013; Nguyen and Masthoff, 2007; 2009; Olafsson et al., 2017; Skalski et al., 2007; Tielman et al., 2017; van Vugt et al., 2006), had a university degree (Creed et al., 2015; Nguyen and Masthoff, 2007; 2009; Tielman et al., 2017; Wissen et al., 2016; Zhou et al., 2017) or had at least some college (Bickmore and Ring, 2010; Bickmore et al., 2009a; 2010; Robertson et al., 2015; Silverman et al., 2001; Wissen et al., 2016; Zhou et al., 2017). Just one study particularly focused on lower-educates (Robertson et al., 2015). Of the articles that reported on the participants' cultural background, participants were *Caucasian* (thirteen articles (Alsharbi and Richards, 2017; Amini et al., 2013; Bickmore et al., 2009a; 2010; Creed et al., 2015; Frost et al., 2012; Olafsson et al., 2017; Robertson et al., 2015; Schmeil and Suggs, 2014; Tielman et al., 2017; Wissen et al., 2016; Yin

et al., 2010; Zhou et al., 2014)), *Afro American* (seven articles (Amini et al., 2013; Bickmore and Ring, 2010; Bickmore et al., 2009a; 2010; Olafsson et al., 2017; Robertson et al., 2015; Zhou et al., 2014)), *Hispanic* (three articles (Amini et al., 2013; Yin et al., 2010; Zhou et al., 2014)) and *Asian* (three articles (Amini et al., 2013; Olafsson et al., 2017; Zhou et al., 2017)).

Of the evaluations performed, no evaluation was in stage IV, technical efficacy. Just one article reports on stage III, system analysis (Zhou et al., 2014). Some articles performed evaluations in stage II, specific system objectives (Bickmore et al., 2005a; 2009a; 2010; Bickmore and Picard, 2004; 2005; Creed and Beale, 2012; Creed et al., 2015; Nguyen and Masthoff, 2009; Schmeil and Suggs, 2014; Skalski et al., 2007; Tielman et al., 2017; Yin et al., 2010; Zhou et al., 2017). However, the majority of the articles report on evaluations in the stage I, technical efficacy. In addition, no article described an observational analytic study and few articles describe qualitative studies (two articles (Nguyen and Masthoff,

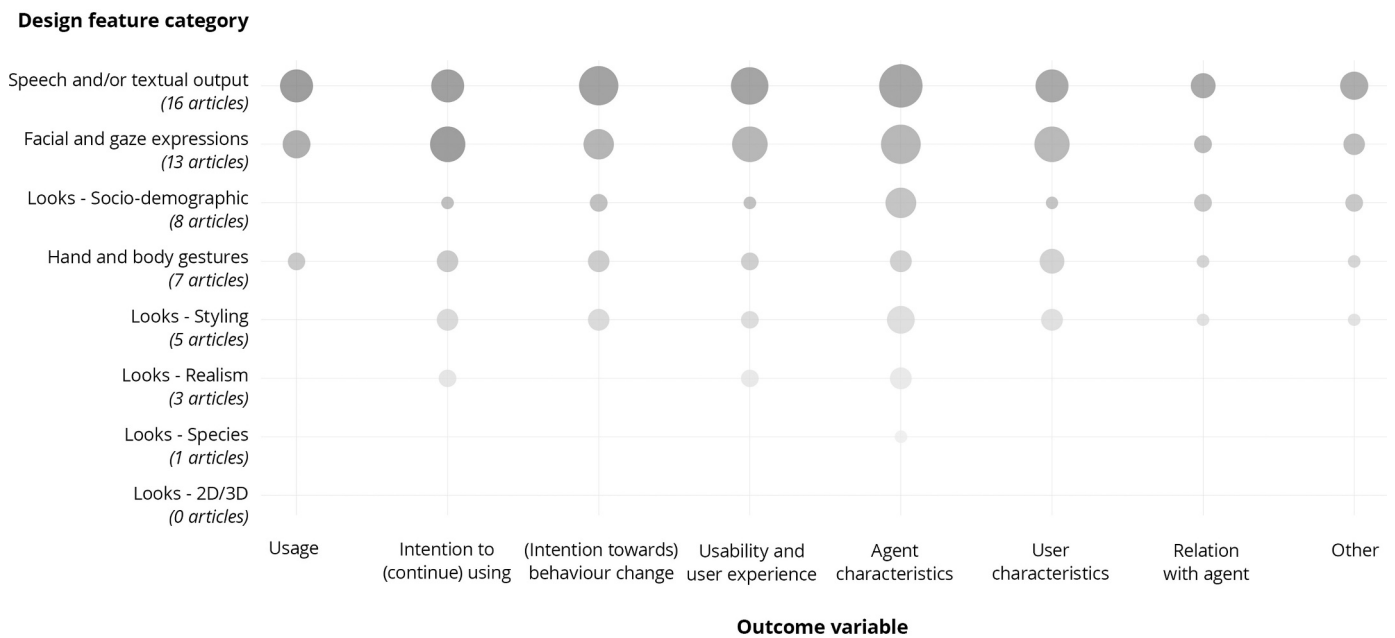


Fig. 2. Frequency of design features and outcome variables in the articles found. The width of the bubble corresponds to the number of articles that research a particular outcome variable for a particular design feature category.

2007; Robertson et al., 2015)) and survey studies (four articles (Alsharbi and Richards, 2017; Forlizzi et al., 2007; Nguyen and Masthoff, 2007; Parmar et al., 2018)). The majority of the studies performed were experimental studies; they compared multiple variants of a particular design feature.

3.1. The Design Features and Outcome Variables Researched

Table 6 provides information about the design features researched, corresponding outcome variables and results for each article in the review. All articles were grouped on design feature category. In addition, Figure 2 provides an overview of the frequencies of the design feature categories and outcome variables identified in articles included in the review.

Some articles research design features in multiple categories. Most of the research is performed on the categories speech and/or textual output and facial and gaze expressions. The categories species and 2D/3D are researched the least.

The thematic analysis of the outcome variables resulted in the following categories: usage, intention to (continue) using, (intention towards) behaviour change, usability and user experience, agent characteristics (e.g. demographics, personality, styling), relation with agent, user characteristics and other. The majority of the articles provide outcomes regarding the users' perception of the agent characteristics (Alsharbi and Richards, 2017; Amini et al., 2014; 2013; Bickmore and Ring, 2010; Bickmore and Schulman, 2007; Bickmore et al., 2009a; 2010; 2005b; Bickmore and Picard, 2005; Creed and Beale, 2012; Forlizzi et al., 2007; Grillon and Thalmann, 2008; Lisetti et al., 2013; Malhotra et al., 2016; Nguyen and Masthoff, 2007; 2009; Olafsson et al., 2017; Parmar et al., 2018; Ring et al., 2014; Robertson et al., 2015; Silverman et al., 2001; Skalski et al., 2007; Tielman et al., 2017; van Vugt et al., 2006; Wissen et al., 2016; Yin et al., 2010; Zhou et al., 2014; 2017). In addition, many articles report on the users' perception of the relation with the agent (Alsharbi and Richards, 2017; Amini et al., 2014; 2013; Bickmore et al., 2005b; Bickmore and Picard, 2004; 2005; Creed et al., 2015; Kang and Gratch, 2011; Lisetti et al., 2013; Olafsson et al., 2017; Parmar et al., 2018; Skalski et al., 2007; van Vugt et al., 2006; Zhou et al., 2014), usability and user experience (Amini et al., 2014; 2013; Bickmore and Ring, 2010; Bickmore and Schulman, 2007; Bickmore et al., 2009a;

2010; 2005b; Lisetti et al., 2013; Nguyen and Masthoff, 2009; Olafsson et al., 2017; Ring et al., 2014; Silverman et al., 2001; Tielman et al., 2017; Wissen et al., 2016; Zhou et al., 2014; 2017), intention to use (Amini et al., 2014; 2013; Bickmore and Schulman, 2007; Bickmore et al., 2010; 2005b; Bickmore and Picard, 2004; 2005; Creed and Beale, 2012; Lisetti et al., 2013; Olafsson et al., 2017; Parmar et al., 2018; Ring et al., 2014; Schmeil and Suggs, 2014; van Vugt et al., 2006; Wissen et al., 2016; Zhou et al., 2014) and system usage (Bickmore et al., 2009a; 2010; 2005b; Bickmore and Picard, 2005; Creed and Beale, 2012; Creed et al., 2015; Olafsson et al., 2017). Also, many articles already provide information about the effect of the design of a particular feature on users' (intention towards) behaviour change (Bickmore and Schulman, 2007; Bickmore et al., 2010; 2005b; Bickmore and Picard, 2005; Creed et al., 2015; Kang and Gratch, 2011; Nguyen and Masthoff, 2009; Parmar et al., 2018; Schmeil and Suggs, 2014; Silverman et al., 2001; Skalski et al., 2007; Tielman et al., 2017; Yin et al., 2010; Zhou et al., 2017). Just a few articles provide results related to the users' perception of their own characteristics (Creed et al., 2015; Frost et al., 2012; Tielman et al., 2017; van Vugt et al., 2006; Yin et al., 2010; Zhou et al., 2014) or report on other outcome variables (Bickmore and Ring, 2010; Bickmore and Picard, 2005; Creed and Beale, 2012; Nguyen and Masthoff, 2007; Olafsson et al., 2017; Tielman et al., 2017; Zhou et al., 2014).

Looking at the relation between the design feature categories and outcome variables specifically, we see that for realism, species and 2D/3D just a few outcome variables are researched, whereas for the other categories, almost all outcome variables have been researched.

In the remainder of this section, the research and outcomes are grouped by the design features categories. We start with describing research related to speech and/or textual output, facial and gaze expressions and hand and body gestures, followed by research on the agent's looks.

3.1.1. Speech and/or Textual Output, Facial and Gaze Expressions and Hand and Body Gestures

Table 4 provides a summary of the effects found for the different outcome variables with respect to design features in the categories speech and/or textual output, facial and gaze expressions and hand and body gestures.

Table 4

Summary of the effects of the design features on the outcome variables, either a positive effect (+), negative effect (-), no effect (0) or an effect that depends on the context (~). For every row, symbols having the same number in superscript are researched within the same study.

	Design feature	Usage	Intention to (continue) using	(Intention towards) behaviour change	Usability and user experience	Agent characteristics	Relation with agent	User characteristics	Preference
Speech/Text, Output, Facial & Gaze Exp., Hand & Body Gest.	Emotion (vs. no)		+ ²	- ³ 0 ³ + ¹	+ ¹	+ ¹ + ² + ³			
	Relational, empathic behaviour (vs. no)		+ ¹ + ² + ³ + ⁴ + ⁵ + ⁶	0 ³ 0 ⁴ 0 ⁵ 0 ⁷ + ⁸	+ ¹ + ² + ⁶ + ⁷ + ⁸	+ ¹ + ² + ³ + ⁴ + ⁵ + ⁶ + ⁷	+ ¹ + ² + ³ + ⁴ + ⁵ + ⁶		
	Providing personal information (vs. no)	+ ²		+ ¹	+ ²		+ ¹		
	Variable behaviour (vs. no)	+ ¹	+ ¹	- ¹ ~ ²					
	User control prosody & facial expressions (vs. no)				+ ¹				
	Interactivity (vs. no)							+ ¹	
	Rap (vs. no)				0 ¹ + ¹	- ¹ 0 ¹	+ ¹		
	Agent message: in text (vs. verbally)	+ ¹		+ ¹					
	Linguistic tailoring (vs. no)			0 ¹					
Looks	Rendering style: human-like (vs. cartoon)		0 ⁴			- ¹ - ² + ¹ + ² + ³ + ⁴			
	Clothing: professional (vs. casual)		+ ¹			+ ¹	+ ¹		
	Body shape: slim (vs. fat)		- ⁴	- ⁴ + ¹ + ² + ³		- ⁴	- ⁴		
	Gender: female (vs. male)								+ ¹ ~ ²
	Age: young (vs. old)								+ ¹ ~ ²
	Cultural tailoring (vs. no)			- ² - ⁴		+ ¹ + ²	- ³		
	Role: friend (vs. professional)								+ ¹ ~ ²

First, some articles provide research on an agent's emotion. Compared to agents not showing emotion, agents showing emotion are rated higher on several characteristics, such as likeability and believability (Creed and Beale, 2012; Creed et al., 2015; Silverman et al., 2001) and resulted in higher usability (Silverman et al., 2001) and intention to use (Creed and Beale, 2012). However, no clear consensus exist for emotional agents triggering behaviour change; one study found that users interacting with an emotional agent showed a larger behaviour change than users interacting with a non-emotional agent (Silverman et al., 2001). Another study found the opposite: users interacting with a non-emotional agent showed a larger behaviour change than users interacting with an emotional agent. But, on other behaviour variables, they did not find any differences (Creed et al., 2015). It should be noted that the two studies offered different application goals: change in awareness on heart attack scenarios and change in food intake.

Second, some articles provide research on an agent's relational, empathic behaviour. First, relational agents are liked more: they score higher on characteristics, such as likeability, perceived caring, trustworthiness and enjoyment (Amini et al., 2014; 2013; Bickmore et al., 2005a; Bickmore and Picard, 2004; 2005; Lisetti et al., 2013; Nguyen and Masthoff, 2009). In addition, relational behaviour positively affects the users' relation with the agent (Amini et al., 2014; 2013; Bickmore et al., 2005a; Bickmore and Picard, 2004; 2005; Lisetti et al., 2013). Lastly, the use of relational agents leads to higher usability (Amini et al., 2014; 2013; Bickmore and Schulman, 2007; Lisetti et al., 2013; Nguyen and Masthoff, 2009) and intention to use (Amini et al., 2014; 2013; Bickmore et al., 2005a; Bickmore and Picard, 2004; 2005; Lisetti et al., 2013). However, with respect to behaviour change, literature presents mixed results; some studies did not find any effect (Bickmore et al., 2005a; Bickmore and Picard, 2004; 2005; Nguyen and Masthoff, 2009), whereas just one article provided some positive results related to an agents relational behaviour (Bickmore and Schulman, 2007).

Though, the studies researched applications with different goals. Bickmore and Schulman, finding a positive effect of relational behaviour, tested the effect of relation behaviour on mood, whereas the majority of the other studies, not finding any effects, focused on physical activity. A last note with respect to an agent's relational behaviour: as described by Nguyen and Masthoff, people seem not to care or expect whether a system could understand and care for their feelings, but when a system is represented by a human-like agent, its lack of empathy could lead to negative user experience and worsen the user's attitude towards the system.

In addition, some articles research an agent providing personal information. High self-disclosure positively affects the user's (intention towards) behaviour change and its relationship with the agent (Kang and Gratch, 2011), whereas stories told in first person result in high system usage and usability (Bickmore et al., 2009a).

Furthermore, some research on variability in an agent's behaviour has been performed. Variability in an agent's behaviour positively affects system usage (Bickmore et al., 2010) and intention to use (Bickmore et al., 2010), but, with respect to behaviour change, non-variable behaviour is preferred over variable behaviour (Bickmore et al., 2010). When varying the behaviour of an agent, changing its behaviour with respect to human eye contact behaviour seems to be better than randomly changing its behaviour, since an agent changing its behaviour with respect to human eye contact behaviour is perceived to be more normal and realistic (Grillon and Thalmann, 2008).

Some last remarks, based on research presented in single articles. First, allowing users to control an agent's prosody (the stress and intonation patterns of an utterance) and facial expressions when the agent's task is to retell a story results in high satisfaction (Bickmore and Ring, 2010). Second, users rate their characteristics (e.g. intrinsic motivation and self-efficacy) higher after interaction with an interactive coach than after interaction with a non-interactive coach and higher after

interaction with a moving coach than after interaction with a non-moving coach (Frost et al., 2012). Furthermore, adding rap music to a dialogue positively affects engagement and the user's relation with the agent, whereas rap music reduces trust in an agent (Olafsson et al., 2017). The presence or absence of rap music did not influence the perception of the agent's characteristics (e.g. naturalness, knowledgeability, perceived similarity and liking), system usage, intention to (continue) using the agent and the systems usability. Also, presenting psycho-education via text results in higher task adherence than when an agent provides the psycho-education verbally (Tielman et al., 2017), since psycho-education in text was better recollected. Finally, linguistic tailoring had no effect with respect to persuasion in the context of behaviour change (Yin et al., 2010).

3.1.2. Looks

Other research identified in the review focuses on the agent's looks. Table 4 provides a summary of the effects found for the different outcome variables with respect to the agent's looks. Research has been performed on the subcategories species, realism, styling and socio-demographics. *No article in the review presented research on effects of agents in either 2D or 3D.*

Just one article researched the agent's species and a few the agent's realism. Research shows mixed results with respect to the best rendering style. Although stylised agents are rated positively on characteristics such as friendliness (Ring et al., 2014; Robertson et al., 2015), several studies indicate that human agents are preferred over abstract, and stylised (cartoon-like) agents (Forlizzi et al., 2007; Ring et al., 2014; Robertson et al., 2015; Wissen et al., 2016). However, the application goal, participants and study methods of the studies varied a lot. With respect to intention to use, no effect of rendering style was found (Wissen et al., 2016).

Second, some articles research the agent's clothing and body shape. A professional looking agent, dressed in a white coat and wearing a stethoscope, is positively rated on its characteristics (e.g. credibility, trustworthiness, reassurance, caring and friendliness), relation with the user and intention to use (Parmar et al., 2018) compared to a casually dressed agent. Regarding the agent's body shape, literature shows mixed results. With respect to behaviour change, some research shows a preference for attractive agents above unattractive agents (Nguyen and Masthoff, 2007; Schmeil and Suggs, 2014; Skalski et al., 2007), whereas other research shows a preference for non-ideal, fatter characters above ideal, slim characters (van Vugt et al., 2006). Also, with respect to the perception of the agent's characteristics, relation with the agent and intention to use, results show positive effects for non-ideal body shapes (van Vugt et al., 2006). Although the studies show different results, the target users and application goal were similar.

Lastly, some articles research the agent's demographics. Literature does not show a clear consensus when it comes to preference for a particular gender. Some research indicates a preference for female agents (Alsharbi and Richards, 2017), whereas other research shows that the preferred gender depends on the task of the agent (Forlizzi et al., 2007). However, the studies differed in target group (children vs. adults) and application goal (providing medical advice or physical activity training vs. treatment of anxiety and post-traumatic stress disorder). In addition, no clear consensus exists on the age of the agent; some research suggests that young agents are preferred over old agents (Wissen et al., 2016), whereas other research suggests that users prefer agents of the same age or older (Alsharbi and Richards, 2017). Again, the studies differed in target group (older adults vs. children) and application goal (increase physical activity and medication vs. treatment of anxiety and post-traumatic stress disorder). Also, some research indicates an agent having the same cultural background as the user is more positively rated on its characteristics (e.g. perception of caring, general liking) (Alsharbi and Richards, 2017; Yin et al., 2010) and its relation with the user compared to an agent with a different cultural background (Zhou et al., 2014), whereas, with respect to behaviour change, agents with a

different cultural background could be beneficial (Yin et al., 2010; Zhou et al., 2017). However, the studies targeted different users and researched applications with different goals. In addition, the studies were in different stages of change (ranging from I to III). Furthermore, some research indicates a preference for an agent that acts as a friend (Alsharbi and Richards, 2017), whereas other research indicates that the preferred agent role relates to the agent's task (Nguyen and Masthoff, 2009). Though, the studies focused on a different target group (children vs. adults) and application goal (treatment of anxiety and post-traumatic stress disorder vs. mood manipulation).

Two final remarks with respect to the agent's looks. First, several studies stress the importance of aligning the agent's looks to the looks of the user (Malhotra et al., 2016; Robertson et al., 2015). Similarity with the agent seems to influence the perception of the characteristics of and preferences for particular agents (Zhou et al., 2014). It seems that some users prefer agents that are similar to themselves, for example, in age (Alsharbi and Richards, 2017), body shape (van Vugt et al., 2006) and cultural background (Alsharbi and Richards, 2017; Zhou et al., 2017). Second, literature indicates that preference for particular agents and perception of their personalities depend on the task of the agent (e.g. providing medical advice, encouraging to perform physical activity) (Nguyen and Masthoff, 2009; Ring et al., 2014). It seems that people apply human stereotypes to agents, and therefore, for example, have preferences for a particular gender for a particular task (e.g. male agents are preferred for the role of athletic trainer) (Forlizzi et al., 2007).

4. Discussion

In this paper, we provided a state of the art of design features for ECAs in eHealth, showing a field that is immature and without consensus on effective design features. Emotion and relational behaviour seem to have positive effects, but do not necessarily lead to behaviour change in the context of eHealth. In the remainder of this section, we discuss the design features researched, the outcome variables on which the effect of these design features were researched and what the measured effects were.

4.1. Design Features for ECAs in eHealth

The included articles show that most of the research focused on speech and/or textual output, gaze and facial expressions and hand and body gestures, and not on an agent's looks. Therefore, we see an opportunity for future work on the agent's looks. We are not aware of a literature review identifying design features for ECAs in other contexts to compare our results with. However, articles that research speech and/or textual output, facial expressions and hand and body gestures (Acosta and Ward, 2011; Berry et al., 2005; Kim et al., 2007; Lee et al., 2007; Pelachaud, 2009; von der Pütten et al., 2009) or the agent's looks (Baylor and Kim, 2004; Cowell and Stanney, 2003; Guadagno et al., 2007; Khan and Angeli, 2009; Khan and Sutcliffe, 2014; Kim et al., 2003; 2007; Lee et al., 2018; Rosenberg-Kima et al., 2008; Straßmann and Krämer, 2017; Veletsianos, 2010) in general, or in another context than eHealth, do exist. Therefore, we believe that in other contexts similar design features might have been researched.

4.2. Outcome Variables for Measuring the Effect of Design Features

The measured effect of the design features for ECAs in eHealth was often on the perception of the agent's and users characteristics, relation with the agent, system usage, intention to use, usability and behaviour change. Again, although we are not aware of a literature review identifying outcome variables used to evaluate ECAs in other contexts, we do see similar outcome variables researched in other contexts by individual articles. For example, research has been performed on the perception of the agent's characteristics (Kim et al., 2007; Pelachaud,

2009), user's characteristics (Kim et al., 2007), relation with the agent (Acosta and Ward, 2011; Lee et al., 2007; von der Pütten et al., 2009) and behaviour change (Berry et al., 2005; Kim et al., 2007; Lee et al., 2007). Therefore, we believe that outcome variables measured in other contexts might be similar to the outcome variables researched in the health context.

4.3. Measured Effects of the Design Features

4.3.1. Effects of an ECA's Speech and/or Textual Output, Facial and Gaze Expressions and Hand and Body Gestures

Existing literature shows some consensus on the effects of an ECAs emotion and relational behaviour, but no consensus on the effects of other design features with respect to an agent's speech and/or textual output, facial and gaze expressions and hand and body gestures. Research on design features for ECAs in other contexts supports our findings.

First, some positive effects of an *agent's emotion* were found in research in other contexts, such as e-learning. Research shows that an agent's positive emotions positively affect the users perception of an agent's characteristics, such as the ability to facilitate learning (Kim et al., 2007) and that the way emotion is implemented might affect the function of emotion (e.g. attract the user's attention, persuade the user) (Pelachaud, 2009). Whereas our literature review does not show a clear effect of emotion on behavior change, existing research in other contexts shows some positive effects, such as an agent's emotion increasing the user's interest in learning (Kim et al., 2007) and increasing the user's cognitive performance (Berry et al., 2005).

Second, research in other contexts shows positive results for the implementation of *relational behaviour*. For example, research on pedagogical and co-learner agent's shows that an agent's empathy positively impacts students self-efficacy (Kim et al., 2007), adapting the agent's speech to the emotional state of the user results in higher rapport (Acosta and Ward, 2011), that trust in the agent is higher for a caring co-learner compared to a non-caring co-learner agent (Lee et al., 2007) and that a higher mutual awareness when increasing the agent's behavioural realism (von der Pütten et al., 2009). Whereas the articles in the review mainly show no effect of relational behaviour on behaviour change, some articles in other contexts do show a positive effect of relational behaviour on behaviour change. For example, in the context of e-learning, research shows an increased learner interest (Kim et al., 2007) and increased learning (Lee et al., 2007) when implementing relational behaviour.

For *other design features* with respect to speech and/or textual output, facial and gaze expressions and hand and body gestures, results show no clear consensus. Either few studies have been performed on these features, which makes it difficult to generalize the results, or results show contradictory effects. Differences might be caused by the studies involving different target groups, ranging from children to older adults, or the applications having too different goals, ranging from mood manipulation to increase of physical activity. Therefore, more research is needed:

We recommend to perform research on the effect of design features regarding the agent's speech and/or textual output, facial and gaze expressions and hand and body gestures on the same outcome variables and with a similar target group and application goal.

4.3.2. Effects of an ECA's Looks

Until now, just a few studies have been carried out on the agent's looks in the context of eHealth. No consensus exist on the agent's species and rendering style, and no research has yet been performed on the effects of agent's in either 2D or 3D. Also, different opinions on the most appropriate agent demographics exist. Our review does not show consensus with respect to the preferred gender, age, role and cultural background. Also, an agent's clothing and body shape seem to be

factors to take into account when it comes to creating a positive perception of the agent. These results are in line with research in other contexts, for example in the context of e-learning, showing mixed results on the agent's rendering style (Straßmann and Krämer, 2017), gender (Baylor and Kim, 2004; Cowell and Stanney, 2003; Guadagno et al., 2007; Kim et al., 2007; Rosenberg-Kima et al., 2008), age (Cowell and Stanney, 2003; Lee et al., 2018), role (Baylor and Kim, 2004; Kim et al., 2003), cultural background (Baylor and Kim, 2004; Cowell and Stanney, 2003) and the agent's clothing and body shape (Baylor and Kim, 2004; Khan and Angeli, 2009; Khan and Sutcliffe, 2014; Veletsianos, 2010). What should be noted is that some articles included in the review stress the importance of aligning the agent's looks, especially its demographics age and gender, to the looks of the user. Research in other contexts, such as e-learning, supports this note (Baylor, 2009; Guadagno et al., 2007; Gulz and Haake, 2005; Lee et al., 2018; Rosenberg-Kima et al., 2008; Straßmann and Krämer, 2017; Veletsianos, 2010). Thus, it seems important to personalise the agent's looks:

We see opportunities for future work researching the effect of design features regarding the agent's looks in relation to the characteristics of the user.

4.3.3. Transferring Effects to a Long-term, Daily Life Setting

The research area on the design of ECAs in eHealth is relatively immature. Most of the articles describe research in stage I or II of the renewed framework of evaluation for telemedicine (Jansen Kosterink et al., 2016). Therefore, we learned about the effects of the agent's design features in a lab setting, but do not yet know how these effects translate to a daily life setting for which the ECAs are designed. In addition, in most lab studies users interacted with the agent's for a short period of time, which is different from the long-term interaction for which the majority of the ECAs are designed. Therefore, we should be careful with interpreting the results of this literature review:

We see an opportunity for replicating the studies identified in this literature review that were in stage I or stage II of the renewed framework of evaluation for telemedicine, in a long-term, daily life setting (stage III or IV)

4.4. Applying Design Guidelines from Other Contexts

As indicated above, the findings of our review are in line with results of general research on design features for ECAs or research on design features for ECAs in another context. Therefore, we believe that general guidelines, like guidelines for designing personalities for social agent's by Dryer (Dryer, 1999) and design guidelines for other contexts, such as guidelines of the Enhancing Agent Learner Interactions (EnALI) Framework (Veletsianos et al., 2009), might be applicable to agent's in eHealth as well. However, research also indicates that the user's perception of ECAs depends on the agent's task (Baylor, 2009). Whereas the task of agent's in eHealth might be similar to agent's in other contexts, such as being informative, like embodied chatbots on commercial websites, agent's used in museums or pedagogical agent's, the task of the agent might differ from the task of these agent's when supporting behaviour change in the health domain. Therefore, we should research whether results of other contexts are still applicable to eHealth:

We suggest to repeat studies focusing on the effects of design features in a general or other context in the eHealth context.

4.5. Strengths and Limitations Review

The strength of this review was that it focused on design features for ECAs in a specific context, eHealth, since preferences for agent designs

might be context-dependent. However, this narrow focus could also be a limitation of the research. General research – research not restricted to a particular context – on agent design features might still be applicable to a health context. In addition, research regarding design features in other contexts, such as research on design features for pedagogical agent's, is not included in the review. Some of the results might still be applicable to the eHealth context.

5. Conclusion

This literature review identified (1) the *researched design features* for ECAs in eHealth, (2) the *outcome variables* used to measure the effect of these design features and (3) the *found effects* for each variable. Results show that the agent's speech and/or textual output and its facial and gaze expressions were the most common design features, whereas little research was performed on the agent's looks. The measured effect of these design features was often on the perception of the agent's and user's characteristics, relation with the agent, system usage, intention to use, usability and behaviour change. With respect to the effects found, consensus on design features of ECAs in eHealth is far from established.

Appendix

Table 5

General information of the articles included in the review.

Source	Goal application	N	Gender	Age	Education	Cultural background	Other	Eval. stage	Study design
Alsharbi & Richards (2017)	[M] Medical treatment advice for anxiety and post-trauma stress disorders	11	9 females, 2 males	[C] 11-15		[C] 9 Northern-Western European, 2 Oceania	Playing computer games: all > 3h a week	I	S
Amini et al. (2013)	[A] Changing unhealthy alcohol consumption behaviours	51	37% females, 63% males		[S] University students	[C/H/AA/A] 55% White, 27% Hispanic, 16% African American, 2% Asia	Drinking at least one drink a month	I	E
Amini et al. (2014)	[A] Changing unhealthy alcohol consumption behaviours	81						I	E
Bickmore et al. (2005)	[P] Promoting physical exercise	91						II	E
Bickmore et al. (2009)	[P] Promoting walking behaviour	26	21 females, 5 males	[A] 54-67	[C] 92% had some college education	[C/AA] 80% Caucasian, 20% African American		II	E
Bickmore et al. (2010)	[P] Promoting walking behaviour	E1: 24 E2: 26	E1: 17 females, 7 males, E2: 21 females, 5 males	E1: [A] 55-67, E2: [A] 54-67	E2: [C] 92% had some college education	E2: [C/AA] 80% Caucasian, 20% African American		E1: II E2: II	E
Bickmore & Picard (2004)	[P] Promote physical exercise	60	60% female		[S] 69% students			II	E
Bickmore & Picard (2005)	[P] Motivating to performing physical exercises	101						II	E
Bickmore & Ring (2010)	[O] Promoting pre-conception care	9	All female	[A] 20-24	[C] 1: high school level, 8: at least some college	[AA] All African American		I	E
Bickmore & Schulman (2007)	[M] Calming users using negative mood induction	16	9 females, 7 males	[A] 18-30	[S] 15 were students			I	E

(continued on next page)

Table 5 (continued)

Creed & Beale (2012)	[N] Changing nutrition habits	50	28 females, 20 males					II	E
Creed et al. (2015)	[N] Improving fruit and vegetable consumption	45	17 females, 26 males, 1 not specified	[A] M = 30.53 (SD = 9.33)	[S / U] Graduates and undergraduates school of computer science	[C] UK and three other countries	No dietary plan, no history of eating disorders, in contemplation, preparation or action stage	II	E
Forlizzi et al. (2007)	[T / P] Being a medical advisor, athletic trainer	24	15 females, 9 males	[A] 18-36, M = 25.7				I	S
Frost et al. (2012)	[P] Raise intrinsic motivation in children exercising	61	39 girls, 22 boys	[C] M = 9.2 (SD = 0.94)		[C] Dutch		I	E
Grillon & Thalmann (2008)	[M] Reducing social phobia (in public speaking)	12		[A] 25-35				I	E
Kang & Gratch (2010)	[M] Psychotherapy	57	53% females, 47% males	[A] M = 30.68 (SD = 10.08)				I	E
Lisetti et al. (2013)	[A] Changing unhealthy alcohol consumption behaviours	81			[S] University students			I	E
Malhotra et al. (2016)	[O] Assisting people with cognitive disabilities in daily life	22	14 females, 8 males	[E] Elderly			All have Alzheimer's disease	I	E
Nguyen & Masthoff (2007)	[P] Promoting physical exercise	E1: 51 E2: 59			E2: [S/U] University staff and graduate students			I	E1: Q, S E2: E
Nguyen & Masthoff (2009)	[M] Performing behaviour modification via the mood manipulation technique	84	47 females, 37 males	[A] M = 25.4 (SD = 7.9)	[S/U] University students and staff			II	E
Olafsson et al. (2017)	[P / N] Motivating to exercise and good nutrition	84	60% male	[A] M = 22.8 (SD = 2.4)	[S] Students	[A/C/AA] 79% Asian, 18% White, 2% Hispanic, and 1% Black	Like rap	I	E

(continued on next page)

Table 5 (continued)

Parmar et al. (2018)	[T] Discussing the importance of having a healthcare proxy and providing information about receiving life-support via a medical ventilator	E1: 43 E2: 308	E1: 20 females, 23 males, E2: 175 females, 133 males	E1: [A] 16-62, E2: [A] 18-76, M = 38 (SD = 12.32)				I	E1: S E2: E
Ring et al. (2014)	[T] Performing social interactions on user's favourite books and medical counseling (cancer related)	E1: 67, E2: 47	E1: 31 females, 36 males, E2: 16 females, 31 males					I	E1: E E2: E
Robertson et al. (2015)	[T] Helping patients (with lower health literacy and socio-economic status) to understand the benefits and drawbacks of different treatment choices in response to prostate cancer		Location I: Men with prostate cancer and a few wives of patients, location II (target group): All male		Location I: [C] Higher education than location II, location II: [L] Lower education than location I	Location I: [C] Caucasian, location II: [AA] African American	Location I: Higher health literacy and socio-economic status than location II	I	Q
Schmeil & Suggs (2014)	[N / P] Change eating and or physical activity behaviours	512	219 females, 293 males	[A] M = 33		[C] 186 from Germany, 177 from Poland, 149 from the UK	More inclined to healthier weight than the European average	II	E
Silverman et al. (2001)	[T] Helping with dealing with a variety of heart attack scenarios	18			[C / U] Cardiologists and general practitioners			I	E
Skalski et al. (2007)	[T] Providing information about risks and treatment of problems associated with high blood pressure	125	83 females, 42 males	M = 21.56	[S] Undergraduate students at a large Midwestern university			II	E
Tielman et al. (2017)	[M] Providing psychoeducation: expressive writing for Post-Traumatic Stress Disorder therapy	46	12 females, 38 males	M = 22.9 (SD = 3.8)	[S / U] University staff and students	[C] All had Dutch as first language		II	E
van Vugt et al. (2006)	[O] General eHealth context	80	56 females, 24 males	M = 23 (SD = 7.8)	[S] University students of Free University of Amsterdam		Body mass index: M = 22 kg/m2 (SD = 3.8)	I	E
Wissen et al. (2016)	[P / T] Providing information about importance physical activity, reminders for medication intake and appointments with health care professionals	64	34 females, 30 males	[E] < 65: 6.3 %, 65-70: 62.5%, > 70: 31.3%	[C / U] Highly educated: 59.4% => associate degree	[C] 95.3% Caucasian	<ul style="list-style-type: none"> • 54.7% living with spouse • Chronic illnesses with medication intake 	I	E
Yin et al. (2010)	[P] Promoting physical exercise	43	44.2% female	18-65		[H/C] 65.0% Latinos: Speak, read and understand English and Spanish: A) Anglo-American born in U.S., in Spanish-speaking Latin American countries for at least two months, English as first language, or B) Latino/Latina born in Spanish-speaking Latin American country, in U.S. for at least two months, Spanish as first language.		II	E
Zhou et al. (2014)	[T] Performing hospital discharge	764	50.8% female	18-90 (M = 49)		[AA/C/H] 52.6% African American, 29.1% Caucasian, 11.4% Hispanic		III	E
Zhou et al. (2017)	[P] Promoting physical activity to young Chinese adults living in the U.S.	49	46.9% male	[C/A] 9-34, M = 26.1 (SD= 3.6)	[C/U] 85.7% college graduate or higher	[A] Speak and read English, born in China, moved to U.S. at least by age of 16, lived in U.S. for at least six months (time in U.S.: M = 2.8 years, SD = 1.8)		II	E

Goal application: A = alcohol consumption / M = mental health / N = nutrition / P = physical activity / T = medical treatment / O = other
 Age: A = adults / C = children / E = elderly
 Education: L = low / C = some college / S = students / U = university
 Cultural background: A = Asian / AA = African American / C = Caucasian / H = Hispanic
 Evaluation stage: I = technical efficacy / II = specific system objectives / III = system analysis / IV = external validity
 Study design: E = experimental / O = observational analytic / Q = qualitative / S = survey

Table 6

Overview of the articles included in the review. For every article, the feature category, design feature(s) researched, outcome variable(s), methods and outcomes are listed.

	Source	Feature category	Design feature(s)	Outcome variable(s)	Method	Outcomes
	Speech and/or textual output, facial and gaze expressions and hand and body gestures					
Emotion	Creed & Beale (2012)	Speech and/or textual output, facial and gaze expressions	Presence or absence of emotion	<ul style="list-style-type: none"> [AC] Caring [AC] Likeability [AC] Trustworthiness [AC] Intelligence [AC] Feeling of positivity [AC] Feeling of support [AC] Level of emotion agent [U] Time viewing educational resources [U] Number of articles and recipies viewed [IT] Desire to work with agent again [O] General opinion via open questions 	Two conditions: <ul style="list-style-type: none"> Emotional (E): emotional facial expressions, appropriate emotion in audio, response manipulated to be appropriate to responses user Non-emotional (NE): neutral facial expressions, neutral audio, neutral responses 	<ul style="list-style-type: none"> Caring, likeability: E > NE Trustworthiness, intelligence, feeling of positivity, feeling of support: no difference Time viewing educational resources, number of articles viewed: no difference Desire to work with agent again: <ul style="list-style-type: none"> Not interacting again? NE: 12 people, E: 2 people Many people preferred to use the website only Open questions: <ul style="list-style-type: none"> Both conditions: visual appearance was liked NE: perceived less positively, neutral voice disliked NE: comments on visuals E: emotional voice liked E: positive comments related to the emotional expressions (liked that agent was friendly, warm, and cheerful)
	Creed et al. (2015)	Speech and/or textual output, facial and gaze expressions	Presence or absence of emotion	<ul style="list-style-type: none"> [AR] Working Alliance Inventory (WAI) [B] Fruit and vegetable intake baseline and after intervention [B] Average fruit and vegetable intake [B] Days per week eating >=5 portions [U] Dropout [U] Number of days logged in [U] Number of times logged in [U] Amount of articles and recipies viewed [U] Number of times viewed different resources [UC] Self-efficacy [UC] Decisional balance 	Two conditions: <ul style="list-style-type: none"> Emotional (E): facial expressions and recorded speech Non-emotional (NE): neutral expression and voice 	<ul style="list-style-type: none"> WAI: no difference Fruit and vegetable intake compared to baseline: NE > E Average fruit and vegetable intake: no difference Days per week eating >=5 portions: no difference Dropout: no difference Logins, articles viewed and resources accessed: no difference Self-efficacy: no difference Decisional balance: no difference

(continued on next page)

Table 6 (continued)

Emotion	Silverman et al. (2001)	Speech and/or textual output, facial and gaze expressions	Amount of emotional behaviour	<ul style="list-style-type: none"> [B] Intention shift (pre and post intervention) for: <ul style="list-style-type: none"> Self care during attack Waiting to call Calling/seeing a doctor Calling 9-1-1 [B] Scores on knowledge quiz (tests knowledge on heart attack symptoms and delay issues) [UX] Usability [AC] Logic (and believable) [AC] Pedagogy 	Four prototypes (focus groups): <ul style="list-style-type: none"> Multimedia versions (M): <ul style="list-style-type: none"> Full Emotive Basic Agent Didactic versions (D): <ul style="list-style-type: none"> Pop-up Text Didactic/Text 	<ul style="list-style-type: none"> Intention shift: <ul style="list-style-type: none"> Pop-up Text (140.6%) > Full Emotive (125%) > Didactic/Text and Basic Agent (37.5%) Shift for Full Emotive, Basic Agent, Didactic/Text <ul style="list-style-type: none"> Shift in call 9-1-1: no difference D and M Shift in self care: M (140%) > D (56.3%) Intention to call or see a doctor: M > D Knowledge quiz: all: change in knowledge, after-the-fact learning: D best, post-intervention: M > D Usability: <ul style="list-style-type: none"> Overall: positive Highest for Full Emotive Drops for each media-crippled version Didactic/Text lowest ratings (despite being the shortest module with the least complexity) Pedagogy: full Emotive, Basic Agent, Pop-up Text result in: (1) vividness that helps better remember results, (2) learning about symptoms overall, where symptoms occur in the heart, and about delay issues, (3) subjects learning more how time slips by during a heart attack event Logical (and believable): all conditions: logical and believable, least: Basic Agent, Full Emotive > Basic Agent
	Amini et al. (2013)	Facial and gaze expressions	Presence or absence of empathy	<ul style="list-style-type: none"> [UX] Attitude (ATT) [IT] Intention to Use (ITU) [UX] Perceived Enjoyment (PENJ) [UX] Perceived Ease of Use (PEOU) [AC] Perceived Sociability (PS) [UX] Perceived Usefulness (PU) [AR] Social Presence (SP) [AR] Trust (TRUST) [AR] Anxiety (ANX) [AR] Social Influence (SI) [AC] Anthropomorphism (ANT) [AC] Likeability (LIKE) [AC] Animacy (ANIM) [AC] Perceived Intelligence (PI) [AR] Perceived Safety (PSA) 	Two conditions: <ul style="list-style-type: none"> Non-empathic (NE): neutral facial expression, does not empathize with user Empathic (E): facial expressions, head gestures, smile, head posture mimicry, eye brow movement, mutual gaze, lip, synchronized verbal expressions 	<ul style="list-style-type: none"> ATT, ITT, PENJ, PEOU, PS, PU, SP, TRUST, SI, ANT, LIKE, ANIM, PI, PSA: E > NE ANX: No difference E and NE
Relational, empathic behaviour	Amini et al. (2014)	Facial and gaze expressions	Presence or absence of empathy	<ul style="list-style-type: none"> [UX] Attitude (ATT) [IT] Intention to Use (ITU) [UX] Perceived Enjoyment (PENJ) [UX] Perceived Ease of Use (PEOU) [AC] Perceived Sociability (PS) [UX] Perceived Usefulness (PU) [AR] Social Presence (SP) [AR] Trust (TRUST) [AR] Anxiety (ANX) [AR] Social Influence (SI) 	Three conditions: <ul style="list-style-type: none"> Non-empathic (NE): neutral facial expression, does not empathize with user Empathic (E): facial expressions, head gestures, smile, head posture mimicry, eye brow movement, mutual gaze, lip, synchronized verbal expressions Text-only (T) 	<ul style="list-style-type: none"> ATT, ITU, PENJ, PS, PU, SP, TRUST: E > NE, E > T ATT, ITU, PEOU, SP: NE > T PENJ, PS, PU, TRUST, SI: T > NE
	Bickmore et al. (2005)	Speech and/or textual output, facial and gaze expressions, hand and body gestures	Presence or absence of relational behaviour	<ul style="list-style-type: none"> [AR] Working Alliance Inventory (on day 7 and 27) [AC] Likability agent [AC] Closeness relationship agent [UX] Usefulness discussions agent [IT] Desire to continue working with agent [U] Nr. of pages with educational content accessed [B] Nr. of days a week physical activity performed 	Three conditions: <ul style="list-style-type: none"> Control (web forms only, no agent) Non-relational (NR): talks about exercise behavior, negotiates and follows up on exercise goals, suggestions to overcome obstacles to exercise, cognitive-behavioral techniques for health behavior change Relational (R): social dialogue, empathetic feedback, meta-relational communication, humor, close forms of address, nonverbal immediacy 	<ul style="list-style-type: none"> Bond-subscale of WIA: R > NR (on day 7 and 27) Overall WAI: no differences Likability agent, desire to continue: R > NR Educational pages viewed: R and NR > C Physical activity: R and NR > C
	Bickmore & Picard (2004)	Speech and/or textual output, facial and gaze expressions, hand and body gestures	Presence or absence of caring	<ul style="list-style-type: none"> [AR] Working Alliance Inventory (WAI) [IT] Willingness to continue [AR] Selection type of farewell messages 	Two conditions: <ul style="list-style-type: none"> Relational, caring (R): empathetic feedback, concerned facial expressions Non-relation, non-caring (NR): no feedback, no concerned facial expressions 	<ul style="list-style-type: none"> Perception caring: R > NR WAI items "Agent cares about me", "Agent is genuinely concerned", "Agent likes me" and "Agent and I trust in one another": R > NR Willingness to continue working: R > NR Selection of sentimental farewell: R > NR

(continued on next page)

Table 6 (continued)

Relational, empathic behaviour	Bickmore & Picard (2005)	Speech and/or textual output, facial and gaze expressions, hand and body gestures	Presence or absence of relational abilities	<ul style="list-style-type: none"> [AR] Working Alliance Inventory (WAI) [AC] Likability agent [AR] Closeness relationship [IT] Desire to continue [IT] Desire to continue using the overall FitTrack system [AR] Sentimentality farewell [O] Opinion on whom helped best to perform exercises [B] Minutes moderate activity a day [B] Days a week > 30 minutes moderate activity [B] Steps per day [B] Days per week > 10,000 steps [U] Average nr. of educational pages accessed per session 	<p>Three conditions:</p> <ul style="list-style-type: none"> Relational (R): daily social dialogue, frequent meta-relational dialogues, friend role, small social distance, daily empathy exchanges, humour, daily continuity behaviours, high non-verbal immediacy Non-relational (NR): no social dialogue, minimal meta-relation dialogues, stranger role, large social distance, no empathy exchanges, no humour, no continuity behaviours, low non-verbal immediacy Control (C): no agent 	<ul style="list-style-type: none"> Bond subscale WAI: R > NR Likability: R > NR Closeness relationship: R > NR Desire to continue: R > NR Desire to continue system: no difference Sentimentality farewell: R > NR Who helped best: no difference Who helped best: agents > friends, family and workout buddy Physical activity: for both groups increase after intervention Physical activity: for both groups decrease during follow up (decrease R > NR) Days a week > 30 minutes moderate activity : R and NR > C Days per week > 10,000 steps: R and NR > C Pages accessed: R and NR > C
	Bickmore & Schulman (2007)	Speech and/or textual output, facial and gaze expressions	Expressivity user, empathic accuracy	<ul style="list-style-type: none"> [B] Change in positive and negative affect [UX] Satisfaction [IT] Willingness to continue working [AC] Likelihood [AC] perception of caring 	<p>Two conditions:</p> <ul style="list-style-type: none"> EXPRESSIVITY: high expressivity dialogue, free speech input EMPATHY: high empathic accuracy and empathic feedback, including happy facial display and close-up facial display for other feeling states, input limited to predefined answers 	<p>Empathic accuracy more important than user expressivity:</p> <ul style="list-style-type: none"> Change in positive affect from pre-induction to post-intervention: EMPATHY > EXPRESSIVITY Satisfaction: EMPATHY > EXPRESSIVITY Females showed greater desire to continue with EXPRESSIVITY agent than males
	Lisetti et al. (2013)	Facial and gaze expressions	Presence or absence of empathic abilities	<ul style="list-style-type: none"> [UX] Attitude (ATT) [IT] Intention to Use (ITU) [UX] Perceived Enjoyment (PENJ) [UX] Perceived Ease of Use (PEOU) [AC] Perceived Sociability (PS) [UX] Perceived Usefulness (PU) [AR] Social Presence (SP) [AR] Trust (TRUST) [AR] Anxiety (ANX) [AR] Social Influence (SI) [AC] Anthropomorphism (ANT) [AC] Likeability (LIKE) [AC] Animacy (ANIM) [AC] Perceived Intelligence (PI) [AR] Perceived Safety (PSA) 	<p>Three conditions:</p> <ul style="list-style-type: none"> Text-only interface (T) Non-empathic ECA (NE): neutral facial expressions, does not empathize with user, no reflection on user's answers, ignores user's emotional state ECA with some empathic abilities (E): emotional facial expressions (happy, sad, concerned, surprised, and neutral); head gestures (nodding); big/subtle smile; head posture mimicry (pitch, yaw, roll); eyebrow movement and lip synchronized verbal reflections 	<ul style="list-style-type: none"> ATT: E > NE, E > T. All means positive ITU: E > NE, E > T. NE and E positive, T negative mean PENJ: E > NE, E > T. All means positive PEOU: No differences. All means positive PS: E > NE, E > T, T > NE PU: E > NE, E > T. All means positive SP: E > NE, E > T. E positive NE and T negative mean Trust: E > NE, E > T. E and T positive, NE negative mean ANX: No differences. All means positive SI: No differences. All means positive ANT: E > NE LIKE: E > NE, E > T. All means positive ANIM: E > NE. E positive, NE negative mean PI: E > NE, T > NE. All means positive PSA: E > NE, E > T. All means positive

(continued on next page)

Table 6 (continued)

<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Relational, empathic behaviour</p>	<p>Nguyen & Masthoff (2009)</p>	<p>Speech and/or textual output, facial and gaze expressions, hand and body gestures</p>	<p>Modality (animated vs no visuals), level of empathy (non-empathy, empathy, empathy and expression)</p>	<ul style="list-style-type: none"> [B] PANAS-S (short-form positive and negative affect schedule questionnaire) before mood induction, after mood induction and after web credibility judgment task [B] Performance on second and third task (time reading and judging, total number of thoughts recalled) [UX] Judgement of web credibility (biased, comprehensive, caring, trustworthy, expert) [AC] Attitude toward agent: enjoyability, caring, likeability, trustworthiness, empathy [AC] Perception of agent design (language, appearance, gestures and voice) 	<p>2 x 3 experiment:</p> <ul style="list-style-type: none"> Modality: <ul style="list-style-type: none"> Animated (A): animated, synthesized voice No visual (NV): no visual representation, no voice Intervention: <ul style="list-style-type: none"> Empathy (E): small talk, polite and friendly, acknowledging mood change, offering means to correct agent's judgments, conveying sympathy and empathy, acknowledging validation emotional state Empathy and expressivity (EE): same as EMPATHY, encourages subjects to freely express feelings Non-empathy (NE): ignores feelings, asks distracting questions <p>Three steps:</p> <ul style="list-style-type: none"> First: mood manipulation Second: participant looks for info about alcoholism and consequences on a low and high credible website Third: measuring effect of mood on subject's judgements of memory 	<ul style="list-style-type: none"> Change in negative mood: rise for all conditions, no differences between conditions Change in positive mood: <ul style="list-style-type: none"> Decline for all conditions E: no differences between modalities NE: NV better recovery positive mood > A Time reading website: E and EE > NE Total number of thoughts recalled: <ul style="list-style-type: none"> Recall positive thoughts > recall negative thoughts Recall more positive thoughts for website with high credibility and more negative thoughts for website with low credibility E: NV > A Credibility websites: no effect modality or intervention Attitude towards agent: <ul style="list-style-type: none"> The more positive? the more empathic the agent was perceived, and the more positive mood change occurred before mood induction and after intervention Enjoyability, caring, trustworthiness, likability: E and EE > NE EE: Empathy, trustworthiness, enjoyability, caring, likability: A > NV E and NE: no differences modalities A: Enjoyability, caring, likability, trustworthiness: E and EE > NE NV: no differences intervention Perceived empathy? EE < E and NE Perception agent design: no differences
	<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Providing personal information</p>	<p>Bickmore et al. (2009)</p>	<p>Speech and/or textual output</p>	<p>Person form stories (first person or third person)</p>	<ul style="list-style-type: none"> For each completed conversation: <ul style="list-style-type: none"> [UX] Enjoyability stories [AC] Engagement [AC] Perceived dishonesty counselor [U] System usage 	<p>Two conditions:</p> <ul style="list-style-type: none"> 1st person (narratives as being part of the agent's own life story) 3rd person (narratives as being part from the life of a human friend of the agent with a similar role and occupation)
<p>Kang & Gratch (2010)</p>		<p>Speech and/or textual output</p>	<p>Amount of self-disclosure (via providing backstories)</p>	<ul style="list-style-type: none"> [AR] Co-presence [AR] Social attraction [B] User's self disclosure (verbally) 	<p>Three conditions:</p> <ul style="list-style-type: none"> High disclosure Low disclosure Non-disclosure 	<ul style="list-style-type: none"> High disclosure: more co-presence and social attraction than low-disclosure, non-disclosure High disclosure: people revealed a medium level of personal information more often than in low-disclosure, non-disclosure
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Variable behaviour</p>	<p>Bickmore et al. (2010)</p>	<p>Speech and/or textual output</p>	<p>Variability content, person form stories (first person or third person)</p>	<p>Experiment 1:</p> <ul style="list-style-type: none"> [UX] Perceived repetitiveness of the interaction [IT] Desire to continue [U] System usage [B] Number of steps <p>Experiment 2:</p> <ul style="list-style-type: none"> [AC] Self-report engagement [AC] Enjoyability stories [IT] Looking forward to talk [AC] Perceived dishonesty [U] System usage [B] Number of steps 	<p>Experiment 1, two conditions:</p> <ul style="list-style-type: none"> Non-variable (NV): same dialogue structure and language, appearance and setting never change Variable (V): one of five different dialogue structures randomly selected, topic randomly selected, one of five different background images randomly selected <p>Experiment 2, two conditions:</p> <ul style="list-style-type: none"> 1st person 3rd person 	<p>Experiment 1:</p> <ul style="list-style-type: none"> Perceived repetitiveness: NV > V Desire to continue: V > NV System usage: V > NV Number of steps: NV > V <p>Experiment 2:</p> <ul style="list-style-type: none"> Self-report engagement: no difference, decrease over time in both conditions Enjoyability stories: 1st > 3rd Looking forward: no difference Perceived dishonesty: no difference System usage: 1st > 3rd Number of steps: 3rd > 1st
	<p>Grillon & Thalmann (2008)</p>	<p>Facial and gaze expressions, hand and body gestures</p>	<p>Attitude (bored, attentive, random or changed based on eye contact user)</p>	<ul style="list-style-type: none"> [AC] Interested/indifferent [AC] Engaged/distracted [AC] Behave normal /abnormal 	<p>Four attitude conditions:</p> <ul style="list-style-type: none"> Always bored (looks away) Always attentive (positive attitude) Randomly changed attitude (attentive and bored) Changed attitude on eye contact behaviour (user looks = attentive behaviour, user looks away = boring behaviour) 	<p>Character changing behavior to human eye contact behavior seems more normal than a character that randomly changes behavior:</p> <ul style="list-style-type: none"> Mean interested/indifference: bored: 5 / attentive: 1.67 / random: 3.50 / tracked: 3.50 / tracked (only people that looked): 2.20 / tracked (without people that did not look): 2.71 Mean engaged/distracted: bored: 4.92 / attentive: 1.92 / random: 3.58 / tracked: 3.67 / tracked (only people that looked): 2.80 / tracked (without people that did not look): 3.29 Mean normal/abnormal: bored: 3.83 / attentive: 2.50 / random: 3.33 / tracked: 3.17 / tracked (only people that looked): 2.80 / tracked (without people that did not look): 3.00 Interviews: tendency from to find the tracked version more realistic than either the bored version or the random version

(continued on next page)

Table 6 (continued)

User control prosody & facial expressions	Bickmore & Ring (2010)	Speech and/or textual output, facial and gaze expressions	Degrees of control over the agent's verbal and nonverbal behavior	<ul style="list-style-type: none"> [AC] Helpfulness in creating a story [AC] Confidence that the agent would properly re-tell the story [O] Possibility to express oneself [UX] Ease of use [UX] Overall satisfaction 	<p>Five conditions:</p> <ul style="list-style-type: none"> TEXT (user enters story in a text box, story displayed as text) AGENT (user enters story in a text box, story is spoken by an ECA with non-verbal behaviour and prosody and neutral face) FACE (same as AGENT, but, in addition, the user can specify the agent's facial display (happy, sad or neutral)) VOICE (same as AGENT, but, in addition, the user can vary the agent's prosody: speed, volume, baseline pitch, emphasis and pauses) ALL (combination of VOICE and FACE) 	<p>Users more satisfied when stories are delivered by a virtual agent compared to plain text, more satisfied with tools to control agent's prosody compared to facial display of emotion and most satisfied when they can control all aspects of the agent's delivery:</p> <ul style="list-style-type: none"> For all measures: TEXT rated lowest, ALL highest, and VOICE second highest Helpfulness: difference between TEXT and other conditions and AGENT and other conditions, but not between FACE, VOICE and ALL Satisfaction: difference between TEXT and FACE, TEXT and VOICE, and TEXT and ALL Confidence, possibility expression and ease of use: no differences between conditions
Interactivity	Frost et al. (2012)	Speech and/or textual output, facial and gaze expressions, hand and body gestures	Movement agent, interactivity	<ul style="list-style-type: none"> [UC] Intrinsic motivation (IM) [UC] Co-variates: parasocial interaction (PSI) and Self-efficacy (SE) 	<p>Three conditions:</p> <ul style="list-style-type: none"> Non-interactive virtual coach (NIVC); no compliments, looks away from viewer, directive instructions in impersonal manner and remains stationary Non-interactive moving virtual coach (NIVMC): same audio track as in NIVC one, moves along with viewer to model behavior Interactive virtual coach (IVC): compliments, looks directly at camera, personal and non-directive feedback cues, moves, models behavior 	<p>Parasocial interaction fully explains variation in intrinsic motivation, self-efficacy does not:</p> <ul style="list-style-type: none"> IM, PSI: higher in IVC than other conditions SE: higher in IVC and NIMVC compared to NIVC SE: NIMVC higher than NIVC Model IM: higher IVC compared to NIMC Model IM & SE: IM different for IVC and NIVC Model IM & SE & PSI: no difference IM for NIMVC and NIVC
Rap	Olafsson et al. (2017)	Speech and/or textual output	Presence of rap music in the dialogue	<p>Self-reports:</p> <ul style="list-style-type: none"> [UX] Satisfaction [IT] Desire to continue working with agent [AC] Trust [AR] Closeness relationship [AC] Naturalness [AC] Knowledgeability [AC/UC] Perceived similarity [AC] Liking of the agent <p>Behavioural measures:</p> <ul style="list-style-type: none"> [O] Selection agent to have third conversation with [U] Number of turns of talk with agent in third conversation <p>Interview:</p> <ul style="list-style-type: none"> [O] Concepts [AC] Opinion most engaging agent 	<p>Four conditions:</p> <ul style="list-style-type: none"> RN (rap nutrition) RE (rap exercise) NoRE (no rap exercise) NoRN (no rap nutrition) 	<p>Self-reports:</p> <ul style="list-style-type: none"> Satisfaction: no difference Desire to continue working: no difference Trust? NoR > R Closeness relationship: R > NoR Naturalness, knowledgeability, perceived similarity, liking of the agent: no difference <p>Behavioural measures:</p> <ul style="list-style-type: none"> Selection agent third conversation: R > NoR (62% vs 38%) Nr. of turns third conversation: no difference <p>Interview:</p> <ul style="list-style-type: none"> Concepts reliability and engagement: R > NoR, information and knowledge: NoR > R Most engaging agent: R > NoR (65% vs 35%)
Agent message	Tielman et al. (2017)	Speech and/or textual output	Presentation mode psychoeducation (ECA providing info verbally or by showing text)	<p>Primary measures:</p> <ul style="list-style-type: none"> [O] Suggestion adherence (stressfulness chosen memory) [O] Task adherence (quality of memory description) [O] Subjective adherence (subjective experience recollecting memory) <p>Mediating measures:</p> <ul style="list-style-type: none"> [B] Recollection of information [AC] Attitude to agent <p>Exploratory measures:</p> <ul style="list-style-type: none"> [UC] Psychoeducation attitude [UX] Function usefulness 	<p>Psychoeducation by:</p> <ul style="list-style-type: none"> A) a virtual agent that asks to read psychoeducation in text on the screen B) a virtual agent that provides the psychoeducation verbally <p>Then, participants perform an expressive writing task (describing one of the worst memories in a digital diary)</p>	<ul style="list-style-type: none"> Agent attitude, psychoeducation attitude, usefulness diary, usefulness psychoeducation: significantly higher than the neutral Subjective adherence, usefulness agent, usefulness of the virtual agent speaking: no significant difference from neutral One group considered agent and voice a useful addition, another group did not No effect presentation mode on agent attitude and recollection Suggestion adherence: positively correlated with subjective and task adherence Subjective and task adherence: not correlated Agent attitude and psychoeducation attitude: positively correlated Correlations recollection and task adherence, and agent attitude and task adherence No effect recollection and agent attitude on suggestion adherence and subjective adherence Task adherence: text > VA

(continued on next page)

Table 6 (continued)

Linguistic, cultural tailoring	Yin et al. (2010)	Speech and/or textual output, socio-demographic	Presence or absence of linguistic and cultural tailoring	<p>Independent variables:</p> <ul style="list-style-type: none"> [AC/UC] Linguistic congruity [AC/UC] Cultural congruity [UC] Need for cognition <p>Dependent variables:</p> <ul style="list-style-type: none"> [B] Persuasion: differences in pre- and post ranking task consisting of a list of statements about exercise [AC] Perceived trustworthiness agent 	<ul style="list-style-type: none"> 2 x 2 experiment: <ul style="list-style-type: none"> Spanish (Latina) English (Anglo-American) Low cultural congruity High cultural congruity Agent designs: adapted to culture in appearance, behaviour and values and background Anglo-American: focus on user's well-being, Latina: interest in family and friends 	<ul style="list-style-type: none"> Persuasion: no effects of linguistic congruity Persuasion, trustworthiness: interaction effect of cultural congruity and need-for-cognition Need for cognition: <ul style="list-style-type: none"> People predisposed to be influenced by peripheral cues (low need for cognition)? Trust and persuasion: own > different culture People predisposed to process information deeply (high need for cognition)? Persuasion: different > own culture
Looks						
Rendering style	Forlizzi et al. (2007)	Soci-demographics, species	Gender, species (human or abstract)	<ul style="list-style-type: none"> [AC] Likability agent per task (athletic trainer, medical advisor) 	Rating multiple static images differing in, amongst others, gender and species	<ul style="list-style-type: none"> Athletic trainer: male > female Medical advisor: female > male Both: humans (females and males) > abstract
	Ring et al. (2014)	Realism	Shading styles (realistic, toon shaded), character proportions (of a human, cartoon)	<ul style="list-style-type: none"> [AC] Realism [AC] Appeal [AC] Familiarity [AC] Eeriness [AC] Friendliness [AC] Trustworthiness [UX] Easiness to interact with [IT] Desire to continue working [AC] Likability [AC] Caring [AC] Appropriateness [AC] Quality of motion 	<p>Experiment 1, four conditions:</p> <ul style="list-style-type: none"> Shaded agent - social task (user's favorite books and movies) Realistic agent - social task Shaded agent - medical task (cancer related topics) Realistic agent - medical task <p>Experiment 2, four conditions:</p> <ul style="list-style-type: none"> Proportions human - social task Proportions cartoon - social task Proportions human - medical task Proportions cartoon - medical task 	<p>Experiment 1:</p> <ul style="list-style-type: none"> Familiarity, trustworthiness, appropriateness: medical > social Friendliness: social > medical, shaded > realistic Likability: shaded > realistic (only social task) Caring: trend toward significance: medial task: realistic > shaded, social task: shaded > realistic <p>Experiment 2:</p> <ul style="list-style-type: none"> Familiarity, trustworthiness, appropriateness, caring: medical > social Friendliness: cartoon > human Appropriateness medical task: human > cartoon Appropriateness social task: cartoon > human
Rendering style	Robertson et al. (2015)	Realism	Amount of realism (photographic, highly stylized renderings to abstract)	<p>Early focus groups:</p> <ul style="list-style-type: none"> [AC] Opinion on four rendering styles of static agent images (photograph, vectorized photo, realistic illustration and cartoon) <p>Final focus groups:</p> <ul style="list-style-type: none"> [AC] Satisfaction, feasibility, acceptability, agent as patient decision aid and general opinion on agent prototype (cartoon style) with interaction 	<p>Early focus groups:</p> <ul style="list-style-type: none"> Four rendering styles presented <p>Final focus groups:</p> <ul style="list-style-type: none"> Agent prototype in cartoon style 	<p>Early focus group Emory: negative reaction towards cartoon (unappealing)</p> <p>Early focus group Grady Memorial Hospital:</p> <ul style="list-style-type: none"> Cartoon: dislike, but acceptable Vectorized photo: looked "fake" Photographic agent: "good" and "looked cool" Preference: cartoon > photographic <p>Final focus groups:</p> <ul style="list-style-type: none"> Grady: cartoon is "OK" Emory: cartoon disliked, prefer "real person" Satisfaction: Grady > Emory Feasibility and acceptability: Grady < Emory Agent as a patient decision aid: low rating
	Wissen et al. (2016)	Socio-demographic, realism	Similarity (age: young or old), Familiarity (familiar: a representation of an actual female health coach/nurse, unfamiliar), rendering (photo-realistic, stylized)	<p>User acceptance:</p> <ul style="list-style-type: none"> [AC] Friendliness [AC] Competence [AC] Attractiveness [AC] Likeability [AC] Trustworthiness [IT] Following advice [IT] Continue interaction [AC] Satisfaction <p>Attitude towards use:</p> <ul style="list-style-type: none"> [UX] Easy to use [UX] Liking working with [UX] Useful to support health management [UX] Enable health goal achievement [UX] Easier to manage health [UX] Useful [UX] Make life more interesting [UX] Help when necessary 	<p>3 x 2 design:</p> <ul style="list-style-type: none"> Appearance: 3 between-subject conditions: older (unfamiliar) vs. younger (unfamiliar) vs. familiar Rendering style: 2 within-subject conditions: stylized vs. photo-realistic 	<p>User acceptance:</p> <ul style="list-style-type: none"> All outcome variables: photo-realistic > stylized Friendliness: older > familiar All other outcome variables: no difference between older, younger and familiar <p>Attitudes towards use of VC:</p> <ul style="list-style-type: none"> No effects of rendering style Four of seven variables: effect of appearance Easy to use, like working with and enable health goal achievement: younger > familiar Easy to use, like working with and enable health goal achievement: no differences between younger and older, and older and familiar Easier to manage health and useful: older > familiar

(continued on next page)

Table 6 (continued)

Clothing	Parmar et al. (2018)	Styling	Agent's attire, design virtual environment	<p>Experiment 1:</p> <ul style="list-style-type: none"> [AC] Professionalism <p>Experiment 2:</p> <ul style="list-style-type: none"> [AC] Professionalism [AC] Credibility information [AC] Trustworthiness [AC] Reassurance [AC] Caring [AC] Friendliness [AR] Comfortability to talk to [AR] Perception face-to-face interaction [IT] Willingness to continue [B] Commitment to choose a healthcare proxy 	<p>Experiment 1:</p> <ul style="list-style-type: none"> Pre-ranking 12 agents with different outfits and offices on professionalism Agents differing in 1) Attire: A) dressed to fit health care (professional) / B) not role-appropriate (casual) and 2) Environment: A) professional setting (medical examination room) and B) no setting (empty room) <p>Experiment 2, four conditions (attire and environment):</p> <ul style="list-style-type: none"> CASUAL EMPTY CASUAL FULL PROF-EMPTY PROF-FULL 	<p>Experiment 1:</p> <ul style="list-style-type: none"> Agent's attire: lab coat and stethoscope most professional, casual clothing without a stethoscope least professional Environment: office setting most professional, empty room least professional <p>Experiment 2:</p> <ul style="list-style-type: none"> Professionalism: white coat > casual and full office setting > empty room Credibility, trustworthiness, reassurance, caring, friendliness, comfortability to talk to, perception face-to-face conversation, willingness to continue: white coat > casual Commitment to choosing a healthcare proxy: white coat > casual Reassurance: full office > empty room Comfortability to talk to: white coat in empty room > casual in empty room
	Body shape	Nguyen & Masthoff (2007)	Socio-demographic, styling	Profession	<p>Experiment 1:</p> <ul style="list-style-type: none"> [AC] Perceived age, gender, profession, attractiveness, trustworthiness, expertise [AC] Preferred agent for learning about fitness programs and health benefits of exercise <p>Experiment 2:</p> <ul style="list-style-type: none"> [O] Credibility messages 	<p>Experiment 1:</p> <ul style="list-style-type: none"> Opinion on 16 images varying in age, gender, profession (doctors and sport instructors) and estimated attractiveness, trustworthiness and expertness <p>Experiment 2 (3 x 2 experiment):</p> <ul style="list-style-type: none"> Credibility agent (based on experiment 1): <ul style="list-style-type: none"> High credibility image Low credibility image No image Content message: <ul style="list-style-type: none"> Fitness programs Health benefits of exercise
Schmeil & Suggs (2014)		Styling	Representation health status: body shape, body animations and poses and skin texture	<ul style="list-style-type: none"> [B] Comparison pre-test and post-test: motivation to engage in more physical activity, intention to engage in more physical activity, motivation to engage in healthier diet, intention to engage in a healthier diet [IT] Motivation and intention to access the tool again 	<p>Four conditions:</p> <ul style="list-style-type: none"> (1) Agent tailored to current health status: a non-animated character (2) Two agents: one tailored to current health status, other shows possible future health status (both still poses) (3) Agent tailored to and models current health behavior (animated) (4) Two agents: one tailored to current health status and current behaviour, other shows possible future changes in health status (both animated) 	<ul style="list-style-type: none"> Motivation to engage in more physical activity: positive change for all conditions, largest in (4) Intention to engage in more physical activity: only change for (2) and (4), largest change in (2) Motivation to engage in healthier diet: positive change for (4) Intention to engage in healthier diet: positive change all conditions, largest in (4) Motivation to revisit the tool: no differences between conditions

(continued on next page)

Table 6 (continued)

Body shape	Skalski et al. (2007)	Styling	<p>Interactivity (interactive or not), source attractiveness (attractive or unattractive)</p>	<p>Manipulation check:</p> <ul style="list-style-type: none"> [AC] Attitude toward agent: source credibility (knowledgeable, intelligence, competence), likeable, warm, pleasing, friendly, attractiveness (appealing, attractive, nice looking) [AR] Perceived interactivity <p>Parameters in causal models:</p> <ul style="list-style-type: none"> Interactivity Perceived interactivity Social presence Source thoughts Message processing Attitude toward blood pressure <p>[B] Behavioural intention: sign up for actual appointment to test blood pressure</p>	<p>Four conditions:</p> <ul style="list-style-type: none"> Interactive (I) (ability to talk to source, control order in which message elements are presented), attractive (A) (pleasant-looking woman) Interactive (I), unattractive (UA) (generated by computer programs) Not interactive (NI) (source discussed categories one by one, no ability to select order), attractive (A) Not interactive (NI), unattractive (UA) 	<ul style="list-style-type: none"> Variable check: Attractiveness: A > UA Rest attitude variables: no difference A and UA Perceived interactivity: not high in both conditions, but I > NI Attractive source model: interactivity induction (+)--> perceived interactivity (+)--> social presence (+)--> positive source thoughts / message processing --> (+) attitude (all, but positive source thoughts were significant paths) Unattractive source model: interactivity induction (+)--> perceived interactivity (+) -->social presence --> (-) negative source thoughts / (+) message processing --> (- and +) attitude (all, but message processing were significant paths) Combined model: interactivity induction (+) --> perceived interactivity (+)--> social presence (+)--> positive source thoughts / message processing --> attitude (all, but path from source thoughts to attitude significant) Behavioral intention: 42% signed up for test
	van Vugt et al. (2006)	Styling	<p>Body shape:</p> <ul style="list-style-type: none"> Similarity between user and interface character Ideal (slim) and non-ideal body shapes (fat) 	<p>Dependent variables:</p> <ul style="list-style-type: none"> [AR] User involvement [AR] User distance [IT] Intention to use the agent <p>Co-variables:</p> <ul style="list-style-type: none"> [AC/UC] Similarity [AC] Valence [AC] Affordance [AC] Aesthetics [AC] Perceived realism [AC] Perceived ethics (trustworthiness and credibility) 	<p>A 2 (similarity: similar versus dissimilar body shapes) x 2 (valence: positive (slim) versus negative (fat)) factorial design</p>	<ul style="list-style-type: none"> Similarity, valence, perceived affordance and aesthetics: no effect on dependents Perceived ethics, perceived realism: effect on dependents Two-way interaction effects of similarity and ethics, and of valence and ethics on dependents Effect perceived realism on user involvement and user distance: the more agent perceived as realistic, the higher user involvement and the lower user distance Effect perceived ethics on user involvement and intention to use: the more agent perceived as ethically good, the higher user involvement and intention to use Effect perceived ethics on user distance and intention to use: similar > dissimilar Effect perceived ethics on user involvement: fat > slim Realism: fat > slim

(continued on next page)

Table 6 (continued)

Gender	Alsharbi & Richards (2017)	Socio-demographic	Age, gender, ethnicity, role	<ul style="list-style-type: none"> [AC] Pre-test: Preference age, gender, ethnicity and role [AC] Ranking task: Preferred agent out of 12 agents [AC/AR] Post-test: rapport preferred agent 	<ul style="list-style-type: none"> Pre-test: opinion on age (younger, older or same age), gender (female, male or no preference), ethnicity (same, other, no preference), role (teacher, doctor, parent or friend), looks and environmental context Ranking task: ranking 12 agents differing in gender (male or female), age (young or old) and ethnicity (Asia, white and Mediterranean) Post-test: rapport preferred agent 	<ul style="list-style-type: none"> Frequencies pre-test: age (0, 4, 7), gender (6, 1, 4), ethnicity (8, 1, 2), role (majority: friend) Ranking task: majority preferred young female Mediterranean, majority preferred the old male Asian the least. Post-test: agent was liked, agent was weird, experience was comfortable, interacting with character does not help to solve a problem
	Forlizzi et al. (2007): <i>See rendering style</i>					
	Malhotra et al. (2016)	Socio-demographic	Gender	<p>Comparison of levels of emotional dimensions of respondents with the known levels in literature. The emotional dimensions:</p> <ul style="list-style-type: none"> [AC] Evaluation (valence) [AC] Potency (power/dominance) [AC] Activity/arousal 	<ul style="list-style-type: none"> Six audio-visual prompts (different steps of handwashing task) with five different emotional deliveries by a male and female character 	<ul style="list-style-type: none"> Overall: consensus on E and P, some on A Male and female characters: consensus on E and P Female character: close to consensus on A Male respondents: consensus on E and P, close to consensus on A Female respondents: consensus on E Age group 65-69: consensus on E Age group 70+: consensus on P and A Highest overall match E and P: age group 65-69 Highest overall match A: female respondents
Age	Alsharbi & Richards (2017): <i>See rendering style</i>					
	Wissen et al. (2016): <i>See rendering style</i>					
Cultural background	Alsharbi & Richards (2017): <i>See rendering style</i>					
	Yin et al. (2010): <i>See speech and/or textual output, facial and gaze expressions and hand and body gestures</i>					
	Zhou et al. (2014)	Socio-demographic	Racial concordance, perceived similarity	<ul style="list-style-type: none"> [AC/UC] Similarity [AR] Working alliance [AC] Satisfaction [UX] Easiness to talk [IT] Desire to continue [AC] Likability [AC] Trust [AC] Closeness relationship [AC] Caring [IT] Likelihood of following advice [O] Pref. talking to doctor or agent 	<p>Four conditions:</p> <ul style="list-style-type: none"> A) Black agent, multiple interactions B) Black agent, single interaction C) White agent, multiple interactions D) White agent single interaction 	<ul style="list-style-type: none"> Perceived similarity unrelated to racial concordance Working alliance: racial concordance < discordant Concordance unrelated to any other measure Working alliance, satisfaction, desire to continue, liking agent, trust in agent, relationship with agent, caring, likelihood of following advice: high similarity > low similarity Higher similarity with agent: preferred receiving their discharge instructions from the agent
Cultural background	Zhou et al. (2017)	Socio-demographic	Cultural adaptation (Chinese vs American agent)	<p>Before and after interaction:</p> <ul style="list-style-type: none"> [B] Persuasion: attitude towards regular exercise [B] Decisional balance: one's relative weighing of the pros and cons of exercise <p>After interaction:</p> <ul style="list-style-type: none"> [UX] Satisfaction [AC] Attitudes toward agent 	<ul style="list-style-type: none"> The agent designs: <ul style="list-style-type: none"> English-speaking, animated American (A): US living room, more hand gestures Chinese (C): room with interior design elements traditional Chinese culture Cultural dimensions dialogues: <ul style="list-style-type: none"> (1) power distance. A: recommended, C: use of should/have (2) individualism vs collectivism. A: potential effect for individual, C: potential effect for family (3) uncertainty avoidance. A: step-by-step advice, C: year-long plan (4) long-term vs short-term orientation. A: focus on fun side of exercise, C: importance of persistence and long-term benefits (5) indulgence vs restraint Small talk: A: local sports teams, basketball, C: Chinese Lunar New Year, traditions, trending TV shows Exercises suggested: A: hiking, cycling, weightlifting, simple exercises, C) badminton, jogging, aerobic exercises 	<ul style="list-style-type: none"> Attitudinal change towards exercise: both positive, but American > Chinese Change decisional balance: no differences (both a significant increase) No significant interaction between condition and participants' acculturation level on either persuasion or decisional balance. Both conditions: satisfied with agent, liked agent, trusted agent, agent easy to use Feeling of how interesting agent is: near significant, Chinese > American No other differences
	Role	Alsharbi & Richards (2017): <i>See rendering style</i>				
Nguyen & Masthoff (2009): <i>See relational, empathic behaviour</i>						

Outcome variables: U = usage / IT = intention to (continue) using / B = (intention towards) behaviour change / UX = usability and user experience / AC = agent characteristics (e.g. demographics, personality, styling) / AR = relation with agent / UC = user characteristics / O = other

References

- Acosta, J.C., Ward, N.G., 2011. Achieving rapport with turn-by-turn, user-responsive emotional coloring. *Speech Communication* 53 (9-10), 1137–1148. <https://doi.org/10.1016/j.specom.2010.11.006>.
- Alsharbi, B., Richards, D., 2017. Using virtual reality technology to improve reality for young people with chronic health conditions. *Proceedings of the 9th International Conference on Computer and Automation Engineering*. ACM, pp. 11–15. <https://doi.org/10.1145/3057039.3057080>.
- Amini, R., Lisetti, C., Yasavur, U., 2014. Emotionally responsive virtual counselor for behavior-change health interventions. *International Conference on Design Science Research in Information Systems*. Springer, Cham, pp. 433–437. https://doi.org/10.1007/978-3-319-06701-8_40.
- Amini, R., Lisetti, C., Yasavur, U., Rishe, N., 2013. On-demand virtual health counselor for delivering behavior-change health interventions. *IEEE International Conference on Healthcare Informatics*. IEEE, pp. 46–55. <https://doi.org/10.1109/ICHI.2013.13>.
- Baylor, A.L., 2009. Promoting motivation with virtual agents and avatars: Role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1535), 3559–3565. <https://doi.org/10.1098/rstb.2009.0148>.
- Baylor, A.L., Kim, Y., 2004. Pedagogical agent design: the impact of agent realism, gender, ethnicity, and instructional role. *International Conference on Intelligent Tutoring Systems*. Springer, Berlin, Heidelberg, pp. 592–603. https://doi.org/10.1007/978-3-540-30139-4_56.
- Berry, D.C., Butler, L.T., De Rosis, F., 2005. Evaluating a realistic agent in an advice-giving task. *International Journal of Human Computer Studies* 63 (3), 304–327. <https://doi.org/10.1016/j.ijhcs.2005.03.006>.
- Bickmore, T., Gruber, A., Picard, R., 2005. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling* 59 (1), 21–30. <https://doi.org/10.1016/j.pec.2004.09.008>.
- Bickmore, T., Ring, L., 2010. Making it personal: end-user authoring of health narratives delivered by virtual agents. *International Conference on Intelligent Virtual Agents* 399–405. https://doi.org/10.1007/978-3-642-15892-6_43.
- Bickmore, T., Schulman, D., 2007. Practical approaches to comforting users with relational agents. *Proceedings of ACM CHI 2007: Conference on Human Factors in Computing Systems*. ACM, pp. 2291–2296. <https://doi.org/10.1145/1240866.1240996>.
- Bickmore, T., Schulman, D., Yin, L., 2009. Engagement vs. deceit: virtual humans with human autobiographies. *International Workshop on Intelligent Virtual Agents*. Springer, Berlin, Heidelberg, pp. 6–19. https://doi.org/10.1007/978-3-642-04380-2_4.
- Bickmore, T., Schulman, D., Yin, L., 2010. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence* 24 (6), 648–666. <https://doi.org/10.1080/08839514.2010.492259>.
- Bickmore, T.W., Carusob, L., Clough-Gorr, K., Heeren, T., 2005. 'It's just like you talk to a friend' relational agents for older adults. *Interacting with Computers* (17), 711–735. <https://doi.org/10.1016/j.intcom.2005.09.002>.
- Bickmore, T.W., Pfeifer, L.M., Jack, B.W., 2009. Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. *27th International Conference on Human Factors in Computing Systems*. ACM Press, pp. 1265–1274. <https://doi.org/10.1145/1518701.1518891>.
- Bickmore, T.W., Picard, R.W., 2004. Towards caring machines. *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. ACM, pp. 1489. <https://doi.org/10.1145/985921.986097>.
- Bickmore, T.W., Picard, R.W., 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction* 12 (2), 293–327. <https://doi.org/10.1145/1067860.1067867>.
- Clark Herbert, H., Brennan Susan, E., 1991. Grounding in communication. *Perspectives on Socially Shared Cognition* 13, 222–233. <https://doi.org/10.1037/10096-006>.
- Cowell, A.J., Stanney, K.M., 2003. Embodiment and interaction guidelines for designing credible, trustworthy embodied conversational agents. *International Workshop on Intelligent Virtual Agents*. Springer, Berlin, Heidelberg, pp. 301–309. https://doi.org/10.1007/978-3-540-39396-2_50.
- Creed, C., Beale, R., 2012. User interactions with an affective nutritional coach. *Interacting with Computers* 24 (5), 339–350. <https://doi.org/10.1016/j.intcom.2012.05.004>.
- Creed, C., Beale, R., Cowan, B., 2015. The impact of an embodied agent's emotional expressions over multiple interactions. *Interacting with Computers* 27 (2), 172–188. <https://doi.org/10.1093/iwct/064>.
- Dryer, D.C., 1999. Getting personal with computers: how to design personalities for agents. *Applied Artificial Intelligence*, 13 (3), 273–295. <https://doi.org/10.1080/088395199117423>.
- Forlizzi, J., Zimmerman, J., Mancuso, V., Kwak, S., 2007. How interface agents affect interaction between humans and computers. *Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces* 209–221. <https://doi.org/10.1145/1314161.1314180>.
- Frost, J., Boukris, N., Roelofsma, P., 2012. We like to move it move it! Motivation and parasocial interaction. *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. ACM, pp. 2465–2470. <https://doi.org/10.1145/2212776.2223820>.
- Grillon, H., Thalman, D., 2008. Eye contact as trigger for modification of virtual character behavior. *2008 Virtual Rehabilitation*, IWVR 205–211. <https://doi.org/10.1109/ICVR.2008.4625161>.
- Guadagno, R., Blascovich, J., Bailenson, J., McCall, C., 2007. Virtual humans and persuasion: the effects of agency and behavioral realism. *Media Psychology* 10, 1–22. doi: 10.108/15213260701300865
- Gulz, A., Haake, M., 2005. Social and visual style in virtual pedagogical agents. *Proceedings of The Workshop on Adaptation to Affective Factors, the 10th International Conference on User Modeling (UM2005)*. pp. 24–29.
- Jansen Kosterink, S., Vollenbroek Hutten, M.M., Hermens, H.J., 2016. A renewed framework for the evaluation of telemedicine. *eTELEMED 2016: The Eighth International Conference on eHealth, Telemedicine, and Social Medicine*. pp. 57–62.
- Kang, S.-H., Gratch, J., 2011. People like virtual counselors that highly-disclose about themselves. *Annual Review of Cybertherapy and Telemedicine* 167 (143-148).
- Kaptein, M., De Ruyter, B., Markopoulos, P., Aarts, E., 2012. Adaptive persuasive systems. *ACM Transactions on Interactive Intelligent Systems* 2 (2), 1–25. <https://doi.org/10.1145/2209310.2209313>.
- Khan, R., Angeli, A.D., 2009. The attractiveness stereotype in the evaluation of conversational agents. *IFIP Conference on Human-Computer Interaction*. Springer, Berlin, Heidelberg, pp. 85–97. https://doi.org/10.1007/978-3-642-03655-2_10.
- Khan, R.F., Sutcliffe, A., 2014. Attractive agents are more persuasive. *International Journal of Human-Computer Interaction* 30 (2), 142–150. <https://doi.org/10.1080/10447318.2013.839904>.
- Kim, Y., Baylor, A.L., Reed, G., 2003. The impact of image and voice of pedagogical agents. *E-Learn 2003 - World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Springer, Berlin, Heidelberg, pp. 2237–2240. https://doi.org/10.1007/978-3-540-30139-4_56.
- Kim, Y., Baylor, A.L., Shen, E., 2007. Pedagogical agents as learning companions: the impact of agent emotion and gender. *Journal of Computer Assisted Learning* 23 (3), 220–234. <https://doi.org/10.1111/j.1365-2729.2006.00210.x>.
- Kramer, L.L., ter Stal, S., Mulder, B.C., de Vet, E., van Velsen, L., Lifestyles, H., Kramer, L.L., 2019. Developing embodied conversational agents for coaching people towards a healthy lifestyle : a scoping review. *Journal of Medical Internet Research* (forthcoming/in press). <https://doi.org/10.2196/14058>.
- Laranjo, L., Dunn, A.G., Tong, H.L., Kocaballi, A.B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A.Y., Coiera, E., 2018. Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association* 25 (9), 1248–1258. <https://doi.org/10.1093/jamia/ocy072>.
- Lee, J.E.R., Nass, C., Brave, S.B., Morishima, Y., Nakajima, H., Yamada, R., 2007. The case for caring colearners: The effects of a computer-mediated colearner agent on trust and learning. *Journal of Communication* 57 (2), 183–204. <https://doi.org/10.1111/j.1460-2466.2007.00339.x>.
- Lee, Y.-H., Xiao, M., Wells, R.H., 2018. The effects of avatars age on older adults self-disclosure and trust. *Cyberpsychology, Behavior, and Social Networking* 21 (3), 173–178. <https://doi.org/10.1089/cyber.2017.0451>.
- Lisetti, C., Amini, R., Yasavur, U., Rishe, N., 2013. I can help you change! An empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems* 4 (4), 1–28. <https://doi.org/10.1145/2544103>.
- Malhotra, A., Hoey, J., König, A., van Vuuren, S., 2016. A study of elderly people's emotional understanding of prompts given by virtual humans. *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 13–16.
- Nguyen, H., Masthoff, J., 2007. Is it me or is it what I say? Source image and persuasion. *International Conference on Persuasive Technology*. Springer, Berlin, Heidelberg, pp. 231–242. https://doi.org/10.1007/978-3-540-77006-0_29.
- Nguyen, H., Masthoff, J., 2009. Designing empathic computers: the effect of multimodal empathic feedback using animated agent. *Proceedings of the 4th International Conference on Persuasive Technology*. ACM, pp. 1541948. <https://doi.org/10.1145/1541948.1541958>.
- Nijland, N., 2011. *Grounding eHealth*. University of Twente, Enschede, The Netherlands Ph.D. thesis. doi: 10.3990/1.9789036531337
- Olafsson, S., Kimani, E., Asadi, R., Bickmore, T., Science, I., 2017. That's a rap. *International Conference on Intelligent Virtual Agents*. Springer, Cham, pp. 325–334. https://doi.org/10.1007/978-3-319-67401-8_41.
- Parmar, D., Olafsson, S., Utami, D., Bickmore, T., 2018. Looking the part: the effect of attire and setting on perceptions of a virtual health counselor. *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, pp. 301–306. <https://doi.org/10.1145/3267851.3267915>.
- Pelachaud, C., 2009. Studies on gesture expressivity for a virtual agent. *Speech Communication* 51 (7), 630–639. <https://doi.org/10.1016/j.specom.2008.04.009>.
- Provoost, S., Lau, H.M., Ruwaard, J., Riper, H., 2017. Embodied conversational agents in clinical psychology: A scoping review. *Journal of Medical Internet Research* 19 (5). <https://doi.org/10.2196/jmir.6553>.
- von der Pütten, A.M., Krämer, N.C., Gratch, J., 2009. Who's there? Can a virtual agent really elicit social presence? *The 12th Annual International Workshop on Presence*. pp. 1–7. doi: 10.1.1.1020.291
- Ring, L., Utami, D., Bickmore, T., 2014. The right agent for the job? The effects of agent visual appearance on task domain. *International Conference on Intelligent Virtual Agents*. Springer, Cham, pp. 374–384. https://doi.org/10.1007/978-3-319-09767-1_49.
- Rist, T., André, E., Baldes, S., Gebhard, P., Klesen, M., Kipp, M., Rist, P., Schmitt, M., 2004. A Review of the Development of Embodied Presentation Agents and Their Application Fields. pp. 377–404. https://doi.org/10.1007/978-3-662-08373-4_16.
- Robertson, S., Solomon, R., Riedl, M., Gillespie, T.W., Chociemski, T., Master, V., 2015. The visual design and implementation of an embodied conversational agent in a shared decision-making context (eCoach). *Learning and Collaboration Technologies* 427–437. <https://doi.org/10.1007/978-3-319-20609-7>.
- Rosenberg-Kima, R.B., Baylor, A.L., Plant, E.A., Doerr, C.E., 2008. Interface agents as social models for female students: The effects of agent visual presence and

- appearance on female students' attitudes and beliefs. *Computers in Human Behavior* 24 (6), 2741–2756. <https://doi.org/10.1016/j.chb.2008.03.017>.
- Ruttkay, Z., Dormann, C., Noot, H., 2004. Embodied conversational agents on a common ground: A framework for design and evaluation. *From Brows to Trust: Evaluating Embodied Conversational Agents* 27–66. https://doi.org/10.1007/1-4020-2730-3_2.
- Schmeil, A., Suggs, S., 2014. "How am I doing?" - Personifying health through animated characters. *International Conference of Design, User Experience, and Usability*. Springer, Cham, pp. 91–102. https://doi.org/10.1007/978-3-319-07635-5_10.
- Scholten, M.R., Kelders, S.M., Van Gemert-Pijnen, J.E., 2017. Self-guided Web-based interventions: Scoping review on user needs and the potential of embodied conversational agents to address them. *Journal of Medical Internet Research* 19 (11), 1–19. <https://doi.org/10.2196/jmir.7351>.
- Silverman, B.G., Holmes, J., Kimmel, S., Branas, C., Ivins, D., 2001. Modeling emotion and behavior in animated personas to facilitate human behavior change: the case of the HEART-SENSE game. *Health Care Management Science* 4 (3), 213–228. <https://doi.org/10.1023/A:1011448916375>.
- Skalski, P., Tamborini, R., Skalski, P., 2007. The role of social presence in interactive agent-based persuasion. *Media Psychology* 10 (3), 385–413. <https://doi.org/10.1080/15213260701533102>.
- Straßmann, C., Krämer, N.C., 2017. A categorization of virtual agent appearances and a qualitative study on age-related user preferences. *International Conference on Intelligent Virtual Agents*. Springer, Cham, pp. 413–422. https://doi.org/10.1007/978-3-319-67401-8_51.
- Tielman, M.L., Neerincx, M.A., van Meggelen, M., Franken, I., 2017. How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. *Technology and Health Care* 25, 1081–1096. <https://doi.org/10.3233/THC-170899>.
- Vardoulakis, L.P., Ring, L., Barry, B., Sidner, C.L., Bickmore, T., 2012. Designing relational agents as long term social companions for older adults. *International Conference on Intelligent Virtual Agents*. Springer, Berlin, Heidelberg, pp. 289–302. https://doi.org/10.1007/978-3-642-33197-8_30.
- Veletsianos, G., 2010. Contextually relevant pedagogical agents: visual appearance, stereotypes, and first impressions and their impact on learning. *Computers and Education* 55 (2), 576–585. <https://doi.org/10.1016/j.compedu.2010.02.019>.
- Veletsianos, G., Miller, C., Doering, A., 2009. Enali: a research and design framework for virtual characters and pedagogical agents. *Journal of Educational Computing Research* 41 (2), 171–194. <https://doi.org/10.2190/EC.41.2.c>.
- van Vugt, H.C., Konijn, E.a., Hoorn, J.F., Veldhuis, J., 2006. Why fat interface characters are better e-health advisors. *Intelligent Virtual Agents, Proceedings* 4133, 1–13. https://doi.org/10.1007/11821830_1.
- Wissen, A.V., Vinkers, C., Halteren, A.V., 2016. Developing a virtual coach for chronic patients: a user study on the impact of similarity, familiarity and realism. *International Conference on Persuasive Technology*. Springer, Cham, pp. 263–275. https://doi.org/10.1007/978-3-319-31510-2_23.
- Yin, L., Bickmore, T., Cortés, D.E., 2010. The impact of linguistic and cultural congruity on persuasion by conversational agents. *International Conference on Intelligent Virtual Agents*. Springer, Berlin, Heidelberg, pp. 343–349. https://doi.org/10.1007/978-3-642-15892-6_36.
- Zhou, S., Bickmore, T., Paasche-Orlow, M., Jack, B., 2014. Agent-user concordance and satisfaction with a virtual hospital discharge nurse. In *International Conference on Intelligent Virtual Agents*. Springer, Cham, pp. 528–541. https://doi.org/10.1007/978-3-319-09767-1_63.
- Zhou, S., Zhang, Z., Bickmore, T., 2017. Adapting a persuasive conversational agent for the Chinese culture. *International Conference on Culture and Computing (Culture and Computing)*. IEEE, pp. 89–96. <https://doi.org/10.1109/Culture.and.Computing.2017.42>.