Computers in Human Behavior 121 (2021) 106787

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Full length article

Tailoring coaching strategies to users' motivation in a multi-agent health coaching application

Tessa Beinema [a,b,*], Harm op den Akker [a,b], Lex van Velsen [a,b], Hermie Hermens [a,b]

[a] eHealth Group, Roessingh Research and Development, Postbox 310, 7500 AH, Enschede, the Netherlands
[b] Biomedical Signals and Systems Group, Faculty EEMCS, University of Twente, Postbox 217, 7500 AE, Enschede, the Netherlands

ARTICLE INFO

Keywords:
Embodied conversational agents
Health behaviour change
Coaching strategies
Tailoring
Motivation

ABSTRACT

Embodied conversational agents are often included in health behaviour change applications as intelligent virtual coaches. A major challenge in their development is tailoring coaching dialogues to user profiles. Agents should collect information about the user and consequently adapt the strategy that guides their interactions.

Previous research discovered relations between users' motivation profiles and potential effective coaching strategies. In the current paper, we describe an experiment with multiple agents that tests if users with certain motivation profiles prefer certain (tailored) strategies.

Participants were classified into four motivation groups (Intrinsic Motivation, External Regulation, Dual Motivation, A-motivation), following their responses to a questionnaire on motivation towards healthy living. Then, two coaches suggested a positively and a negatively tailored strategy. Participants rated these and chose their favourite.

Results ($N = 108$) show that the Dual Motivation group appreciated their positively tailored strategy more than their negatively tailored strategy, while intrinsically motivated participants appreciated both strategies. Furthermore, agents' likeability does not seem to influence strategy appreciation, while there was an effect of participant's age and gender.

We conclude that coaching strategies for dialogues with agents can be tailored to personal motivation to live healthy. Future research should focus on performing a long-term study in a real-life setting.

1. Introduction

Embodied Conversational Agents (ECAs) are employed in many domains, ranging from customer service agents (Xu et al., 2017) to health coaches (van Wissen et al., 2016; Kramer et al., 2020) or digital healthcare professionals (Zhou, Bickmore, Paasche-Orlow, & Jack, 2014). ECAs are *"more or less autonomous and intelligent software entities with an embodiment used to communicate with the user"* (Ruttkay et al., 2004). While some ECA applications are designed for relatively short single session interactions (e.g., customer support agents), others aim to have the user interact with an agent in several sessions that are spread out over a longer period of time. Such long-term interaction is, for example, relevant for ECAs that are incorporated in health behaviour change applications (Bickmore et al., 2010, Bickmore, Trinh, Asadi, & Olafsson, 2018).

Health behaviour change applications are generally developed to assist a user in adopting a healthier lifestyle by adjusting modifiable behaviours, such as being physically active or making dietary choices (World Health Organization, 2018). With an ageing society and the increased pressure on healthcare providers this brings with it (World Health Organization, 2015, pp. 3–4), the always accessible, scalable and low-cost properties of such health applications make them an interesting candidate for provision of continous support and coaching. While changing behaviour might prevent or relieve health conditions, it has been shown that behaviour change, especially in the long-term, tends to be difficult (Bouton, 2014). Digital health applications can assist users in this process, but adherence to these applications and the support they provide can be a problem (Wangberg et al., 2008; Nijland, 2011).

The inclusion of ECAs in health behaviour change applications serves multiple purposes. One is that ECAs can make the use of a health application easier, more satisfying, and less frustrating (André & Pelachaud, 2010; Bickmore et al., 2016). Second, their potential to engage the user can help improve users' adherence and participation (Scholten et al., 2017; Bickmore et al., 2010) and can potentially increase an

* Corresponding author. Biomedical Signals and Systems Group, Faculty EEMCS, University of Twente, Postbox 217, 7500 AE, Enschede, the Netherlands.
E-mail address: t.c.beinema@utwente.nl (T. Beinema).

https://doi.org/10.1016/j.chb.2021.106787
Received 14 April 2020; Received in revised form 17 October 2020; Accepted 16 March 2021
Available online 31 March 2021

application's effect in changing the user's behaviour (Ma et al., 2019). ECAs can give the system social ability, which is an important factor for "establishing and maintaining a collaborative relationship between user and system" in such applications (Kamphorst, 2017). Third, an ECA's credibility, similarity, authority, power, and social attractiveness can all contribute to the persuasion of a user (Pickard, 2012, pp. 73–74), which can be useful when providing behaviour change support. Social attractiveness specifically can influence persuasion by enhancing the positive or negative perception of the message.

In a health behaviour change application, ECAs can take on the role of a virtual coach (Kramer et al., 2020). From counselling literature, we know that the quality of a working alliance between a councillor and client is a factor in therapeutic change and adherence (Castonguay et al., 2006) and that it has three key aspects, namely goal agreement, task agreement, and development of a personal bond (Horvath & Greenberg, 1989; Hatcher & Gillaspy, 2006). To ensure that the advice given and topics addressed by an ECA serve such long-term purposes as building up a working alliance and supporting a user's behaviour change, coaching dialogues with an ECA are guided by a *coaching strategy*. These coaching strategies are carefully designed by digitising existing interventions for the target population and domain, and combining those with communication, persuasive design and human-computer interaction principles, such as tailoring to the individual user (Krebs et al., 2010; Ryan et al., 2019; Wangberg et al., 2008).

A promising and novel approach for tailoring coaching content from the digital health (eHealth) field that seems suitable for tailoring a coaching strategy in our ECA application, is tailoring to the user's motivation towards a certain domain of healthy living. A recent study found that persuasive features can be tailored to a user's motivation to live healthy (van Velsen et al., 2019). We investigate whether this connection between the persuasive features and motivation holds when incorporated in a coaching setting with ECAs. If this is the case, the motivation to live healthy can enrich the user model in our ECA application, and the relation between motivation and persuasive features can be incorporated in the design and tailoring of the coaching strategies that our ECAs follow.

2. Theoretical background

2.1. Tailoring

An ECA's capability to communicate with a user, through natural language, is important in health coaching applications, as it provides an intuitive human-computer interaction (HCI) metaphor. Interactivity (Rafaeli, 1988) is generally acknowledged to be a natural attribute of human face-to-face communications, but is also an attribute that is fundamental in HCI concepts such as the computers-are-social-actors paradigm (CASA) (Nass et al., 1994), which states that people apply social rules and expectations to computers. Whether ECAs for behaviour change are built as (part of) eHealth systems, HCI applications, or persuasive technologies, they should not just interact, but they should engage the user. As stated by Bickmore et al. (2010): "Engagement is crucial, because it is typically a prerequisite for other system objectives: If a user stops interacting with a system, then it cannot have any further impact." Once the user is engaged however, they need to stay engaged and goal agreement, task agreement and development of a personal bond become important. Implementing the dialogues with the agent to follow a carefully designed strategy can fulfil these needs, but not all users have the same preferences or respond well to the same approach. Therefore, we investigate how ECA's high-level coaching strategies can be tailored.

Tailoring an application to its users has proven to be an effective approach in both embodied conversational agents (e.g., by adjusting non-verbal behaviour (Krämer et al., 2010)) and digital health applications (Krebs et al., 2010; Ryan et al., 2019; Wangberg et al., 2008). *Tailoring* can be seen as the adjustment of a communication's timing, intention, content and representation to a user (op den Akker et al., 2014). Communications can be tailored to many aspects of a user's personal profile, such as big five personality traits (de Vries et al., 2016), measured level of physical activity (Achterkamp et al., 2013), or susceptibility to persuasive strategies (Kaptein et al., 2012). For coaching specifically, Kamphorst (2017) states that it is important for a system that it "asks questions, gives feedback, and offers advice that is tailored to the individual user". There are many examples of tailoring approaches that can be used for this purpose. Examples include personalization, adaptation, content matching, feedback, inter-human interaction, goal setting, user targeting, context awareness, and self-learning (Dijkstra, 2008; Hawkins et al., 2008; op den Akker et al., 2014).

2.2. Tailoring of coaching dialogues and strategies

Coaching dialogues can be tailored on different levels, ranging from sentence level changes (e.g., using the user's first name), to deciding which combination of actions make up the best high-level coaching strategy to follow for a user (Beinema et al., 2018). In most ECA coaching applications tailoring is performed at the lower levels, and participants generally are all provided with the same high-level strategy or are being assigned personalised approaches by a human coach, which the agent then follows (e.g., Fadhil et al. (2019)). We focus on tailoring higher-level strategies in interactions with our agents and distinguish *social actions* and *coaching actions*. Social actions are used to build up e.g. trust and rapport between the user and an agent (Bickmore et al., 2005, p. 7) and can involve e.g. introductions, small talk or discussing background stories (Bickmore & Picard, 2005; Bickmore et al., 2009). Coaching actions are actions designed around behaviour change techniques (Michie et al., 2013) and persuasive features. For example, where one user might benefit from having an emphasis on actions that are focused on informing on health benefits, another might be best supported by receiving tips on which steps to take (e.g., as demonstrated in (Abdullah et al., 2018)); of course, the emphasis might shift over time and different types of actions can be mixed.

Tailoring of coaching dialogues and strategies is a process that involves research on multiple facets of health application development. As Paramythis et al. (2010) describe, each layer of an interactive adaptive system should carefully be designed and evaluated. A first aspect is the design of the user model, which should include information needed to make tailoring decisions. This information can come from various sources e.g. sensors, the user's responses in dialogues, build in knowledge or models on behaviour change and the domain in question. How to retrieve the relevant information for tailoring and when to update information (e.g., the user's stage of change or preferred type of physical activity) should be taken into account when designing the user model. Second, as previously mentioned, strategies should be carefully constructed based on available interventions and previous results, and available dialogues should be adjusted and extended to be able to execute these strategies. Tailoring of these strategies to a user's profile, in turn, then also needs to be fine-tuned and carefully evaluated; first for individual strategies and later for combinations of carefully balanced strategies. Ultimately, this evaluation needs to be performed for the intended result - improved health behaviour. But in the initial development phases, an evaluation of suitability and appreciation of strategies is essential to ensure that the content and strategies also have the potential to fulfil adherence and engagement requirements. The latter should be evaluated separately, but should always be preceded by a study that acknowledges the technical strategy.

2.3. Coaching strategies and motivation

Motivation is an integral element in changing (health) behaviour and a construct that returns in many theories of behaviour change. Examples range from classic Operant Conditioning (Miltenberger, 2008, p. 141),

the Information-Motivation-Behavioural Skills Model (Fisher & Fisher, 1992), Protection-Motivation Theory (Rogers, 1989; Norman et al., 2005, chap. 3), and the Transtheoretical Model (Prochaska & Velicer, 1997) to Self-Determination Theory (Ryan & Deci, 2000) and the Fogg Behaviour Model (Fogg, 2009). To take the Transtheoretical Model as an example, without (a change in) motivation people are less likely to move from one stage to a next stage, and are thus less likely to ultimately end up in the desired maintenance stage.

In health coaching applications, there has been some investigation into how users can be effectively motivated. Examples of approaches include the translation of a human-human approach such as motivational interviewing to the ECA setting (Olafsson et al., 2019) or sending motivational messages that are tailored to a user's personality (de Vries et al., 2016), and a recent scoping review studied which behaviour change techniques and persuasive system design principles have been reported in relation to motivation and adherence in weight loss applications (Asbjørnsen et al., 2019). Tailoring dialogues with ECAs can also have a positive effect on motivation, for example, as found for tailoring explanations to students' personal context and goals in order to change planning-behaviour for stress reduction (Abdulrahman & Richards, 2019). However, most studies either do not tailor their approach or do not tailor it to a user's motivation specifically. Recently, van Velsen et al. (2019) investigated if persuasive features can be tailored to a user's motivation to live healthy in the context of digital health (eHealth) applications.

In the present study, our ECAs present participants with a subset of two out of four coaching strategies based on persuasive features originally tested in the study by van Velsen et al. (2019). These four features are the following: *Self Goal Setting*, in which a user sets his or her own goal; *Health Education*, in which a user receives information about the benefits of healthy behaviour; *Showing Progress*, in which a user is provided with information on their progress towards their health goals; and *Implementation Intentions* in which a user defines when, where and how they will perform an activity. In the earlier study, these strategies were appreciated differently by participants with three types of motivation. These three types of motivation are *intrinsic motivation*, *external regulation*, and *a-motivation*. As originally defined in Self-Determination Theory (Ryan & Deci, 2000): intrinsic motivation is motivation to perform an action because performing the activity is rewarding in itself; external regulation is motivation to perform an action so that external requirements are fulfilled or a reward is gained; and a-motivation means that there is no intent to act. While most research has focused on one type of motivation per participant, there is evidence that users can have combined types of motivation (Gourlan et al., 2016). We therefore also include a Dual Motivation group based on the results from van Velsen et al. (2019), who found that participants can have both intrinsic motivation and external regulation.

2.4. Multi-agent coaching

Furthermore, most health coaching applications that use ECA technology, provide a single ECA as a coach who then provides coaching on one domain (e.g., for physical activity (Watson et al., 2012; King et al., 2017)). Health, however, is a construct that often includes multiple domains (World Health Organization, 2006; Huber et al., 2016). Thus, changing behaviour to lead a healthy lifestyle often requires a holistic approach. To that end, recent research has been investigating coaching on a combination of domains (e.g., Gardiner et al. (2017); Klaassen et al. (2018)) and even multiple coaches (op den Akker et al., 2018; Das et al., 2019; Hurmuz et al., 2020). Interaction with multiple agents at the same time provides opportunities for vicarious persuasion (Kantharaju, De Franco, Pease, & Pelachaud, 2018) and engagement (André & Rist, 2001). It also allows for a decentralised presentation of domain information, for example, by casting the agents as coaches that each have their own expertise, thus providing the possibility to include multiple viewpoints without an agent contradicting itself (Kantharaju et al., 2019). We therefore perform our experiment in a setting where multiple ECAs are present and interact with the user, so that the results can be incorporated in strategies for a broad range of ECA health coaching systems.

2.5. Research question and hypotheses

In the present paper, we present an explorative study in which participants interact with ECAs in a multi-agent setting. In this study, we investigate whether we can tailor coaching strategies presented by ECAs to users' motivation to live healthy, as a first step in extending methods for effectively coaching people to lead a healthy lifestyle. Specifically, we investigate if coaching strategies that are positively tailored to a participant's motivation profile are preferred over negatively tailored strategies.

Our underlying assumption is that a first prerequisite for developing well-tailored strategies is that these also need to be preferred by users and that a user's appreciation of a strategy will contribute to their engagement with the application. Therefore, evaluating whether a certain tailoring approach is appreciated by the target population is a first step towards developing well-tailored strategies that, supported by use of the application, have the potential to lead to long-term behaviour change.

This leads to the following research question:

RQ. *Can we tailor coaching strategies to a participant's motivation profile?*

To answer this question, we conducted an online experiment. In the experiment, participants interacted with a group of four ECAs, following a speech-bubble and reply-buttons paradigm. Two of these ECAs each presented a strategy; a positively tailored strategy and a negatively tailored strategy. This leads to the following hypothesis:

H1. *Participants appreciate the strategy that is positively tailored to their motivation profile more than the negatively tailored strategy, and consequently will choose that strategy.*

As stated above, participants in our experiment participated in a group conversation with multiple ECAs and two of these ECAs will present a strategy to the user. However, differences between agents might cause users to perceive them and the presented strategies differently. We therefore wanted to take possible influences of agent perception on strategy appreciation into account. Previous research has shown that user's perception of agents can influence their perception of a message (e.g., Schulman and Bickmore (2009); Ruijten et al. (2014)), the likeliness of them following an agent's advice (ter Stal, Tabak, op den Akker, Beinema, & Hermens, 2019), and that it can also have an influence on the answers that are given to survey questions (Kim et al., 2019). Research on human-human persuasion has shown that *source likeability* influences persuasiveness of a message (Chaiken, 1980) and for ECAs specifically, Pickard (2012, p. 74) defined likeability to be a dimension of an ECA's social attractiveness that influences the persuasiveness of an ECA's message. This leads to the following hypothesis:

H2. *The perceived likeability of the source (the coach suggesting the strategy) affects the participant's appreciation of that strategy.*

In addition to the questions on strategy appreciation and the likeability of the ECAs, we asked participants for their demographics. Various demographics have been shown to have an influence on engagement and appreciation for eHealth applications (Hardiker & Grant, 2011; Perski et al., 2017) and ECAs (e.g., Pezzullo et al. (2017); Payne et al. (2013); Krämer et al. (2010)). Level of education and living situation specifically were influences for the appreciation of the persuasive features on which we based our strategies (van Velsen et al., 2019). This leads us to our last hypothesis (an overview of our hypotheses can be found in Fig. 1):

H3. *A participant's demographics affect their appreciation of a strategy.*

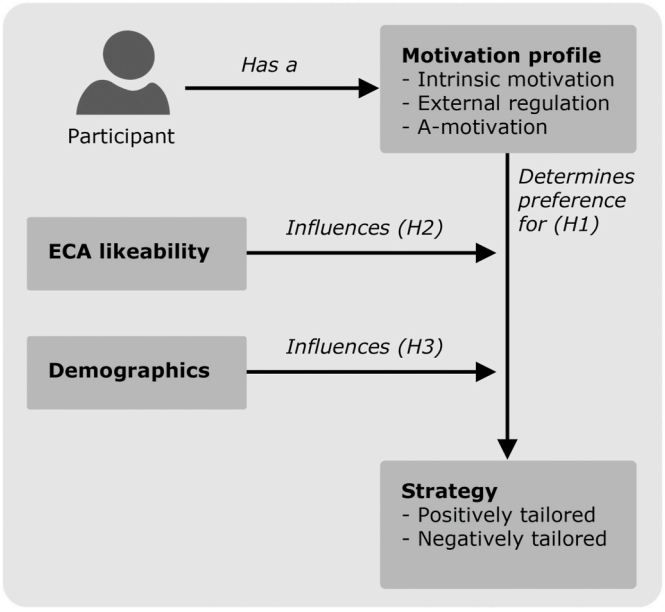

Fig. 1. Overview of the hypotheses.

3. Methods

We conducted an online experiment that consists of six steps, in which the participant interacted with our agent application. A schematic overview of these steps is depicted in Fig. 2, and they can be described as follows:

Step 1 The participant meets a virtual robot (named "Coda"), who in a one-on-one conversation introduces the experiment and poses the questions on demographics.
Step 2 Coda introduces the participant to the three coaches (Alexa, François, and Helen) in a group conversation. The participant gets to interact with the coaches and learn about the coaches' backgrounds (e.g., their origin and hobbies) and expertise (physical activity, nutrition, and cognition, respectively).
Step 3 The participant has a one-on-one interaction with coach Helen in which she asks them to respond to the statements from the questionnaire on personal motivation to live healthy (van Velsen et al., 2019). The results from this questionnaire are used in the background to classify the participant into one of our four motivation groups. This classification will determine the two strategies that are presented to the participant in step 5.
Step 4 While the coaches are 'deliberating', the participant has a one-on-one intermezzo with Coda. Coda asks the participant to rate the likeability of each of the coaches and he asks them to indicate their preference for a coach.
Step 5 After a short introduction by Coda and Helen (who asked the motivation questions), coaches Alexa and François both propose a coaching strategy. One strategy is the positively tailored strategy for the participant's motivation group, the other strategy is the negatively tailored strategy. The presentation of these two strategies is randomized over Alexa and François, as is the order in which they get to present their strategy.
Step 6 In this last step, Coda asks the participant to again rate the likeability of the coaches and to indicate their preference for a coach (as was done in step 4). He also asks participants to rate both strategies, to choose their preferred strategy from these two, and to rate the usability of the system (a control question to ensure that the system had a high enough usability to conduct the experiment).

The dialogues between the ECAs and the participant in these steps contain both social actions and coaching actions. That is, the introduction dialogue in step 2, involves social actions such as 'getting acquainted talk' and 'introductions with background stories' (Bickmore & Picard, 2005). Examples of coaching actions are, for example, the questions by coach Helen in step 3 (gathering relevant domain information to tailor advice) and Alexa's and François' presentation of the strategies (suggestion of a tailored coaching approach).

3.1. Implementation

The experiment was conducted using a fully-functional multi-party conversational agent system. The system's server provided user and dialogue management. The interface consisted of a webpage that showed the agents, their speech-bubbles, and the reply options. A screenshot of this multi-agent interface can be found in Fig. 3. The dialogues between the coaches and the user were specified using the WOOL Dialogue Framework (Roessingh Research and Development, 2020).

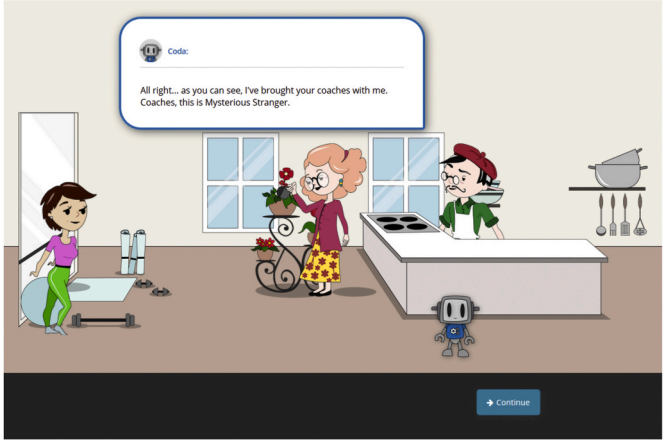

Fig. 3. A screenshot of the multi-agent interface. From left to right the agents shown are: Alexa (physical activity coach), Helen (cognitive coach), François (nutrition coach), and the assistant robot Coda in front of the counter.

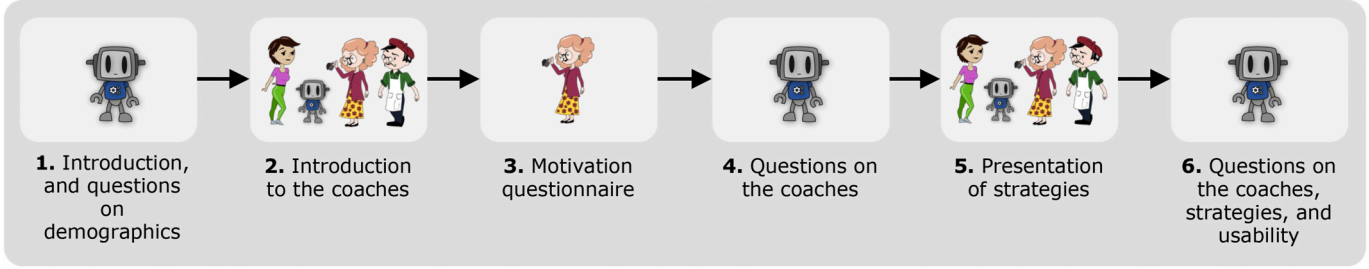

Fig. 2. Overview of the six steps that participants go through in the experiment.

3.2. Motivation group classification

The classification of participants into motivation groups was a three-step process. First, participants answered the questionnaire on motivation to live healthy (van Velsen et al. (2019), adapted from the revised Sports Motivation Scale (SMS-II) by Pelletier et al. (2013)). Coach Helen asked participants to indicate their agreement with these 11 statements on a seven-point Likert scale (ranging from Completely disagree (1) to Completely agree (7)).

Second, we created a motivation profile for each participant by calculating the normalised score for each of the three motivation *types* (intrinsic motivation, external regulation and a-motivation). These three normalised scores combined formed the participant's motivation *profile*.

Third, we used the motivation profile to classify participants into a motivation *group*. Four motivation groups were defined, namely: Intrinsic Motivation, External Regulation, A-motivation and Dual Motivation. We decided to add the Dual Motivation group since previous research shows that people may have combined types of motivation (Gourlan et al., 2016) and that people can be both intrinsically motivated and externally regulated (van Velsen et al., 2019). The specific classification rules can be found in Table 1. A threshold value of 0.2 was used for the classification into the Intrinsic Motivation and External Regulation groups.

3.3. Strategy design

For each of the motivation groups we designed a positively tailored and a negatively tailored strategy. The positively tailored strategies were based on persuasive features for which a motivation type had a large influence on their appreciation in the van Velsen et al. (2019) study. The negatively tailored strategies were based on persuasive features for which a motivation type had a small or no influence on their appreciation. For example, in the dialogue for the coaching strategy based on the 'Self Goal Setting'-feature the coach would propose a coaching approach in which they would help the participant to set a personal goal. The final selection of coaching strategies can be found in Table 2. We presented the Dual Motivation group with strategies suitable for both the intrinsic motivation and external regulation types. Since there were no significantly preferred persuasive features for the a-motivation type, but we wanted to include the A-motivation group for completeness, we present the A-motivation group with two random strategies. We selected these from the set of four strategies that were selected for the other three groups.

3.3.1. Coaching strategy dialogue creation

We converted the four persuasive features that we selected as a basis for our strategies into dialogues. In these dialogues, the virtual coach shortly explains the strategy, and proposes to the participant to follow it. The process of translating the user interface mock-ups into dialogues was as follows:

Table 1
The rules used for classifying participants into motivation groups based on the normalised scores (between 0 and 1) from their motivation profile. If the prerequisites for the first group (A-motivation) were not met, the prerequisites for the second group were checked, and so on.

Group	Classified as group when
A-motivation	A-motivation score is highest; else
Intrinsic Motivation	Intrinsic motivation score is higher than the external regulation score plus a threshold value of 0.2; else
External Regulation	External regulation score is higher than the intrinsic motivation score plus a threshold value of 0.2; else
Dual Motivation	Scores do not meet the prerequisites for the three other groups.

Table 2
The four motivation groups and the positively tailored and negatively tailored strategy presented for each group.

Group	Positively tailored strategy	Negatively tailored strategy
Intrinsic Motivation	Self Goal Setting	Health Education
External Regulation	Health Education	Implementation Intentions
Dual Motivation	Showing Progress	Implementation Intentions
A-motivation	*Random*	*Random*

1. Four researchers (with experience in writing motivational content) each wrote a short dialogue in which one of the strategies was presented.
2. One of these four writers examined the resulting four strategy dialogues and created one general dialogue structure suitable for presenting the strategies to a participant. The four dialogues from step 1 were then all adjusted to fit this structure.
3. Each of the four original writers reviewed all four of the resulting dialogues to verify that they provided a good representation of the strategy and were understandable by the study participants).

Furthermore, we ensured during this process that the coach presenting the strategy and the order in which the strategies were presented could be randomized. That is, since the two presenting coaches were experts in the physical activity and nutrition domain, we presented the strategies in the context of physical activity and nutrition. We also ensured that strategies could be presented both as a first or second suggestion. The resulting strategy dialogues can be found in Appendix A.

3.4. Measurements

In addition to the motivation questionnaire, we collected a number of parameters by means of questions that were posed by the agents in our experiment. These demographic, coach preference, coach likeability, strategy appreciation, and usability questions were asked by the robot agent, Coda.

3.4.1. Demographics

Each participant was asked for their age, gender, educational level (primary school, high school, vocational education, college, university), and living situation (with spouse, with friend/family member/other, alone).

Participants also indicated their self-reported physical activity level (not at all, not at all but thinking about beginning, less than 2.5h a week, more than 2.5h a week in the last six months, more than 2.5h a week for more than six months). Finally, participants answered the three questions (on a seven-point Likert scale) that make up the health literacy scale by Chew et al. (2004).

3.4.2. Coach preference and coach likeability rating

We asked participants two sets of questions about the coaches. The first set of questions asked participants to indicate their agreement with the statement "[COACH_NAME] is likeable." using a seven-point Likert scale (ranging from Completely disagree (1) to Completely agree (7)) (item taken from the scale used in Acosta and Ward (2011)). The second set of questions asked participants to indicate their first and their second preference for a coach. In this manner, each participant created a ranking for all three coaches. The questions were asked both before and after strategy presentation.

3.4.3. Strategy preference and strategy appreciation rating

We asked participants to indicate their agreement with the statement "This coaching approach would motivate me to lead a healthy lifestyle." on a seven-point Likert scale (ranging from Completely disagree (1) to Completely agree (7)). They were asked to do this for both of the strategies that were presented to them. In addition, we asked participants to

choose which of the two strategies that they were presented with was their preferred one.

3.4.4. Usability

The last question that participants answered was a one-question post-task usability questionnaire, formulated as: "How easy or difficult was it to use this system?" with answers on a seven-point Likert scale (ranging from Very difficult (1) to Very easy (7)) (Tedesco & Tullis, 2006). This question was added as a control question, since low usability of an application could distract from the application's content and might influence the ratings given to strategies and coaches.

3.5. Data analysis

The data collected in the experiment were stored in three models on the server in a NoSQL database (a user model, interaction model and questionnaire model). After completion of the experiment the data was extracted from the database as.json files, anonymized and converted to. csv using Python scripts. The resulting.csv-file was imported in the SPSS 25.0 statistics program, which was used for statistical analyses. Tests were performed using 95% confidence intervals.

We started testing with strategy appreciation ratings after checking the distribution of these ratings for normality (which they were for the four presented strategies, the chosen strategy, and the not-chosen strategy). We then checked the randomisation of the two strategies that were presented to the participant by performing a paired-samples t-tests. There was no significant difference in rating between the scores for the first strategy ($M = 4.92$, $SD = 1.64$) and the scores for the second strategy ($M = 4.85$, $SD = 1.59$) ($t = 0.368$, $p = .714$).

To compare strategy appreciation ratings between the chosen and not chosen strategy we performed paired-samples t-tests. The same method was applied when comparing the appreciation ratings for the positively tailored and negatively tailored strategy within motivation groups.

To test if one strategy was chosen more often than the other by participants in the strategy preference question, Binomial tests were conducted. The distribution of the number of choices for both strategies was compared to a 50-50 chance distribution.

To compare likeability ratings among the three coaches in the before condition, a repeated measures ANOVA was conducted. When Mauchly's test for Sphericity was significant ($\chi^2 = 10.335$, $p = .006$), a Huynh-Feldt correction was applied. Post hoc tests were conducted through paired-samples t-tests using a Holm-Bonferroni correction (resulting in significance levels set at 0.05, 0.025 and 0.017) to infer which coaches' ratings differed significantly. The same tests were performed for the likeability ratings in the after strategy presentation condition, but since Mauchly's test did not indicate any issue with sphericity (χ^2=0.905, $p = .636$) no correction was applied. Comparison of a coach's likeability rating before and after strategy presentation was performed using paired-samples t-tests.

Appreciation ratings for a strategy between two different presenting coaches were compared using an independent-samples t-test.

To determine the influence of demographics on strategy appreciation we conducted linear regression analyses including the age, gender, education, self-reported physical activity, health literacy, and living situation as main effects using the Backward method. The categorical variables were recoded using dummy variables with two levels (0, 1). The gender variable already had two levels (0, Male; 1, Female). The variables education, self-reported physical activity and health literacy were recoded to represent low and high scores. Specifically, education and health literacy were recoded into variables with two levels (0, 1) using their median and mean. This led to an 'other' (0) and 'university' (1) group, and lower (0) and higher (1) than 4.12 groups, respectively. Self-reported physical activity was recoded into 'active for less than 2.5h per week' (0) and 'active for more than 2.5h per week' (1). Living situation was recoded into 'alone' (0) and 'with partner/other' (1).

3.6. Recruitment

Participants were recruited through a Dutch panel of older adults who had indicated that they are interested in participating in eHealth research. In addition, we used a combination of snowball and convenience sampling (through social media, using flyers, and through personal connections). To be included, subjects had to be 18+, proficient in either Dutch or English, and they had to complete the full experiment.

3.7. Ethics

The performed online experiment does not require formal medical ethical approval according to Dutch law. Digital informed consent was obtained from each participant.

4. Results

4.1. Demographics

108 participants completed the experiment. 52 of them were male (48.1%) and 56 were female (51.9%). 67 completed the Dutch version (62.0%) and 41 the English version (38.0%). Their age ranged from 18 to 84 years ($M = 46.94$, $SD = 19.69$). Their educational background was primary school (0.9%), high school (13.9%), vocational education (7.4%), college (15.7%), and university (62.0%).

Self-reported health literacy scores were high, with a mean of 4.12 ($SD = 0.62$) on a five-point Likert scale. 13 participants reported 'not being physically active at all' (12.0%), 3 indicated 'not at all, but thinking about beginning' (2.8%), 31 participants answered 'less than 2.5h a week' (28.7%), 24 answered 'more than 2.5h a week in the last six months' (22.2%), and 37 indicated that they were physically active for 'more than 2.5h a week for more than six months (34.3%). Most participants lived with a spouse (66, 61.1%), some with a friend/family member/other (19, 17.6%), and others lived alone (23, 21.3%).

The distribution of the age, health literacy, language, education, self-reported physical activity, and living situation demographics was similar between the overall set of participants and within the motivation groups specifically. For gender however, the male-female ratio in the groups differed from the ratio in the full set of participants. The Intrinsic Motivation group had less males ($N = 20$) than females ($N = 36$), while the Dual Motivation group had more males ($N = 27$) than females ($N = 12$).

4.2. Usability

The ratings given by participants on the usability question were high in general ($M = 5.86$, $SD = 1.33$), as well as for the Intrinsic Motivation ($M = 5.95$, $SD = 1.41$) and Dual Motivation ($M = 5.67$, $SD = 1.31$) groups. We therefore assume that the usability of the application was sufficient to allow participants to focus on the interaction with the ECAs without being distracted by usability issues.

4.3. Motivation groups

Of the 108 participants, 56 (51.9%) were classified into the Intrinsic Motivation group, 3 (2.8%) into the External Regulation group, 39 (36.1%) into the Dual Motivation group and 10 (9.3%) into the A-motivation group. A scatterplot showing the normalised intrinsic motivation score and external regulation score for all participants can be found in Fig. 4. As can be seen, there are few participants who solely have a high score on external regulation. Furthermore, there are a few participants in the A-motivation group who have relatively high scores on intrinsic motivation and external regulation. On the other hand, some participants in the Dual Motivation group have quite low scores on

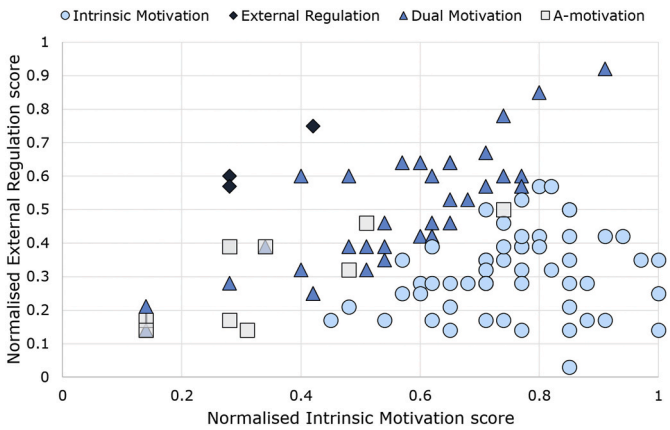

Fig. 4. A scatter plot showing the normalised score for the intrinsic motivation and external regulation types per participant. The marker shape indicates the motivation group.

intrinsic motivation and external regulation.

4.4. Strategy preference

We presented participants in the Intrinsic Motivation, External Regulation and Dual Motivation groups (98 out of 108 participants) with a positively tailored and a negatively tailored strategy. We asked these participants to a) rate both strategies and b) chose between those strategies.

4.4.1. Strategy appreciation rating

We tested the appreciation ratings given to the positively tailored and negatively tailored strategies within the Intrinsic Motivation and Dual Motivation groups (see Table 3). No significant difference in appreciation rating was found between the two strategies presented to the Intrinsic Motivation group, with the positively tailored Self Goal Setting strategy receiving similar ratings as the negatively tailored Health Education strategy. There was however a significant difference between the appreciation rating for the positively tailored and the negatively tailored strategy within the Dual Motivation group. The positively tailored Showing Progress strategy was appreciated more by participants than the negatively tailored Implementation Intentions strategy.

When comparing the appreciation rating for the strategy that participants chose with the appreciation rating for the not chosen strategy, we found a significant difference between the appreciation rating for the strategy that was chosen and the appreciation rating of the not chosen strategy (see Table 4). This was also the case for the chosen and not chosen strategy within the Intrinsic Motivation group, and the chosen and not chosen strategy within the Dual Motivation group.

We did not look into the appreciation ratings given by the A-motivation group in much detail since the group contained just 10 participants and they were presented with two random strategies, but we shortly report the distributions of the appreciation ratings to provide some insight. These were the following: Self Goal Setting was appreciated with a median of 5.00 ($N = 6$, $IQR = 4.75$–6.25); Health Education was appreciated with a median of 4.00 ($N = 5$, $IQR = 3.00$–5.50); Showing Progress was appreciated with a median of 5.00 ($N = 5$, $IQR = 4.50$–6.50); and Implementation Intentions was appreciated with a median of 5.50 ($N = 4$, $IQR = 5$–6).

Table 3
Results of the paired-samples t-test comparing the appreciation rating for the positively tailored strategy with the rating for the negatively tailored strategy for the Intrinsic Motivation and Dual Motivation groups.

Group	Pos. tailored	Neg. tailored	*t*	*p*
	M (*SD*)	*M* (*SD*)		
Intrinsic	5.02 (1.58)	5.05 (1.58)	0.152	.880
Dual	4.97 (1.71)	4.28 (1.65)	−2.610	.013

Table 4
Results of the paired-samples t-test comparing the appreciation rating for the chosen strategy with the rating for the not chosen strategy within all participants, and the Intrinsic Motivation and the Dual Motivation groups.

Group	Chosen	Not chosen	*t*	*p*
	M (*SD*)	*M* (*SD*)		
All	5.31 (1.48)	4.45 (1.62)	−5.543	<.001
Intrinsic	5.34 (1.49)	4.73 (1.60)	2.757	.008
Dual	5.18 (1.60)	4.08 (1.65)	4.882	<.001

4.4.2. Strategy choice

After indicating their appreciation for both presented strategies, participants were asked to choose one. Out of the 98 participants, 55 chose the strategy that was positively tailored (56.1%) and 43 chose the strategy that was negatively tailored (43.9%). This distribution did not indicate a significant preference for either strategy ($p = .266$). This was also the case within the Intrinsic Motivation (30 positively tailored, 26 negatively tailored, $p = .689$) and Dual Motivation (24 positively tailored, 15 negatively tailored, $p = .200$) groups specifically.

These results partly support our first hypothesis (*"Participants appreciate the strategy that is positively tailored to their motivation profile more than the negatively tailored strategy, and as a consequence will also choose that strategy."*). Participants who can be classified into the Dual Motivation group appreciate the strategy that is positively tailored to their motivation profile more than the negatively tailored strategy. However, the more appreciated strategy is not chosen more often. Within the Intrinsic Motivation group, individual participants appreciated their chosen strategy more than their not chosen strategy, but as a group they appreciated both strategies equally.

4.5. Coach preference

Before and after the presentation of the strategies, participants were asked to rate the three coaches on likeability and to choose their most preferred coach.

4.5.1. Coach likeability rating

Participants were asked to rate the likeability of the coaches before and after strategy presentation. The mean likeability ratings for all three coaches (Alexa, François, and Helen) can be found in Tables 5 and 6.

When comparing the ratings for the three coaches given before the strategies were presented, we found that they differed significantly ($F(1.860, 199.044) = 16.079$, $p<.001$). The post hoc tests revealed that there was a significant difference between the ratings for Alexa and François ($t = 2.302$, $p = .023$), and there was a significant difference between the ratings for François and Helen ($t = -5.482$, $p<.001$), and Alexa and Helen ($t = -3.698$, $p<.001$).

When comparing the ratings for the three coaches given after the strategies were presented, we found that they differed significantly (F(2,

Table 5
Likeability rating of the coaches before strategy presentation (the External Regulation group was omitted since *N* was 3).

Group	*N*	Alexa	François	Helen
		M (*SD*)	*M* (*SD*)	*M* (*SD*)
All	108	3.70 (1.00)	3.45 (1.05)	4.01 (0.92)
Intrinsic	56	3.98 (0.96)	3.63 (1.02)	4.21 (0.78)
Dual	39	3.41 (1.02)	3.28 (1.05)	3.79 (1.08)
A-motivation	10	3.30 (0.68)	3.60 (0.97)	3.60 (0.70)

Table 6
Likeability rating of the coaches after strategy presentation (the External Regulation group was omitted since *N* was 3).

Group	*N*	Alexa	François	Helen
		M (*SD*)	*M* (*SD*)	*M* (*SD*)
All	108	3.77 (0.93)	3.53 (1.04)	3.88 (0.96)
Intrinsic	56	3.98 (0.86)	3.70 (1.03)	4.00 (0.93)
Dual	39	3.62 (1.04)	3.38 (1.07)	3.79 (1.08)
A-motivation	10	3.20 (0.63)	3.30 (0.95)	3.50 (0.53)

214) = 7.501, $p = .001$). Post hoc tests revealed that there was no significant difference between the ratings for Alexa and Helen (t = -1.254, $p = .212$), but there was a significant difference between the ratings for Alexa and François ($t = 2.517$, $p = .013$), and François and Helen (t = -3.734, $p<.001$).

Comparing the before with the after rating for each coach, we found that for coach Helen there was a significant change in rating (t = -2.383, $p = .019$) in the overall set of participants. This was also the case in the Intrinsic Motivation group ($t = 2.271$, $p = .027$). In the Dual Motivation group there was only a significant change in rating between the before and after measurement for coach Alexa (t = -2.731, $p = .010$).

To summarise, the likeability of the two strategy presenting coaches (Alexa and François) was significantly different before strategy presentation and there was a significant difference in rating between the two presenting coaches after strategy presentation.

4.5.2. Coach choice

In addition, participants were asked to indicate their preferred coach both before and after strategy presentation. Before strategy presentation, 58 participants chose Helen (53.7%), while 30 participants chose Alexa (27.8%) and 20 participants chose François (18.5%). After strategy presentation this distribution was 50 (46.3%), 34 (31.5%) and 24 (22.2%), respectively. This matches the likeability ratings for the coaches.

4.6. Strategies and coach influence

To assess a possible influence of presenting coach on strategy appreciation, we compared the appreciation ratings for the strategies presented by coach Alexa with the ratings for the same strategies when presented by coach François. This was done within the Intrinsic Motivation and Dual Motivation groups. The distribution of ratings and results can be found in Table 7.

For strategies presented to the Intrinsic Motivation group, tests showed that there was no significant difference between the ratings given to the positively tailored Self Goal Setting strategy presented by Alexa as opposed to the same strategy presented by François. There was also no difference between the ratings for the negatively tailored Health Education strategy presented by Alexa and the ratings when François presented the strategy.

For strategies presented to the Dual Motivation group, tests showed that there was no significant difference between the ratings given to the positively tailored Showing Progress strategy presented by Alexa and the ratings for the same strategy presented by François. For the negatively tailored Implementation Intentions strategy the ratings when presented by Alexa and the ratings for the strategy when presented by François also did not differ significantly.

Finally, we compared the appreciation ratings for the chosen strategy and not chosen strategy between the two coaches. There was no difference between a chosen strategy presented by Alexa ($M = 5.25$, SD = 1.49) and a chosen strategy presented by François ($M = 5.41$, $SD = 1.48$) ($t = -0.546$, $p = .894$). There was also no difference in appreciation rating between a not chosen strategy presented by Alexa ($M = 4.72$, SD = 1.54) and a not chosen strategy presented by François ($M = 4.07$, SD = 1.68) ($t = 2.082$, $p = .283$).

Since the likeability of the two strategy presenting coaches did differ significantly before strategy presentation, we assume that there was not an 'equal starting point' when the participants were presented with the strategies. Even though the rating between the two coaches also differed after the presentation of the strategies, we found no significant difference in appreciation rating for a strategy presented by Alexa as opposed to a strategy presented by François. Therefore, we also assume that the difference in rating for the coaches did not influence the rating that participants gave to the strategies. These findings lead us to reject our second hypothesis in the context of this experiment (*"The perceived likeability of the source (the coach suggesting the strategy) affects the participant's appreciation of that strategy."*).

4.7. Strategies and demographics

To assess the influence of demographics on the appreciation of strategies we conducted a linear regression for the appreciation rating of the strategy that participants chose and appreciation rating of the strategy that they did not choose. We included the variables age, gender, education, self-reported physical activity, health literacy and living situation. We also conducted this test for each of our four strategies (Self Goal Setting and Health Education for the Intrinsic Motivation group, Showing Progress and Implementation Intentions for the Dual Motivation group). We only report significant results. The results (Table 8) show that older participants tend to appreciate their chosen strategy less than younger participants when observing the whole population and the Intrinsically Motivation group. In the Dual Motivation group, females tend to appreciate their chosen strategy more than males. Furthermore, older participants in the Intrinsic Motivation group tend to rate the tailored Self Goal Setting strategy lower than younger participants (Table 9).

These results support our third hypothesis (*"A participant's demographics affect their appreciation of a strategy."*) for the age and gender demographics.

5. Discussion

Tailoring coaching strategies to motivation would be a valuable method to incorporate in the design of user profiles and tailored coaching strategies for health coaching applications with ECAs. Previous research on health coaching applications has found tailoring to be

Table 7
Results of the independent-samples *t*-test comparing the appreciation rating for the positively and negatively tailored strategies between participants when Alexa was presenting a strategy with the same strategy being presented by François for the two motivation groups.

Group	Strategy	Alexa	François	*t*	*p*
		M (*SD*)	*M* (*SD*)		
Intrinsic	Self Goal Setting (positively tailored)	4.70 (1.73)	5.31 (1.39)	−1.452	.238
	Health Education (negatively tailored)	4.97 (1.52)	5.15 (1.66)	−0.430	.451
Dual	Showing Progress (positively tailored)	4.65 (1.79)	5.32 (1.60)	−1.224	.350
	Implementation Intentions (negatively tailored)	4.42 (1.90)	4.15 (1.42)	0.507	.149

Table 8
Results of regression analyses for the chosen strategy.

Group	Demographic	Beta	*t* (*df*)	*p*	R^2
All	Age	-.36	−3.97 (106)	<.001	.129
Intrinsic	Age	-.43	−3.51 (54)	.001	.186
Dual	Gender	.38	2.50 (37)	.017	.145

Table 9
Results of regression analyses for the Self Goal Setting strategy (positively tailored strategy for Intrinsic Motivation).

Group	Demographic	Beta	*t* (*df*)	*p*	R^2
Intrinsic	Age	-.31	−2.42 (54)	.019	0.098

effective in various ways (Krebs et al., 2010; Ryan et al., 2019; Wangberg et al., 2008), but tailoring coaching strategies for health coaching dialogues based on persuasive features to user's motivation to live healthy is a novel approach. The present study found that coaching strategies using persuasive features can be tailored to an individual's type of motivation to live healthy. Specifically, we found that participants in the Dual Motivation group (who are both intrinsically motivated and externally regulated) appreciated a positively tailored strategy more than a negatively tailored strategy, which is as hypothesised.

However, the Intrinsic Motivation group did not appreciate a positively tailored strategy over a negatively tailored strategy, but indicated similar ratings for both strategies in response to the statement "This coaching approach would motivate me to lead a healthy lifestyle". This effect might be explained by reflecting on our method of strategy construction. In a previous study by van Velsen et al. (2019), all persuasive features were appreciated to some degree by intrinsically motivated participants. We therefore selected a persuasive feature with a relatively low appreciation by participants with a high intrinsic motivation score for the construction of our negatively tailored strategy. Combined with the notion that we presented participants with dialogues based on the persuasive features, we conclude that intrinsically motivated participants seem to appreciate both strategies when presented by ECAs.

Furthermore, our experiment resulted in three secondary observations. First, the present study verifies conclusions from previous research that people can have multiple types of motivation (Gourlan et al., 2016; van Velsen et al., 2019), which was the main reason for including a Dual Motivation group. This is a recent insight in the health coaching domain and goes against the general assumption that people are either intrinsically motivated, externally regulated or a-motivated. We can therefore support the notion that inclusion of multiple types of motivation is necessary when defining a participant's motivation profile and that adaptation of coaching strategies to such combined profiles should be investigated further in future research.

Second, while we rejected our hypothesis on the influence of coach likeability on strategy appreciation, we did observe changes in coach likeability before and after strategy presentation. For coach Helen, her 'before strategies'-likeability was higher than her 'after strategies'-likeability (and higher than that of the other two coaches), while for the other two coaches (Alexa and François), this was the other way around. A possible explanation for this effect could be a difference in type or number of interactions between the coaches and the participants at two moments in the experiment.

In the experiment, the coaching dialogues included social and coaching actions. The social actions we included were, for example, 'getting acquainted talk' (Bickmore & Picard, 2005) and introductions with background stories (Bickmore et al., 2009). The social parts of the dialogues were designed to be present in similar amounts for all of the coaches. Coaching actions we included were, for example, asking the user for relevant information (the motivation questionnaire) and suggesting a tailored approach (the strategy presentation), as suggested for the development of effective e-coaching and persuasive design systems (Kamphorst, 2017; Oinas-Kukkonen & Harjumaa, 2009). For these coaching actions, there was a clear difference in interactions with the coaches in step 3 (Helen poses motivation questions) and step 5 (Alexa and François present the strategies) of the experiment. The timing of these steps, one before and the other after the first likeability measurement, and thus the difference in number of interactions could explain the differences on the perception of the respective coaches. Whether the found differences occurred solely due to the number of interactions in the various steps and participants' contribution of social values to these interactions (Nass et al., 1994), or if the perception of the coaches was also influenced by the content of these interactions is an interesting direction for future research.

Third and last, we found that there was an influence of participants' age and gender on the appreciation of the strategies. Namely, older participants rated their chosen strategy lower than younger participants in general, and female participants in the Dual Motivation group rated their chosen strategy higher than male participants in that group. These findings are in line with the trend towards more engagement from female participants that Perski et al. (2017) reported for health coaching applications, but seem to be in contrast with the trend towards more engagement from older participants that they report. The effect found for gender also seems to be in contrast with the reported effect by de Vries et al. (2017) that male participants rated certain types of motivational messages higher than female participants. Combined with the fact that living situation and level of education were found to be relevant factors in the study that inspired our strategy design (van Velsen et al., 2019), the inclusion of user-related parameters remains relevant for future research.

Future research performed in the context of our health coaching system will involve a long-term study. The study will let users interact with a group of ECAs over several weeks, while collecting a large amount of user parameters, interactions and contextual data. We hypothesise that from such data we can gain valuable insights to further fine-tune the tailoring of coaching dialogues. Furthermore, future research should also investigate how multiple coaching strategies could be combined, and how to balance these based on available knowledge about e.g. the user and the domain. This balancing might, for example, be done through selecting the topics that are discussed by a coach based on their relevance for certain strategies. With the further investigation of tailoring coaching strategies, extension of the user profile to include other user parameters than those included in the user model so far will also need to be considered.

5.1. Limitations

Although the study provided valuable insights, there were some limitations. First, the coaching strategies were presented to our participants by means of dialogues in an interactive multi-agent application. In these dialogues an ECA presents the strategy to the user, but the length of this interaction is limited. Thus, users are asked to indicate the appreciation for these strategies based on a limited experience. While this approach allowed us to explore the suitability of tailoring strategies to motivation for use in coaching dialogues with ECAs, the effects of long-term interaction using these strategies on e.g. health outcome should be investigated in future research.

Second, the size of the External Regulation group ($N = 3$) made it difficult to draw any specific conclusions with regards to this group. This is partly due to the introduction of our Dual Motivation group. There was a number of participants who had a normalised external regulation score that was higher than their normalised intrinsic motivation score, but the difference between these scores did not exceed the set threshold. We see this, however, as a confirmation that the introduction of a combined motivation group was necessary.

Furthermore, to thoroughly test the influence of a participant's impression of an ECA on the interpretation of the content, the inclusion of more parameters would be preferable (e.g., perceived expertise, authority, similarity to the participant, etc. (Pickard, 2012)). Because of the presence of multiple agents however, every additional question on the participant's opinion of the agents would have to be asked at least three times (once for each ECA) at two moments (before and after strategy presentation). Since the online experiment already contained quite a number of questions and we did not want participants to drop

out, we limited the number of questions about the agents.

6. Conclusions

Tailoring the high-level coaching strategy to a user of a health behaviour change application involves making choices that can have a major influence on the resulting dialogues between the virtual coach (an ECA) and a user. The finding that coaching strategies can be tailored to personal motivation to live healthy when it comes to coaching dialogues is therefore an important step in creating intelligent health coaching applications. Future research should investigate how to fine-tune the tailoring of these strategies based on long-term studies performed in more realistic settings, that is in the daily life of users, and should investigate how to combine and balance multiple strategies.

Declaration of competing interest

There is no conflict of interest with this paper.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation programme under Grant Agreement #769553 (the Council of Coaches project).

Appendix A

A Coaching strategy dialogues

This Appendix contains the text from the dialogues that were used to present the coaching strategies to the users. The presenting coach and order of presentatation were randomized for strategy one and two. The names of the coaches were inserted where it says '$StrategyOneCoach' and '$StrategyTwoCoach'. The strategy presentation started with a few introduction sentences, which were the following:

Coda: Coaches, now that you have finished your discussion: What's the plan?
Helen: Well, we have discussed and there are two coaching approaches that we think could be suitable.
Helen: We'd like to tell you about both. $StrategyOneCoach and $StrategyTwoCoach will both explain one.
User: Ok.
Helen: Helen: Good! Let's start with the first one … $StrategyOneCoach?

Then the first strategy was presented using the dialogue for that strategy (see subsections A.1-A.4). This was followed by a connecting comment by Helen:

Helen: Well said! And now the second coaching approach … $StrategyTwoCoach?

Which was followed by the presentation of the other strategy for the participants' motivation group by the other coach. Then there were some concluding remarks, namely:

Helen: Well, that's that! I hope you enjoyed talking to us. This is it for now, but we hope to see you again sometime.

Coda: I would like to ask you some final questions to finish the experiment.

A.1 Self Goal Setting

Coach: I suggest that we will support you in setting your own health goals.
Coach: For example, you can tell us how many steps you would like to take every day, or how many glasses of water you would like to drink.
(Option 1).

User: How many steps?
Coach: Well, the number of steps is just an example, you can also set as a goal to do exercises or go for walks a couple of times a week.

(Option 2).

User: How much water?
Coach: Drinking more water, eating less sugar, or eating more fruit … these are just examples. In any case, setting a healthy diet goal is important.

Coach: The most important thing is that you set a concrete goal for yourself. Then, together we will try to make sure that you stick to your goals!
Coach: We can help you choose a goal. After all, we are familiar with all the healthy activity lifestyle guidelines.
Coach: And we know a lot about healthy eating as well. But in the end, you have to set your own goals.
(Optional, can be skipped.)

User: But, why can't you set a goal for me?
Coach: If you set your own goals, chances are higher that you actually stick with them, compared to when we will try to commit you to something.

User: Okay, I think I understand?
Coach: Great! I hope this coaching approach will work well for you!

A.2 Health Education

Coach: I suggest we will provide you with tips and advice on healthy behaviour, and explain how this contributes to your health.
Coach: For example, why is it important to be physically active? And why is a healthy eating pattern good for you.

(Option 1).

User: Why is physical activity important?
Coach: Many people don't move enough. They say sitting is the new smoking. We will explain you what the benefits of a healthy active lifestyle are!

(Option 2).

User: Why is healthy eating important?
Coach: By eating healthy you make sure you are getting your required nutrients. We will explain in which way a healthy diet can contribute to your general health and you feeling fit.

Coach: We will start by having a conversation about your general health knowledge - just to see what you already know about healthy eating and activity.
Coach: We will also talk about your reasons for wanting to live a healthier life. That way, we can better tailor our information to your wishes and needs.
Coach: And of course, there is always the possibility of asking questions!

(Optional, can be skipped.)

User: But, why is all of this important?
Coach: It may not always be easy to adopt a healthy lifestyle. If you know exactly what the potential benefits are, this will help you in achieving your goals.

User: Okay, I think I understand?
Coach: Great! I hope this coaching approach will work well for you!

A.3 Showing Progress

Coach: I suggest we will regularly keep you up to date on the progress that you are making.

Coach: We could, for example, say that every day we will provide you some feedback on the number of steps that you've been walking, and how healthy you have been eating.

(Option 1).

User: How do you know how much I'm walking?
Coach: We have automatic step counters for that. If you agree to share this data, it should be no problem keeping track of your steps.

(Option 2).

User: How do you know what I'm eating?
Coach: We could keep track of that in a food diary, for example, but we'll figure that out when we get there.

Coach: By regularly checking your progress, you can choose to increase your set goals, or rather adjust them a little bit.

Coach: Let's imagine you want to walk for 10 km every week. We could then tell you every day how much you've already walked, and how far you still need to walk.

Coach: Or, if you decide to drink an additional 3 glasses of water every day, we can tell you how you're managing to do so over time.

(Optional, can be skipped.)

User: But, why is knowing this so important?
Coach: Insight! A healthy lifestyle starts with understanding your current behaviour. If we show you how you're doing, you can see how close you are to reaching your health goals.

User: Okay, I think I understand?
Coach: Great! I hope this coaching approach will work well for you!

A.4 Implementation Intentions

Coach: I suggest that you link your intentions for healthy behaviour or activities to specific moments in time.

Coach: For example, if you would like to go for regular short walks, you could select three different days of the week to do so.

(Option 1).

User: Why three days?
Coach: It could be fewer, or less. That is up to you.

(Option 2).

User: Should it be the same every week?
Coach: No, you can select different days every week.

Coach: The most important thing is that you define a schedule for yourself. Then, together we will try to make sure that you stick to it!

Coach: And of course, it's not just about physical activity. You can also schedule moments for healthier eating.

Coach: Or, if you decide to drink an additional 3 glasses of water every day, we can tell you how you're managing to do so over time.

(Optional, can be skipped.)

User: But, why do I have to schedule all that?
Coach: Well, by creating a schedule, you can select those moments that suit you best. If I don't plan, I know I have the tendency to postpone things.

User: Okay, I think I understand?
Coach: Great! I hope this coaching approach will work well for you!

References

Abdullah, A. S., Gaehde, S., & Bickmore, T. W. (2018). A tablet based embodied conversational agent to promote smoking cessation among veterans: A feasibility study. *Journal of Epidemiology and Global Health, 8*, 225–230. https://doi.org/10.2991/j.jegh.2018.08.104

Abdulrahman, A., & Richards, D. (2019). Modelling working alliance using user-aware explainable embodied conversational agents for behavior change: Framework and empirical evaluation. *40th International Conference on Information Systems, ICIS 2019*, 1–17.

Achterkamp, R., Cabrita, M., op den Akker, H., Hermens, H. J., & Vollenbroek-Hutten, M. M. (2013). 2013. Promoting a healthy lifestyle: Towards an improved personalized feedback approach. In *2013 IEEE 15th international conference on e-health networking, applications and services, healthcom* (pp. 725–727). https://doi.org/10.1109/HealthCom.2013.6720772

Acosta, J. C., & Ward, N. G. (2011). Achieving rapport with turn-by-turn, user-responsive emotional coloring. *Speech Communication, 53*, 1137–1148. https://doi.org/10.1016/j.specom.2010.11.006

op den Akker, H., Jones, V. M., & Hermens, H. J. (2014). Tailoring real-time physical activity coaching systems: A literature survey and model. *User Modeling and User-Adapted Interaction, 24*, 351–392. https://doi.org/10.1007/s11257-014-9146-y

op den Akker, H., op den Akker, R., Beinema, T., Banos, O., Heylen, D., Pease, A., … Hermens, H. (2018). Council of coaches: A novel holistic behavior change coaching approach. In *Proceedings of the 4th international conference on information and communication technologies for ageing well and e-health (ICT4AgeingWell2018)* (pp. 978–989). Scitepress.

André, E., & Pelachaud, C. (2010). Interacting with embodied conversational agents. *Speech Technology: Theory and Applications*, 123–149. https://doi.org/10.1007/978-0-387-73819-2_8

André, E., & Rist, T. (2001). Presenting through performing: On the use of multiple lifelike characters in knowledge-based presentation systems. *Knowledge-Based Systems, 14*, 3–13. https://doi.org/10.1016/S0950-7051(00)00096-4

Asbjørnsen, R. A., Smedsrød, M. L., Nes, L. S., Wentzel, J., Varsi, C., Hjelmesæth, J., & van Gemert-Pijnen, J. E. (2019). Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: Scoping review. *Journal of Medical Internet Research, 21*. https://doi.org/10.2196/14265

Beinema, T., op den Akker, H., & Hermens, H. (2018). Creating an artificial coaching engine for multi-domain conversational coaches in eHealth applications. In E. André, T. W. Bickmore, S. Vrochidis, & L. Wanner (Eds.), *Proceedings of the AAMAS workshop on intelligent conversation agents in home and geriatric care applications co-located with the federated AI meeting (FAIM 2018), CEUR* (pp. 35–39).

Bickmore, T. W., Gruber, A., & Picard, R. (2005). Establishing the computer - patient working alliance in automated health behavior change interventions. *Patient Education and Counseling, 59*, 21–30. https://doi.org/10.1016/j.pec.2004.09.008

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. In *ACM transactions on computer-human interaction* (pp. 293–327). ACM.

Bickmore, T., Schulman, D., & Yin, L. (2009). Engagement vs. deceit: Virtual humans with human autobiographies. In *International workshop on intelligent virtual agents* (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*, 648–666. https://doi.org/10.1080/08839514.2010.492259

Bickmore, T., Trinh, H., Asadi, R., & Olafsson, S. (2018). Safety first: Conversational agents for health care. *Studies in Conversational UX Design, 33*. https://doi.org/10.1007/978-3-319-95579-7_3

Bickmore, T. W., Utami, D., Matsuyama, R., & Paasche-Orlow, M. K. (2016). Improving access to online health information with conversational agents: A randomized controlled experiment. *Journal of Medical Internet Research, 18*, e1. https://doi.org/10.2196/jmir.5239

Bouton, M. E. (2014). Why behavior change is difficult to sustain. *Preventive Medicine, 68*, 29–36. https://doi.org/10.1016/j.ypmed.2014.06.010

Castonguay, L. G., Constantino, M. J., & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy: Theory, Research, Practice, Training, 43*, 271–279. https://doi.org/10.1037/0033-3204.43.3.271

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Chew, L. D., Bradley, K. A., & Boyko, E. J. (2004). Brief questions to identify patients with inadequate health literacy. *Family Medicine, 36*, 588–594.

Das, K. S. J., Beinema, T., op den Akker, H., & Hermens, H. (2019). Generation of multi-party dialogues among embodied conversational agents to promote active living and healthy diet for subjects suffering from type 2 diabetes. In *ICT4AWE 2019 - proceedings of the 5th international conference on information and communication technologies for ageing well and e-health* (pp. 297–304). https://doi.org/10.5220/0007750602970304

Dijkstra, A. (2008). The psychology of tailoring-ingredients in computer-tailored persuasion. *Social and Personality Psychology Compass, 2*, 765–784.

Fadhil, A., Wang, Y., & Reiterer, H. (2019). Assistive conversational agent for health coaching: A validation study. *Methods of Information in Medicine, 58*, 9–23. https://doi.org/10.1055/s-0039-1688757

Fisher, J. D., & Fisher, W. A. (1992). Changing AIDS-risk behavior. *Psychological Bulletin, 111*, 455–474. https://doi.org/10.1037/0033-2909.111.3.455

Fogg, B. (2009). A behavior model for persuasive design. In *Proceedings of the 4th international conference on persuasive technology - persuasive '09*. https://doi.org/10.1145/1541948, 1541999 http://portal.acm.org/citation.cfm?doid=1541948.1541999.

Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., Jack, B. W., & Bickmore, T. W. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient Education and Counseling, 100*, 1720–1729. https://doi.org/10.1016/j.pec.2017.04.015

Gourlan, M., Trouilloud, D., & Boiché, J. (2016). Motivational profiles for physical activity practice in adults with type 2 diabetes: A self-determination theory perspective. *Behavioral Medicine, 42*, 227–237. https://doi.org/10.1080/08964289.2014.1001810

Hardiker, N. R., & Grant, M. J. (2011). Factors that influence public engagement with eHealth: A literature review. *International Journal of Medical Informatics, 80*, 1–12. https://doi.org/10.1016/j.ijmedinf.2010.10.017

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16*, 12–25. https://doi.org/10.1080/10503300500352500

Hawkins, R. P., Kreuter, M., Resnicow, K., Fishbein, M., & Dijkstra, A. (2008). Understanding tailoring in communicating about health. *Health Education Research, 23*, 454–466. https://doi.org/10.1093/her/cyn004

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *Journal of Counseling Psychology, 36*, 223.

Huber, M., Van Vliet, M., Giezenberg, M., Winkens, B., Heerkens, Y., Dagnelie, P. C., & Knottnerus, J. A. (2016). Towards a 'patient-centred' operationalisation of the new dynamic concept of health: A mixed methods study. *BMJ Open, 6*, 1–11. https://doi.org/10.1136/bmjopen-2015-010091

Hurmuz, M. Z. M., Jansen-Kosterink, S. M., op den Akker, H., & Hermens, H. J. (2020). User experience and potential health effects of a conversational agent-based electronic health intervention: Protocol for an observational cohort study. *JMIR Research Protocols, 9*. https://doi.org/10.2196/16641

Kamphorst, B. A. (2017). E-coaching systems: What they are, and what they aren't. *Personal and Ubiquitous Computing, 21*, 625–632. https://doi.org/10.1007/s00779-017-1020-6

Kantharaju, R. B., De Franco, D., Pease, A., & Pelachaud, C. (2018). Is two better than one? Effects of multiple agents on user persuasion. In *Proceedings of the 18th international conference on intelligent virtual agents (IVA2018)* (pp. 255–262). ACM. http://arxiv.org/abs/1904.05248.

Kantharaju, R. B., Pease, A., Reidsma, D., Pelachaud, C., Snaith, M., Bruijnes, M., … op den Akker, H. (2019). Integrating argumentation with social conversation between multiple virtual coaches. In *Proceedings of the 19th ACM international conference on intelligent virtual agents (IVA2019)*. ACM.

Kaptein, M., de Ruyter, B., Markopoulos, P., & Aarts, E. (2012). Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems, 2*. https://doi.org/10.1145/2209310.2209313

Kim, S., Lee, J., & Gweon, G. (2019). Comparing data from chatbot and web surveys effects of platform and conversational style on survey response quality. *Conference on Human Factors in Computing Systems - Proceedings*, 1–12. https://doi.org/10.1145/3290605.3300316

King, A. C., Campero, I., Sheats, J. L., Castro Sweet, C. M., Garcia, D., Chazaro, A., Blanco, G., Hauser, M., Fierros, F., Ahn, D. K., Diaz, J., Done, M., Fernandez, J., & Bickmore, T. (2017). Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: The COMPASS trial design, methods, and baseline characteristics. *Contemporary Clinical Trials, 61*, 115–125. https://doi.org/10.1016/j.cct.2017.07.020

Klaassen, R., Bul, K. C., op den Akker, R., van der Burg, G. J., Kato, P. M., & Di Bitonto, P. (2018). Design and evaluation of a pervasive coaching and gamification platform for young diabetes patients. *Sensors, 18*, 1–27. https://doi.org/10.3390/s18020402

Krämer, N. C., Hoffmann, L., & Kopp, S. (2010). Know your users! Empirical results for tailoring an agent's nonverbal behavior to different user groups. In *International conference on intelligent virtual agents* (pp. 468–474). https://doi.org/10.1007/978-3-642-15892-6_50

Kramer, L. L., ter Stal, S., Mulder, B. C., de Vet, E., & van Velsen, L. (2020). Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *Journal of Medical Internet Research, 22*, Article e14058. https://doi.org/10.2196/14058

Krebs, P., Prochaska, J. O., & Rossi, J. S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. *Preventive Medicine, 51*, 214–221. https://doi.org/10.1016/j.ypmed.2010.06.004

Ma, T., Sharifi, H., & Chattopadhyay, D. (2019). Virtual humans in health-related interventions: A meta-analysis. In *Conference on human factors in computing systems - proceedings* (pp. 1–6).

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine, 46*, 81–95. https://doi.org/10.1007/s12160-013-9486-6

Miltenberger, R. G. (2008). *Behaviour modification: Principles and practices*.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computer are social actors. In *Conference on human factors in computing systems - proceedings* (pp. 72–78). https://doi.org/10.1145/259963.260288

Nijland, N. (2011). *Grounding eHealth - towards a holistic framework for sustainable eHealth technologies*. Ph.D. thesis. University of Twente. https://doi.org/10.3990/1.9789036531337

Norman, P., Boer, H., & Seydel, E. R. (2005). Protection motivation theory. In *Predicting health behaviour: Research and practice with social cognition models* (pp. 81–126). Maidenhead: Open University Press.

Oinas-Kukkonen, H., & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems, 24*, 485–500.

Olafsson, S., O'Leary, T., & Bickmore, T. (2019). Coerced change-talk with conversational agents promotes confidence in behavior change. *ACM International Conference Proceeding Series, 31*. https://doi.org/10.1145/3329189.3329202

Paramythis, A., Weibelzahl, S., & Masthoff, J. (2010). Layered evaluation of interactive adaptive systems: Framework and formative methods. *User Modeling and User-Adapted Interaction, 20*, 383–453. https://doi.org/10.1007/s11257-010-9082-4

Payne, J., Szymkowiak, A., Robertson, P., & Johnson, G. (2013). Gendering the machine: Preferred virtual assistant gender and realism in self-service. In *International workshop in intelligent virtual agents* (pp. 106–115). https://doi.org/10.1007/978-3-642-40415-3_9

Pelletier, L. G., Rocchi, M. A., Vallerand, R. J., Deci, E. L., & Ryan, R. M. (2013). Validation of the revised sport motivation scale (SMS-II). *Psychology of Sport and Exercise, 14*, 329–341. https://doi.org/10.1016/j.psychsport.2012.12.002

Perski, O., Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine, 7*, 254–267. https://doi.org/10.1007/s13142-016-0453-1

Pezzullo, L. G., Wiggins, J. B., Frankosky, M. H., Min, W., Boyer, K. E., Mott, B. W., Wiebe, E. N., & Lester, J. C. (2017). "Thanks Alisha, keep in touch": Gender effects and engagement with virtual learning companions. In *International conference on artificial intelligence in education* (pp. 299–310). https://doi.org/10.1007/978-3-319-61425-0_25

Pickard, M. D. (2012). *Persuasive embodied agents: Using embodied agents to change people's behavior, beliefs, and assessments*. Ph.D. thesis. University of Arizona.

Prochaska, J. O., & Velicer, W. F. (1997). *The transtheoretical model of health behaviour change*. https://doi.org/10.4278/0890-1171-12.1.38

Rafaeli, S. (1988). Interactivity: From new media to communication. In *Sage annual review of communication research: Advancing communication science* (Vol. 16, pp. 110–134).

Roessingh Research and Development. (2020 March). *The WOOL Dialogue Platform*. Roessingh Research and Development Accessed: http://wwww.woolplatform.eu.

Rogers, R. W. (1989). Cognitive and psychological processes in fear appeals and attitude change: A revised theory of protection motivation. In J. Cacioppo (Ed.), *Social psychophysiology: A sourcebook* (pp. 153–176). Guildford Press.

Ruijten, P. A., Ham, J., & Midden, C. J. (2014). Investigating the influence of social exclusion on persuasion by a virtual agent. In *International conference on persuasive technology* (pp. 191–200). https://doi.org/10.1007/978-3-319-07127-5_17

Ruttkay, Z., Dormann, C., & Noot, H. (2004). Embodied conversational agents on a common ground: A framework for design and evaluation. In *From brows to trust: Evaluating embodied conversational agents* (pp. 27–66). Kluwer.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation. *American Psychologist, 55*, 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, K., Dockray, S., & Linehan, C. (2019). A systematic review of tailored eHealth interventions for weight loss. *Digital Health, 5*, 1–29. https://doi.org/10.1177/2055207619826685

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*. https://doi.org/10.2196/jmir.7351

Schulman, D., & Bickmore, T. (2009). Persuading users through counseling dialogue with a conversational agent. *Proceedings of the 4th international conference on persuasive technology, 350*. https://doi.org/10.1145/1541948.1541983

ter Stal, S., Tabak, M., op den Akker, H., Beinema, T., & Hermens, H. (2019). Who do you prefer? The effect of age, gender and role on users' first impressions of embodied conversational agents in eHealth. *International Journal of Human-Computer Interaction, 36*, 881–892. https://doi.org/10.1080/10447318.2019.1699744

Tedesco, D. P., & Tullis, T. S. (2006). A comparison of methods for eliciting post-task subjective ratings in usability testing. In *Usability professionals association* (pp. 1–9). UPA).

van Velsen, L., Broekhuis, M., Jansen-Kosterink, S., & op den Akker, H. (2019). Tailoring persuasive eHealth strategies for older adults on the basis of personal motivation: An online survey. *Journal of Medical Internet Research, 21*, 1–16. https://doi.org/10.2196/11759

de Vries, R. A., Truong, K. P., Kwint, S., Drossaert, C. H., & Evers, V. (2016). Crowd-designed motivation: Motivational messages for exercise adherence based on behavior change theory. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 297–308). ACM. https://doi.org/10.1145/2858036.2858229.

de Vries, R. A., Truong, K. P., Zaga, C., Li, J., & Evers, V. (2017). A word of advice: How to tailor motivational text messages based on behavior change theory to personality and gender. *Personal and Ubiquitous Computing, 21*, 675–687. https://doi.org/10.1007/s00779-017-1025-1

Wangberg, S. C., Bergmo Trine, S., & Johnson, J. A. K. (2008). Adherence in internet-based interventions. *Patient Preference and Adherence, 2*, 57–66.

Watson, A., Bickmore, T., Cange, A., Kulshreshtha, A., & Kvedar, J. (2012). An internet-based virtual coach to promote physical activity adherence in overweight adults: Randomized controlled trial. *Journal of Medical Internet Research, 14*. https://doi.org/10.2196/jmir.1629

van Wissen, A., Vinkers, C., & van Halteren, A. (2016). Developing a virtual coach for chronic patients: A user study on the impact of similarity, familiarity and realism. In A. Meschtscherjakow, B. De Ruyter, V. Fuchsberger, M. Murer, & M. Tscheligi (Eds.), *International conference on persuasive technology* (pp. 263–275). Cham: Springer. https://doi.org/10.1016/B978-0-12-373932-2.00008-9.

World Health Organization. (2006). *Constitution of the world health organization. Technical report*.

World Health Organization. (2015). *World report on ageing and health. Technical Report*.

World Health Organization. (2018). *Noncommunicable diseases fact sheet* Accessed: March 2020 http://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In *Conference on human factors in computing systems - proceedings 2017-may* (pp. 3506–3510). https://doi.org/10.1145/3025453.3025496

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* (pp. 528–541). https://doi.org/10.1007/978-3-319-09767-1_63