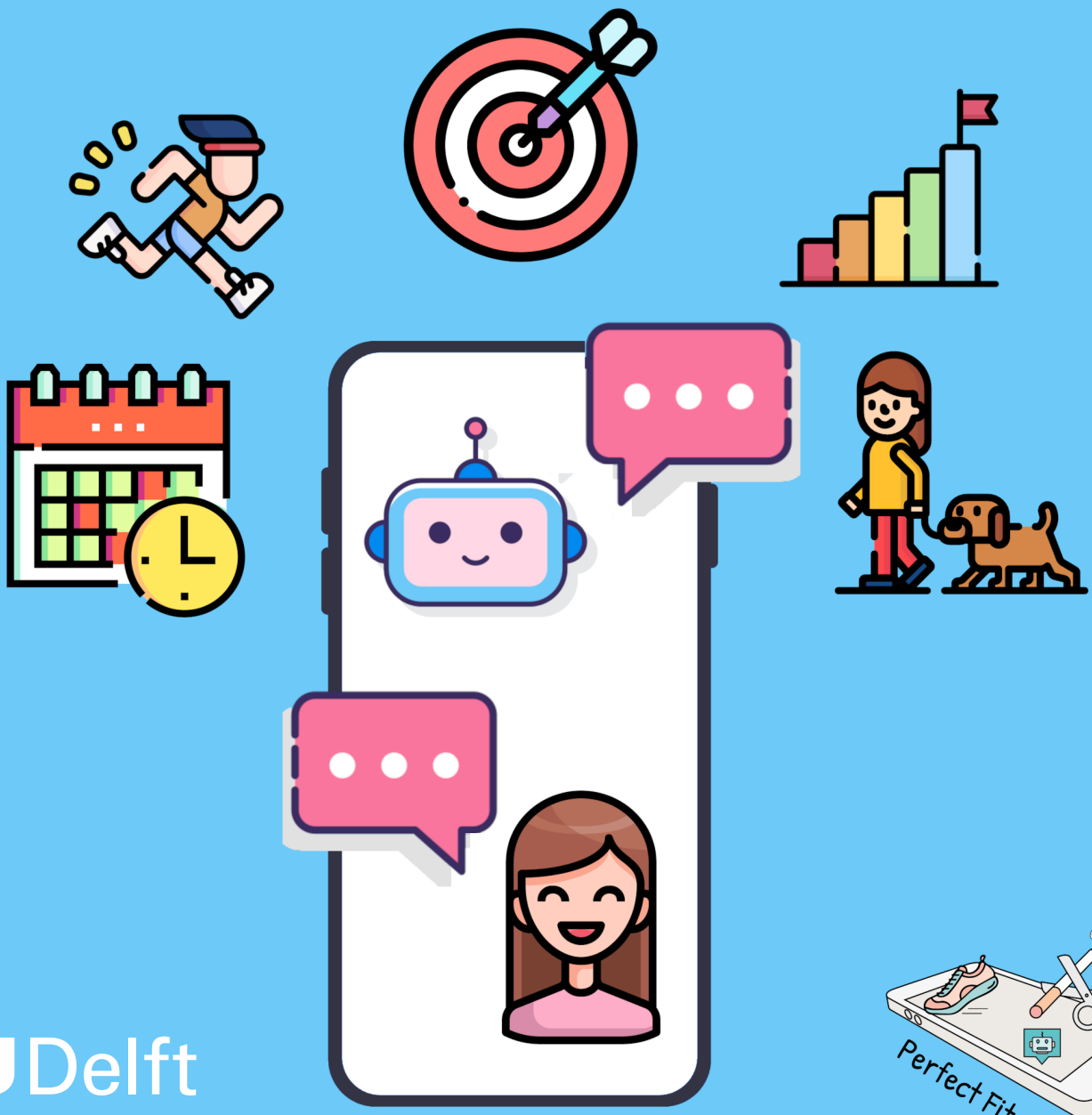
Goal-setting dialogue for physical activity with a virtual coach

Beyza Hizli



Delft University of Technology

TU Delft

Goal-setting dialogue for physical activity with a virtual coach

by

Beyza Hizli

Supervisor:	Dr.ir. W.P. Brinkman
Daily supervisor:	Ir. N. Albers
Institution:	Delft University of Technology
Place:	Faculty of Electrical Engineering, Mathematics and Computer Science, Delft
Project Duration:	December, 2020 - June, 2022

TU Delft

Acknowledgements

"No one who achieves success does so without acknowledging the help of others. The wise and confident acknowledge this help with gratitude."

This work was completed as part of the Computer Science master's at Delft University of Technology, in cooperation with Perfect Fit.

Nele, my daily supervisor, you are the best supervisor I have had in my life. No one gives feedback as fast as you do, and your support and guidance helped me a lot. Thank you to my main supervisor Willem-Paul for guiding me during this thesis project and taking the time to do so.

To all the students at the Monday morning meetings, I am happy that you guys were there and helped out whenever you could. I feel like you were the ones that could relate to the things I was going through with regard to my studies. It is great that, after meeting online for so many months because of covid, we were able to meet each other in real life this year.

I would like to thank my family. Mom and dad for always supporting me and for giving me the opportunity to study. My sister Berna for always being there for me and dealing with my high stress points. Michiel for always supporting me and being patient with me. Melda for always brightening my day. I love you all and I truly appreciate having you in my life.

Thank you to my friends (Plululup group and Mandy) for giving me the most fun and happy times, and distracting me from stressing about my studies when I needed it. I am so happy to have you guys in my life. I should have more time to hang out and play video games with you now :).

I spent a big part of my life studying at Delft University of Technology. I am grateful that I was able to study here and that I have met such amazing people that mean so much to me. Even though studying was very stressful at times, I will look back at my time here with a good feeling in my heart.

Beyza Hizli
Rijswijk, June 2022

Abstract

Lack of physical activity is one of the main risk factors for the development of cardiovascular diseases. Although most people are aware of the risks of this poor health behavior, they have a hard time improving this behavior. Goal-setting is an effective step in the process of changing behavior. It can motivate a person and help them to remain focused on the desired outcome, which increases the chance of successfully achieving a goal. In this thesis, we studied the design of a goal-setting dialogue for a virtual coach to motivate people in the context of physical activity. The dialogue was designed to support people in setting SMART goals and to raise their self-efficacy. Part of the design was vicarious experiences in the form of examples of people who successfully achieved a physical activity goal. We gathered these examples and fit a model to predict which examples to show to a user during the conversation with the virtual coach. An experiment was conducted to evaluate whether there was an increase in users' self-efficacy after the conversation with the virtual coach, how motivating the given examples were, and users' attitudes towards the virtual coach. The results indicated that users' self-efficacy was lower after the interaction with the virtual coach. However, we found that people considered the given examples motivating and had a positive attitude toward the virtual coach.

Contents

1 Introduction

1.1. Motivation

According to the World Health Organization, cardiovascular disease is the number one cause of death worldwide, with a death rate of approximately 17.9 million lives a year [19]. Lack of physical activity is one of the main risk factors for the development of cardiovascular disease [28]. Although most people are aware of the risks of this poor health behavior, they have a hard time improving this behavior. Three factors that are important to change behavior are capability, opportunity, and motivation [71]. To increase motivation, one can set goals [60]. Goal-setting is a common step in the process of changing behavior [29]. It can motivate a person and help them to remain focused on the desired outcome, which increases the chance of successfully achieving a goal. However, setting realistic and well-defined goals can be a difficult task [64]. Often, people aim for unrealistic goals or set vague goals where it is hard to measure whether they actually achieved the goal. Locke and Latham are the founders of the goal-setting theory. They identified principles that affect the chance of achieving a goal successfully, such as goal commitment and goal clarity [66].

Goal setting is used in both traditional behavior change interventions and smartphone applications, such as apps to promote physical activity [72]. There are many apps that can help someone keep track of their physical activity and set target goals, such as losing weight or taking a number of steps per day [84]. The goals that are set with such apps specify a certain behavior or activity, have a deadline, and in some cases can be measured.

Previously evaluated physical activity applications are web tools and mobile apps to set goals as part of an intervention. However, the goals that are set in these apps are limited in terms of relevance and level of attainability. Relevance and attainability are attributes of the SMART goal-setting framework [36], a framework based on the goal-setting theory [76]. SMART goals are Specific, Measurable, Attainable, Relevant, and Time-bound. Making sure that the goal conforms to these principles can make the goal more effective and result-oriented [77]. It is important to ensure that a goal matters to a person, that it aligns with other relevant goals of that person [75] and that the goal is realistic and achievable by that person [76]. To enhance the relevance and attainability, certain questions could be

asked to the user, such as *"Why is this goal significant to your life?"* or *"Is achieving this goal realistic with effort and commitment?"* [61, 62]. When people are supported to realize why they want to achieve a certain outcome, they are able to create more personalized goals and therefore achieve better results [62]. Thus, by enhancing the relevance, personalization is also increased. Previous studies have shown that personalized goal-setting significantly improves physical activity [23].

Current physical activity applications are only focused on the goal being specific, time bound, and measurable. Therefore, there is room for improvement for goal-setting apps that support people to set relevant and achievable goals. Furthermore, in current approaches, users do not set goals with a conversational agent. Conversational agents are increasingly playing an important role in health care, assisting clinicians during consultations and supporting patients in their behavior change process [59]. Setting goals together with a virtual coach instead of using web tools or humans can be beneficial in several ways. A dialogue with a virtual coach can serve people where traditionally human coaches would have provided support. Chatbots are easy to understand and communication is similar to how people interact with each other through computers or other smart devices [102]. Moreover, mobile apps are useful for people who are reluctant to make an appointment with counselors. A chatbot would be available at all times, so the person is not restricted by working hours and would have the freedom to start the intervention whenever suits them best [102]. Additionally, chatbots are able to reach a broad audience and decrease the workload of clinicians and human operators where automation is possible [31]. Using a conversational agent instead of a web tool could increase adherence to the intervention due to social interaction [42]. Research shows that using a conversational agent instead of normal text interfaces creates more arousal, trust, and commitment [56].

To address current limitations, this research proposes to design and develop a goal-setting dialogue for a conversational agent to motivate users in the context of physical activity. The aim is to provide a dialogue that supports users in setting SMART goals and motivate them in with regard to physical activity because of the integral role motivation plays in changing health behavior [10].

In addition, this research offers a new approach to the goal-setting process by creating a chatbot with the purpose of setting physical activity goals. The aim of this project is to create a goal-setting dialogue framework that can be applied to set physical activity goals with a virtual coach.

1.2. Research question

The aim of this thesis is to design and implement a goal-setting dialogue that a virtual coach can use to support people in setting physical activity goals. The conversation with the virtual coach should motivate the user with regard to physical activity. The main research question can be defined as follows:

How can a goal-setting dialogue for a virtual coach be designed to motivate people in the context of physical activity?

To optimize the goals that are set, the goal must adhere to the principles of the goal-setting theory. The goals that are set should be SMART, but especially relevant and achievable, since these are the aspects that are currently lacking in goal-setting interventions. Furthermore, it should motivate the user to commit to the goal and change their behavior. It is important to understand how the virtual coach should interact with the user. This relates to the type of questions that are asked by the virtual coach, the language that is being used, and also the input that is expected from the user. When these requirements are understood and defined, we need to know how to design such a goal-setting

dialogue. Finally, the goal-setting dialogue that is built needs to be evaluated. First, we could evaluate the level of motivation because we want to increase motivation as it positively affects behavior change [71]. Second, self-efficacy is relevant because it affects motivation and behavior [89]. Finally, it is useful to know whether the virtual coach is liked and accepted by the user. Provoost et al. show that acceptability is used in many studies to evaluate the effectiveness of a conversational agent [81]. The following sub-questions are defined based on the previous matters:

- *What are the requirements for a goal-setting dialogue with a virtual coach in the context of physical activity?*
- *How can a goal-setting dialogue with a virtual coach be designed in the context of physical activity?*
- *How effective is the designed goal-setting dialogue?*
  - Self-efficacy
  - Motivation
  - Establishing a positive attitude towards the virtual coach

1.3. Approach

To answer the above-mentioned research questions, the first step was to explore and research the current state of the art. What is the goal-setting theory, how do we set relevant and achievable goals, how do virtual coaches interact with the user, and how can they motivate the user? In addition, we explored how conversational agents have been designed and used up to now, and what existing physical activity applications do. From this literature research, requirements were defined (Chapter 2). In addition, the advice of experts was desired to make design choices. This expert consultation aimed to find out and understand why experts would choose certain design choices over others, to aid in making the final design decision. After the literature research and expert consultation, we designed a dialogue for the conversational agent (Chapter 3). Subsequently, an experiment was set up to evaluate the created design (Chapter 4). Finally, the evaluation results were discussed, limitations were identified, future work was proposed, and final remarks were given (Chapter 5).

2

Foundation

This chapter answers the first sub-question:
What are the requirements for a goal-setting dialogue with a virtual coach in the context of physical activity?

To define the requirements for a goal-setting dialogue with a virtual coach, we explored previous work and current practices and applications to have a good understanding of the state-of-the-art. Additionally, experts in the field of psychology were consulted for further discussion on design choices. At the end of this chapter, a list of requirements is presented for the goal-setting dialogue.

2.1. Related work

In this section, previous research is explored. The topics that are covered are goal-setting, personalization, conversational agents, and current applications for physical activity. At the end of each topic, the key takeaways and requirements that can be derived are summarized.

2.1.1. Goal-setting

Locke and Latham have been researching goal-setting and motivation since the 1960s [75]. In the goal-setting theory they developed, they consider human actions to be purposeful and directed by conscious goals [75]. Content and intensity are two attributes of goals that have been studied extensively. The content concerns the difficulty of the goal and how specific or vague a goal is. The difficulty of a goal differs per person, as we all have different knowledge, skills, and abilities. Locke and Latham have found a linear relationship between difficulty and performance. Given that a person can achieve a certain goal, the more challenging the goal, the better they adjust their efforts to the difficulty of the task and the better they perform [75, 66]. Furthermore, the authors found that more specific goals lead to better performance. The main reason for that is that when people set vague goals, they tend to be satisfied more easily by their performance, even if this is lower than what they can achieve. Another

aspect of specific goals is that the more specific the goal is, the less variable the goal becomes. This means that there are fewer possible outcomes for that goal, which is desirable because it makes it clear when a goal is achieved. All of these findings are based on studies that focus on many tasks, from work-life-related tasks to sports. [75]. This shows that the founders of the goal-setting theory have researched goal-setting in many fields, including goal-setting for physical activity, which is what we are interested in.

The second attribute of goal-setting that has been studied extensively is intensity, which is related to goal commitment. The main aspects of intensity are the scope, clarity, and mental effort involved in mental processes [82]. Gollwitzer, Heckhausen, and Ratajczak found that the more intensely a person tries to solve a problem, the more likely they are to become committed to the goal and achieve the goal [39]. Commitment refers to how important a goal is considered, how determined a person is to achieve it, how attracted they are to the goal and how well they stick to the goal when faced with obstacles [75]. The degree to which a person feels committed to a goal is enhanced when people believe that they can achieve a goal and when they believe that achieving the goal is important. An individual's belief in his or her capacity to perform an action is referred to as self-efficacy. Self-efficacy includes the ability of a person, their experience, training, the information they have obtained, previous successes, and internal attributions [66]. For example, if a person has experience or past successes in completing a task, they have more self-efficacy. Increasing self-efficacy has a positive effect on successfully completing a task and increases motivation to do a task [71]. Additionally, self-efficacy can be enhanced by encouragement [101] and vicarious experience [5]. Vicarious experiences deliver a feeling or experience from someone else. They are most effective when a person observes someone similar to themselves successfully achieves a goal [9]. In addition to increasing self-efficacy, understanding the purpose of the goal and why the goal is important can also be used to enhance the commitment to a goal [62].

Another relevant aspect of goal-setting is that receiving feedback on goals is effective and leads to higher performance [74, 75]. Feedback provides information on the degree to which a person met the standards. If the performance meets or exceeds the standard, the same performance is generally maintained. If the performance is below standards, a person may be dissatisfied. If that person has high enough self-efficacy, they might set higher goals to meet the standards.

Based on the goal-setting theory, the SMART goal-setting framework was created [36]. SMART goals are Specific, Measurable, Attainable, Relevant, and Time-bound [27, 61]. Specific means that the goal provides a detailed description of what is to be accomplished. Measurable goals are quantifiable, which makes it possible for progress and targets to be measured and having benchmarks. Attainable means that the goal should be realistic and achievable by the person who sets the goal. Relevant goals are consistent with the vision of the person and what the person considers important. Finally, time-bound goals have a target achievement date, so there is a deadline to focus on. Making sure that the goal is consistent with these SMART principles can make the goal more effective and result-oriented [77]. This is desirable because it increases the chance of achieving a goal. The SMART goal-setting framework captures the important aspects of the goal-setting theory. However, one thing that is not considered is self-efficacy.

We can derive two requirements from the goal-setting topic:

- R1: The virtual coach should motivate the user to enhance their self-efficacy.
- R2: The dialogue with the virtual coach should support the user in creating specific, measurable,

attainable, relevant and time-bound (SMART) goals.

2.1.2. Personalization

Personalization is the process of making something suitable for the needs of a particular person [57]. 'Personalization' is a term with many definitions [37]. For clarity, we use the definition used by Fan et al. [32], defining personalization as "a process that changes the functionality, interface, content, or distinctiveness of the system to increase its personal relevance." It is important to consider what will be personalized and for whom it will be personalized. Moreover, the type of personalization should be considered, which can be implicit or explicit. In implicit personalization, the required information is obtained automatically already by the observations of the system, and in explicit personalization, the user needs to actively participate in providing information, such as answering questions. In digital technology, personalization can be achieved by changing the interface, functionality, or access to specific content. Examples of this are personalized feedback, daily health reports, personalized reminders, alerts, warnings, and recommendations [57].

Rimer and Kreuter use the term 'tailoring' to define a process to create personalized communications [83]. This type of tailoring is included in the definition we use for the term personalization. Tailoring uses data from an individual to determine which information or strategies meet that individual's needs best. It can enhance motivation by matching people's interests and needs, by framing information in a context that is meaningful to the person, by using design elements to capture the person's attention, and to provide the amount and type of information the user prefers [83]. This could consequently lead to increased attention and ultimately to an increased likelihood of behavior change. As an example, the Woebot applies personalization by showing specific content to individuals depending on their mood state [35].

The study by Lee et al. proposes a different way of personalizing health services [62]. They argue for a reflective strategy that helps people realize what matters to them and enables them to better personalize services themselves. This not only increases personalization, but also helps with commitment to the goal. They achieve this by asking specific questions, in particular *why* questions.
Key takeaways from this section are that personalization can be useful in motivating the user, increasing the user's attention, and ultimately enhancing the chance of successful behavior change. This leads to the following requirement:

- R3: The dialogue with the virtual coach should be personalized.

2.1.3. Current approaches

Conversational agents

Conversational agents are computer programs that interact with a person through speech and text. Laranjo et al. researched 14 different conversational agents in healthcare [59] and found that they are mainly task-oriented, for example, assisting clinicians during consultations and supporting users in their behavior change process. With advances in voice recognition, natural language processing, and artificial intelligence, conversational agents are increasingly playing an important role in health care [59]. Car et al. researched 47 articles on the topic of conversational agents in healthcare [18]. They found that conversational agents in healthcare are aimed at treatment, monitoring, and health service support, and that most conversational agents are text-based, delivered through smartphone applications. The role of a virtual coach is to teach the user new skills, provide a sense of companionship by establishing

an effective relationship based on trust, and provide the relevant and accurate information that the user requests [54].

An example of an automated conversational agent that is used in healthcare is Woebot [35]. It is designed to provide cognitive behavioral therapy (CBT), supporting people who self-identify as having symptoms of anxiety and depression. Another example is the virtual coach created by Watson et al. to increase levels of physical activity in overweight adults [97]. The authors also mention that virtual coaches provide an interactive relationship and can form a social bond with the user, which is absent in the use of web tools. Using a conversational agent instead of a web tool could increase adherence to the intervention due to social interaction [42].

Physical activity applications

In an extensive study on mobile health applications by Sama et al., most of the applications were fitness or training apps (174 of 400 apps) [85]. An interesting finding in their research is that only 8 of 400 mobile health apps that the authors evaluated used goal-setting as their main engagement method, indicating that goal-setting is either not at all used in the apps or only used as a secondary engagement method. We explored apps in the context of physical activity. These apps were found in previous literature, but also in the Google Play Store and Apple app store [40, 3] . In most of the physical activity apps [86, 44, 98, 95], the user begins by creating his profile by filling in relevant demographics such as their name, gender, and weight. The apps offer the ability to set goals for a given set of activities or to select one of the offered goals. Moreover, users often have to select a deadline for a goal. Most physical activity apps offer an overview of the activities they have completed, some apps offer guides for physical exercises, and some for diets. When setting goals, users are limited to what can be tracked by the app.

This section shows that mobile applications have been used to successfully increase the level of physical activity of people and that the key features they offered were setting target goals for specific activities, showing an overview of the activities, and providing guidelines for physical activities.

2.2. Expert consultation

With the information obtained from the literature research, a general idea of the requirements is obtained. In addition to the literature, the advice of experts in the field of psychology was desired to make design choices for the goal-setting dialogue. To make these decisions, experts were consulted and asked for input on scenarios. Scenarios are examples of use situations of the goal-setting dialogue to help people understand the context of the proposed technology [92]. They describe how people would use (part of) a system to perform tasks or activities. It is a relatively lightweight method to envision future use cases; it allows people to imagine a use case [20]. Scenarios are useful because they are easy to create compared to actually implementing examples, and when people can envision a use case, they can make more accurate decisions about them than without them. Moreover, scenarios are useful in the design process because they capture the consequences and trade-offs of potential designs [20]. The scenarios were presented together with claims to emphasize the differences in the scenarios and to raise a discussion. The aim of the expert consultation was to find out and understand why experts would choose certain design choices over others.

2.2.1. Method

Participants

To further refine the requirements and obtain additional insights, the scenarios were discussed with the following four experts:

- a psychologist and assistant professor, expert on changing health behavior and smoking cessation.
- a senior eHealth researcher and scientific manager
- a PhD student at unit Health, Medical, and Neuropsychology and medical psychologist at a cardiology clinic.
- a PhD student in the field of biomedical signaling and systems, researching the use of sensors for the physical activity running.

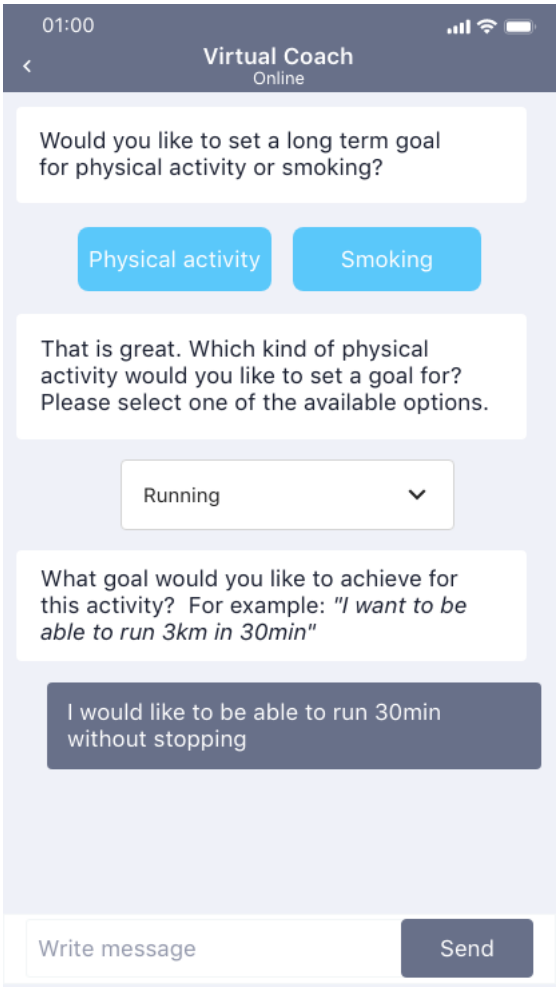
The first three experts were all related to the psychology field, the fourth was more involved in the development of the app itself, with a focus on physical activity and the use of sensors. Throughout the following sections, when experts are named without explicitly mentioning their background, we refer to psychology experts.

Materials

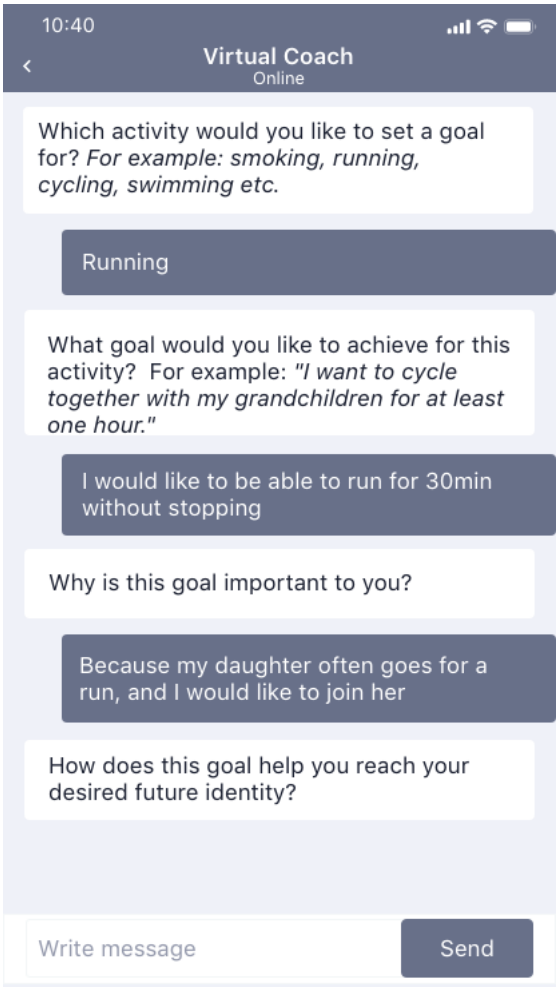
Four scenarios were created, which contain use case examples for parts of the goal-setting dialogue. Each scenario was presented with two or three options on how parts of the dialogue could be approached. Additionally, each scenario was presented together with claims to emphasize the different approaches for a part of the dialogue and raise a discussion. The first scenario is shown in Figure 2.1 as an example. All scenarios can be found in Appendix A.

Procedure

The four experts were invited to one of the three scenario meetings. In each meeting, the focus of the discussion and the questions was adjusted to the relevant expertise. The purpose of the discussions was to find out why experts would choose certain design choices over others and why they considered certain aspects of the dialogue important. Additionally, questions were asked where their input and feedback were desired.



Long-term goal-setting option A



Long-term goal-setting option B

Figure 2.1: This figure shows an example of the scenarios. In this scenario, two options are given to approach the long-term goal-setting part. In option A, the user needs to select the type of activity for which they are setting a goal and write down the long-term goal they would like to achieve. In option B, the user writes down the type of activity instead of selecting it. In addition, extra reflective questions are asked about why the goal is important to the user.

2.2.2. Findings of expert consultation

In the following, we report the main findings of the discussions with experts. The proposed claims are explained based on the past literature, and the discussion results are given.

Claim: The virtual coach should ask reflective questions

This claim states that the virtual coach should ask the user why the goal is important to them and how this goal helps to reach their future identity. These reflective questions posed in scenario B of the scenarios in Figure 2.1 are based on the findings of Lee et al. [62], where they use a reflective approach to enhance the personalization and commitment that the user experiences. The experts believed that reflective questions are useful, explaining that the user will be more involved and more committed in this way. However, they were concerned that those questions might be very hard to answer. To solve this problem, they came up with several ideas. One was to make use of questionnaire-type questions such as the Likert scale, so that users can answer the questions on, for example, a scale from 1-10.

Another way that was proposed to make the process easier and to ensure that all the elements of goal setting are included was by giving the user sentences in which they have to fill in the blanks.

An example that one of the PhD students mentioned is the ”Wat er toe doet”, in English: ”What matters”, questionnaire [96]. It presents four questions to help a person realize what is important in their lives. In the first question, the user needs to select 3 out of 9 topics that are important to them, which are topics such as family, physical or mental health, and their relationship. By offering options that a user can choose from, the action becomes simpler because the user does not have to come up with their own topics. Next, the user is asked to write a goal for one of the three topics they have selected. When they choose family, a relevant set of examples is given such as ’I want to visit my family members on their birthdays.’ A similar method can also be used in the goal setting dialogue. If users are presented with topics for which they can set a goal and examples of goals are given for those topics, the process of coming up with a goal can be facilitated.

A requirement that can be derived from this discussion is that the virtual coach should ask the user reflective questions, so that the commitment to the goal and perceived personalization are enhanced.

- R4: The dialogue with the virtual coach should make the user realize why the goal they are setting is important to them; the user should become committed to the goal.

Claim: The virtual coach should give examples of goals
For the examples that the virtual coach presents to the user in the dialogue, the PhD student and senior researcher suggested not to put them in the beginning of the chat. Their reasoning for this is that the user might be pushed in a certain direction if they see the example first. Experts agreed that showing examples of people who achieved a goal of physical activity can help to increase the user’s self-efficacy. However, they mentioned that the examples that are shown should suit the user, as it can otherwise have a discouraging effect. The use of examples to improve self-efficacy is in line with what has been discussed previously and belongs to the first requirement (R1).

Claim: The virtual coach should be friendly and use emojis
Since the use of emojis became popular and can be useful, health researchers have started to apply them in health interventions [100]. The use of emojis can make the virtual coach seem more friendly and approachable. Experts mentioned that the use of emojis makes the dialogue look more appealing and helps break the distance between the virtual coach and the user, which can make the user feel more comfortable. We mentioned our concern that users might consider the virtual coach less credible when using emojis [100], however, the senior researcher believed that it is okay to use emojis as long as they are not overused. The language used in one of the scenarios was inspired by the Woebot [35] and other research following guidelines for e-coaching [73, 100]. These studies indicate, among other things, that starting with a compliment has a positive effect on the perceived conversation quality by the user. Experts agree that giving compliments and encouraging the user is good practice. The psychologist said that, in general, it was better to give a little too many encouragements than too few, with the idea that the majority of the users would be happy with the encouragement. When the user is encouraged, their self-efficacy could increase [101]. The results of this discussion are in line with the first requirement (R1).

Claim: The dialogue should guide the user in creating SMART goals
To make sure that the goal is SMART, the goal needs to be specific, measurable, achievable, realistic and time bound. To make the goal achievable, the virtual coach in the third scenario tries to make the user aware of the feasibility of the goal. However, one of the psychology PhD students suggested that the virtual coach should explicitly ask whether the user thinks the goal is achievable, so that they actually consider whether the goal is achievable. The psychologist agreed with asking that question, but suggested letting the virtual coach first say something positive, because of concerns that the user might be demotivated if the coach asks if they believe their goal is achievable.

Conclusions that we drew from this discussion is that it is good to support users in setting SMART goals by asking them specific questions. This is in line with the second requirement that we defined earlier (R2). Furthermore, the virtual coach should explicitly ask the user if they believe that the goal is achievable so that the user thinks about whether the goal they are setting is realistic.

- R5: The dialogue with the virtual coach should make the user consider whether the goal they are setting is achievable and realistic.

In addition to the advice the experts gave regarding the claims, their general advice was to keep the language easy to understand, short and simple. This is in line with the methods of Provoost et al. [81], where they agree to keep text readable by using short and clear sentences. In addition to keeping the language simple, experts also recommend keeping the process as simple as possible. Providing options from which users can select makes it easier for users to complete the goal-setting dialogue.

Additionally, one PhD student mentioned that it might be wise to give the user the ability to change the goal if they want to. The reason for this is that during the dialogue the user thinks about their goal in more depth and might realize that the goal that they initially wrote down is not what they want to achieve. If they are able to change it, they can set a new goal that they would like to achieve. From this last discussion, we derived the following requirement:

- R6: The language used by the virtual coach should be simple, short, and clear.

2.3. Requirements

Based on the literature review and expert consultation, the following requirements for the goal-setting dialogue with the virtual goal are defined (in no particular order of importance):

- R1: The virtual coach should motivate the user to enhance their self-efficacy.
- R2: The dialogue with the virtual coach should be personalized.
- R3: The dialogue with the virtual coach should support the user in creating specific, measurable, attainable, relevant, and time-bound (SMART) goals.
- R4: The dialogue with the virtual coach should make the user realize why the goal they are setting is important to them; the user should become committed to the goal.
- R5: The dialogue with the virtual coach should make the user consider whether the goal they are setting is achievable and realistic.
- R6: The language used by the virtual coach should be simple, short, and easy to understand.

3

Design

This chapter describes the proposed solution to the research question:
How can an effective goal setting dialogue to increase physical activity with a virtual coach be designed?

The designed solution is explained first with its relation to previous research. Next, the designed goal-setting dialogue flow is explained and a more detailed description of the dialogue is given with its relation to the requirements that we established in the previous chapter. The data-driven part of the solution is explained in detail afterwards, followed by an analysis of the collected data for the data-driven part.

3.1. Design overview

To meet the requirements that were established in the previous chapter, a goal-setting dialogue was designed for a conversational agent. In the previous chapter, the goal-setting theory was explained and we found that goal-setting is used to motivate people to change their behavior. We found that self-efficacy affects the goals we choose, how we approach the goal, and goal commitment. In our proposed design, we tried to enhance self-efficacy in two ways: verbal persuasion (encouragement) [5, 101] and vicarious experience [5]. Vicarious experiences are most effective when a person observes someone similar to them successfully achieve a goal [9]. Ashford et al. reviewed interventions aimed at increasing self-efficacy for physical activity to research the relation between the used intervention techniques and the change in self-efficacy [4]. They found that self-efficacy was significantly higher when vicarious experience was used as an intervention technique, supporting the theory that seeing similar people perform a behavior can raise an individual's belief that they can also achieve the same [5, 4, 9]. Based on this theory and review, we proposed the idea of using a model that predicts which examples to show to the user to enhance the user's self-efficacy. The idea of this model was to predict which example the user would find motivating based on what similar users found motivating. To create this model, data was needed. We needed examples to show and data as input for the prediction model. How these data are collected is explained in detail in Section 3.3.

Furthermore, we described the SMART goal-setting framework [36] that was based on the goal-setting theory of Locke and Latham [75]. Making sure the goal is consistent with these SMART principles can make the goal more effective and result-oriented [77]. Therefore, we incorporated this framework into our design. The virtual coach asks questions to make the user consider each aspect of the framework. Two of the principles, how achievable and how relevant the goal is, are especially important aspects of our design because previous physical activity interventions lack these principles. We found that asking reflective questions helps with goal commitment and personalization [62], which helps for the relevance of the goal. Therefore, the virtual coach asks reflective questions, based on the framework [96] that one of the experts proposed in the expert consultation.

The design phase was divided into two parts. The first phase covers the design of the goal-setting dialogue with the conversational agent, and the second part focuses on the data collection, in which the examples were collected and a prediction model was created.

3.2. Goal-setting dialogue

In this section, we explain the goal-setting dialogue in more detail. Figure G.1 gives a high-level overview of the dialogue flow. The goal-setting dialogue is designed to set goals for physical activity, in particular for running or walking. These activities are chosen to fit the scope of the Perfect Fit project.

From now on, we will refer to the virtual coach using the name Jody. We decided to give the virtual coach a gender neutral name to avoid gender bias [34].

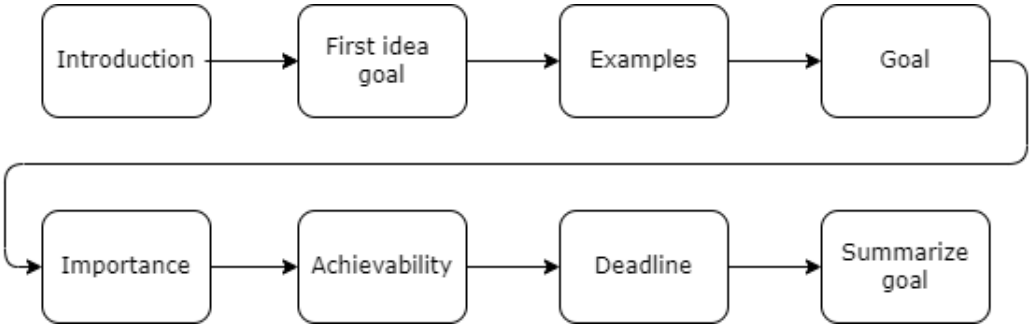


Figure 3.1: A high-level overview of the dialogue flow.

At the beginning of the dialogue, Jody introduces themselves and asks the mood of the user. Asking their mood helps the virtual coach appear friendly and increases the perceived personalization of the conversation. Jody responds appropriately according to the selected mood and continues by explaining the purpose of the conversation. Jody asks whether the user wants to set a running or walking goal and asks the user to think about the goal they want to set. Jody emphasizes that this is not their final goal and that they will make their goal more specific later on. Jody asks the user to think about their goal themselves first, because the psychologist expert warned us about giving examples beforehand. When the user sees an example beforehand, it is more likely that they will copy the example instead of properly thinking about it themselves. Copying an example or creating a goal based on the given example can happen unconsciously. It is important for the user to think about the goal themselves to increase the chance that they are setting a goal that is relevant to them.

After the user has an initial idea of what running or walking goal they would like to achieve, Jody shows two examples of other people who have achieved a running or walking goal. As we explained in the previous section, the examples are shown as vicarious experiences to enhance the user's self-efficacy. The examples consist of two parts. The first part is an introduction of the example person, so that the user has an idea of what type of person the example concerns. The second part of the example describes the walking or running goal that the person achieved, including a short description on how they achieved their goal. Running or walking goal examples are shown because these are the type of goals that the user is setting, and we want to increase the user's self-efficacy with regard to running or walking. Figure 3.2 shows an example of someone who achieved a running goal.

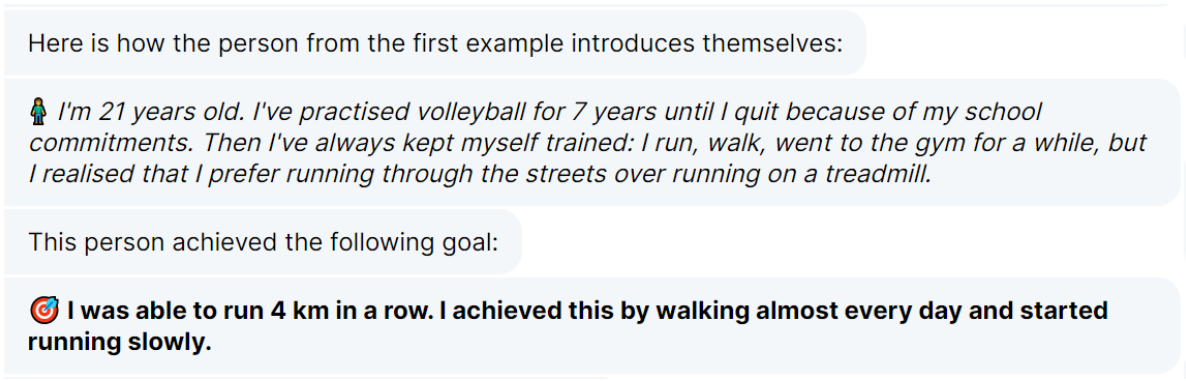


Figure 3.2: Screenshot of part of the dialogue showing an example of a person who achieved a running goal.

Before Jody shows the second example, Jody asks whether the user finished reading the first example. If the user indicates that they have finished reading by pressing the button that becomes available, Jody continues. This is done to avoid showing too much text at once and to minimize the chance that people skip reading parts.

After the user has read the examples, Jody asks the user to think about their goal again and write it down in more detail. This refers to the principle of the SMART goal-setting framework that goals should be specific. The conversation continues with this framework by asking a reflective question: why the running or walking goal the user set is important to them. This improves the commitment to the goal and perceived personalization [62]. One concern that was raised in the previous chapter was that these types of questions might be hard to answer. Our proposed solution guides the user by giving them options to select from (see Figure 3.3), based on the 'What matters' questionnaire [96]. Jody first offers options to choose from: *Family, Health, Relationships, Personal Growth, Work, Friends* (Figure 3.3). Consequently, Jody asks why their goal is important for their chosen option.

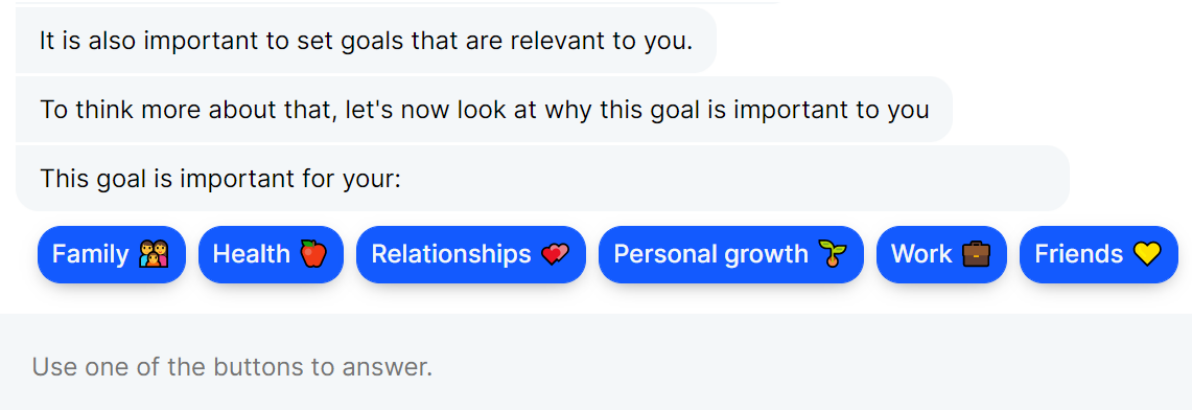
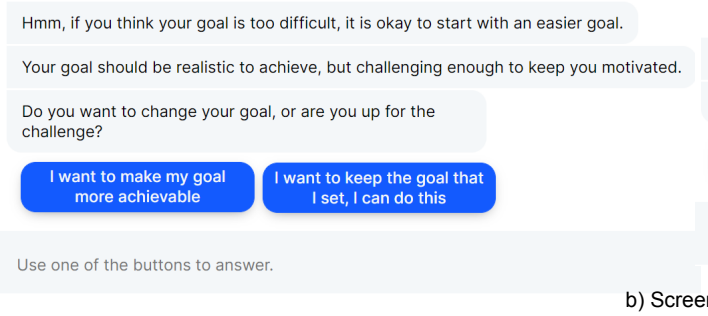


Figure 3.3: Screenshot of part of the dialogue showing buttons to choose why the running or walking goal the user set is important to them.

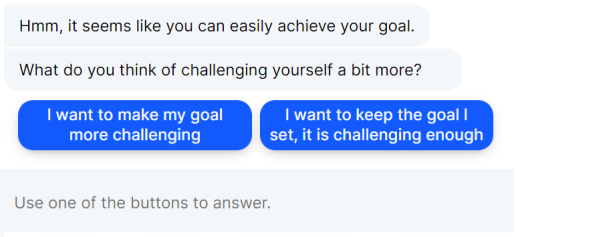
After addressing the relevance principle of the SMART goal-setting framework, Jody continues with the next principle: attainability. Jody asks the user how achievable they consider their goal (on a scale from 1 to 7, where 1 is not achievable and 7 easily achievable). If the user indicates that it is hard for them to reach their goal (1 or 2) or easy for them to reach their goal (6 or 7), Jody gives them the opportunity to change their goal to something that suits them better (Figure 3.4).
The question was asked as follows:

How achievable is your goal [goal] to you?
Select a number between 1 and 7 where 1 is not achievable and 7 is easily achievable.

The questions about whether they want to change their goal were formulated following the advice of psychologist experts, to minimize the risks that the user might be demotivated by the question.



a) Screenshot of the dialogue when the user indicates that their goal is easy to achieve.



b) Screenshot of the dialogue when the user indicates that their goal is hard to achieve.

Figure 3.4: In both cases a) and b), the user is given the option to change or keep their goal.

To make the goal time-bound, which is also a principle of the SMART goal-setting framework, Jody asks when the user wants to achieve their goal and checks the deadline the user sets. If the deadline is in the past or in the near future (within a week), Jody asks the user to pick a different deadline. When the user sets a valid deadline, Jody asks the user whether the deadline they set is correct, and otherwise gives them the opportunity to change it.

When the deadline is set, Jody summarizes the goal and offers the possibility to change the goal or

the deadline for the goal. When the user indicates that they are happy with the goal, Jody congratulates them on setting the goal and gives some encouraging words before saying goodbye. Encouraging and complimenting the user is in line with the e-coaching guidelines [73, 100] that we found in the previous chapter.

3.2.1. Personalization

In Chapter 2 we found that personalization is useful to motivate the user, to increase the user's attention, and ultimately to enhance the chances of successful behavior. In the designed goal-setting dialogue, we found multiple ways to include personalization elements.

Jody provides examples of people who successfully achieved a running or walking goal. The selection of these examples is personalized: the examples that are predicted to be most motivating for that user are shown to the user. The personalization aspect here is the use of examples that are considered motivating by similar people rather than using general examples.

Other personalization elements are present in the dialogue flow. When the user chooses an answer, Jody responds differently depending on their choice. For example, Jody asks the user what their mood is by giving mood options that the user can choose from, and responds differently based on their choice. Another example is that Jody offers the user the opportunity to change their goal if they indicate that their goal is difficult or easy to achieve, which changes the direction of the dialogue.

3.2.2. Language

Attention is paid on the language that Jody used. As explained in the previous chapter, it is desired to keep the language easy to understand, short, and simple. Therefore, the language used by the coach is kept simple and longer text messages are split up into multiple sentences to make it more readable. Another aspect of language is the use of emojis. The use of emojis can make the conversational agent appear more friendly and approachable [100]. Psychology experts were in favor of using emojis in the dialogue, as long as they were not overused. Jody uses emojis every now and then in the messages and for the buttons provided (see Figure 3.3 for emoji examples).

In addition to using easy language, experts also recommended keeping the goal-setting process simple. By offering options that the user can choose from, the action becomes simpler, because the user does not have to come up with an answer themselves. Buttons are used for multiple questions throughout the conversation.

Finally, Jody follows the principles of motivational interviewing by expressing empathy (e.g., acknowledging the user's mood), by cheering the user on and complimenting the user (e.g., complimenting them for thinking about their goal), and supporting self-efficacy and optimism (e.g., encouraging the user to achieve their goal) [46].

Table 3.1 below gives a brief overview of the requirements we established in Chapter 2 with the proposed solutions.

Table 3.1: Overview of requirements with a brief description of the design solution.

	Requirement	Design solution
R1	The virtual coach should motivate the user to enhance their self-efficacy.	The virtual coach tries to enhance the users' self-efficacy by showing examples of others that successfully achieved a goal (vicarious experiences) and encouraging them (verbal persuasion).
R2	The dialogue with the virtual coach should be personalized.	The examples that are shown by the virtual coach are personalized and the direction of the dialogue can change based on specific user choices.
R3	The dialogue with the virtual coach should support the user in creating specific, measurable, attainable, relevant and time-bound (SMART) goals.	The virtual coach asks questions to make the user consider each element of the SMART goal setting theory.
R4	The dialogue with the virtual coach should make the user realize why the goal they are setting is important to them; the user should become committed to the goal.	The virtual coach asks reflective questions to make the user realize why their goal is important to them to increase relevance and consequently commitment to the goal.
R5	The dialogue with the virtual coach should make the user consider whether the goal they are setting is achievable and realistic.	The virtual coach asks the user whether they believe their goal is achievable, and supports them with changing their goal if they want to.
R6	The language used by the virtual coach should be simple, short and easy to understand.	The language used by the virtual coach is simple and long sentences are split up for readability

3.3. Example data

Part of our designed dialogue is the examples that are shown to the user to enhance their self-efficacy. We mentioned before that we use a model to predict which examples to show to the user. Two ways were considered to predict these examples:

1. Predict which examples to show based on which examples people with similar user characteristics perceived as similar.
2. Predict which examples to show based on which examples people with similar user characteristics found motivating.

We needed to gather data for both models. For the similarity ratings, it was necessary to obtain ratings on how similar someone perceives the person of the example. For the motivation ratings, we needed to obtain ratings of how motivating the user perceives the example. Thus, we created examples consisting of two parts. The first part was the introduction of the person in which they mention something about their physical activity. We were interested in their physical activity because we wanted users to feel similar to them with respect to physical activity. We were interested in physical activity because we

want to increase the users' physical activity self-efficacy. In the second part of the example, the physical activity goal that the example person has successfully achieved is shown, and a brief description on how they achieved their goal is given. Additionally, describing how the example person achieved their goal instead of only showing the goal they achieved was done to enhance motivation.

Jody shows examples of people who achieved a running or walking goal in the proposed goal-setting dialogue. Therefore, we needed examples of people that achieved a running or walking goal.

3.3.1. Example data collection

To gather the examples and ratings of these examples, the data collection phase was divided into two parts. The first part (A) focused on obtaining the examples. The second part (B) focused on rating these examples on how similar the example people are perceived, and on how motivational the example goals are considered. Both parts A and B were approved by the TU Delft Human Ethics Research Committee (HREC reference number: 1707). These experiments were run in September and October 2021.

3.4. Data collection part A: Introductions and goal examples

The goal of this part was to gather introductions of people who have achieved physical activity goals.These introductions were used as part of the example that the virtual coach used to increase the user's self-efficacy and to motivate the user.

In our goal-setting dialogue, Jody showed two examples of a person who has achieved a walking or running goal, including an introduction of the example person and how they achieved their goal. Three questions were asked to obtain the first part of the examples:

1. *Please introduce yourself in 2-3 sentences as if you are introducing yourself to a new person you meet at your friend's gathering.* ***This person is interested in learning about your physical activity.***
2. *Describe one goal with regards to running or walking you have achieved in the past year.*
3. *Describe in one sentence how you have achieved your goal.*

The first question was asked to obtain information about the participant, expecting the participant to describe themselves. The question specifically mentions that they should introduce themselves to someone who is interested in learning about their physical activity, to ensure that they write about their physical activity. This was desired because the final goal was that people who read the example feel similar to or relate to the given example person with regard to physical activity as well.

The second and third questions asked for a goal the participant achieved with respect to running or walking and how they achieved that goal. The latter question was asked for motivational purposes; besides reading that the person in the example achieved a certain goal, it can be motivating and helpful to read about how they managed to achieve that goal [67]. The answers to the second and third questions were presented together during the conversation with the virtual coach.

3.4.1. Participants

Participants who were invited to take part in the study were fluent English-speaking adults (18 years and older). Prolific, an online crowdsourcing platform, was used to recruit participants. Eligible were

people who had achieved a running or walking goal in the past year.
Prolific offers the ability to prescreen participants based on user variable criteria. Diversity was desired because in the final conversation with Jody, there should be an example of a similar person available for any kind of user. To ensure diversity between the participants, groups with different user variables that were available beforehand for pre-screening in Prolific were created, and participants were recruited based on these groups. The groups filter on gender (2 levels: male/female), age (3 levels: 18-35, 36-55, and 55+), and activity level (3 levels: less than 60 minutes active per week, 60-120 minutes active per week, and more than 120 minutes active per week). The combinations of the three variables gender, age, and activity level resulted in 18 (2x3x3) different groups. The groups and their corresponding user variables can be found in Table D.1 in Appendix D. Four participants were recruited per group, resulting in a total of 72 participants. The number of participants was chosen based on the budget available for this part of the project and to make sure that we have at least 4 people per group.

Table B.1 in Appendix B shows the characteristics of the participants of part A.

3.4.2. Materials

Qualtrics was used to host the online questionnaires.

3.4.3. Measures

We collected data on user variables that served as input for the prediction model. The chosen variables are variables that were expected to have an influence on perceived motivation of people and variables measuring physical activity.

The COM-B model is a behavior model [99]. It identifies three factors that are necessary for behavior to occur: capability, opportunity, and motivation. Capability and opportunity in the COM-B model impact motivation, and together they impact behavior. The chosen user variables and their relationship to these factors are explained below.

Motivation

Motivation refers to mental processes that influence our decision making and behavior.

- Running- or walking self-efficacy (an adaptation of the exercise self-efficacy scale by McAuley [68], instead of exercise to specifically walking or running. The whole scale can be found in Appendix C). Self-efficacy is a motivational construct that affects choices, effort, persistence, and achievement [90].
- Big-5 personality (10-item TIPI questionnaire[41]). The Big-5 personality traits are related to intrinsic and extrinsic motivation [43][14].
- TTM-phase for becoming more physically active (an adaptation based on the Exercise: Stages of Change short form [93]). The TTM-phase that a person is in affects their motivation to change, but also what they perceive as motivating. For example, an already active person most likely does not consider walking 2000 steps every day motivating, because they already walk more than that.
- Number of hours spent sitting on a weekday and on a weekend day [53]. The number of hours spent sitting is related to the TTM-phase. For example, if a person sits a lot, the person is likely in one of the first stages of physical activity change.

- Physical activity self-identity (an adaptation to physical activity instead of exercise of the questionnaire from the paper by Anderson and Cychosz [2]). Physical activity self-identity measures the extent to which someone sees physical activity as a part of their self-concept, which affects their physical activity behavior.
- Need for cognition (the three items used in the paper by Steward et al. [94], which are in turn from the paper by Cacioppo et al. [16]). Need for cognition describes an individuals' tendency to engage in and enjoy effortful cognitive activity [15] and is shown to be positively correlated with intrinsic motivation [78].

Capability

Capability describes the physical and psychological capability that together with opportunity makes behavior possible. This includes, for example, a person's physical and mental functioning. For this study, we are interested in the physical activity level of the users. The following variables were obtained to determine the physical activity level:

- Godin-leisure time activity [38] to measure exercise and categorize based on physical activity, for which guidelines are given in the questionnaire.
- Weekly exercise (number of hours per week).

Opportunity

Opportunity describes the physical and social opportunity that together with capability makes behavior possible. This includes, for example, financial or material resources and social norms.

- Self-evaluated socioeconomic status and personal and household income. Both self-evaluated socioeconomic status and income could affect how much money a person has to spend on physical activity.
- Household size. Household size could, for example, impact whether people have someone to workout with.
- Highest completed education. Education level could impact physical activity in multiple ways, for example, it was found that for low-education people, not working and job loss were associated with reduced physical activity, while for highly educated individuals the opposite was true [30].

Smoking

In addition to the variables mentioned above, we gathered smoking variables because smoking affects a person's physical fitness [25] and is closely related to physical activity [47]. For example, smoking can be associated with significantly reduced odds of being moderately or highly physically active [79]. The following smoking-related variables were obtained only from participants who smoke at least once a day:

- Smoking self-identity, non-smoking self-identity and quitter self-identity (3 items each based on Meijer et al. [70]).
- Smoking group-identity (based on the ingroup ties subscale of the paper by Cameron [17]).
- Non-smoking group-identity (based on the ingroup ties subscale of the paper by Cameron [17]).

- Quitting smoking self-efficacy (4 items based on the paper by Scholl et al. [88]).
- Quitting smoking attempts [1].
- TTM-phase for quitting smoking [1].
- Quit smoking status, smoking status, and the smoking frequency were obtained from the participant's Prolific profile.

Demographics
Finally, we gathered the basic user demographics age and gender.

Part of the user data was retrieved from the Prolific profiles of the participants. This includes their age, gender, household size, personal and household income range, highest completed education level, self-evaluated socioeconomic status, weekly exercise amount in minutes, quit smoking status, smoking status and smoking frequency. The other variables were collected using a Qualtrics questionnaire.

3.4.4. Procedure

The experiment was divided into two parts: the pre-screening questionnaire to check whether the participants are eligible for this experiment, and the goal-questionnaire for the content of the experiment.

1. First, the participant completed the Qualtrics prescreening questionnaire which determined whether they were eligible for the experiment. Eligibility was based on whether the participant consented to participate in the study, whether they were fluent in English and whether they achieved a running or walking goal in the past year or not.

   Only participants who were eligible and completed the pre-questionnaire received an email with an invitation to the goal-questionnaire.

2. Participants had one week to respond to the invitation to the goal-questionnaire. The goal-questionnaire started with asking the participant to introduce themselves, and describe a walking or running goal they achieved in the past year and how they achieved it. Additionally, questions were asked to collect the user data variables described in Section 3.4.3.

   Participants were excluded after the second part if they failed half or more of the attention check questions, did not respond within a week, did not complete the goal-questionnaire, and/or clearly gave nonsensical answers to the free-text questions.

When a participant was excluded, they were replaced by a new participant who was recruited from the same user variable group. Participants who passed part 1 and/or 2 (were not excluded) were paid according to the minimum payment rules on Prolific (i.e., min. 5 GBP/hour).

3.4.5. Data preparation

The introductions, goals and how the goals were achieved were extracted from the responses and checked in several ways. First, the goals were checked to determine whether the collected goals were actual running or walking goals according to the following definition: *"A goal that involves running or walking and is achieved by running or walking"*. All goals were read and labeled "running or walking goal" or "other goal" by three people. Two raters were students with a background in computer science,

and the third rater was a student with a background in computer science and persuasion algorithms. For any disagreement, the majority rating was chosen. To measure inter-rater agreement, Fleiss' Kappa was calculated and resulted in a value of 0.99, which means almost perfect agreement [69]. Afterwards, the introductions, goals and how the goals were achieved sentences were corrected on several criteria, including anonymization, spelling, grammar, and punctuation mistakes. A complete overview of the criteria can be found in Appendix E.

3.4.6. Results

After cleaning and anonymizing the data, we were left with 72 examples of people who achieved a running or walking goal. Two examples of the resulting data are shown below:

- **Introduction:**
  "I'm Patricia and I really like playing tennis and walking at night. I set a challenge for me to walk at least 6.000 steps per day."
  Achieved goal and how they achieved it:
  "I walked every day at least 6.000 steps.
  I achieved this by just going for walks alone, with my boyfriend or my dog."

- **Introduction:**
  "Hello, I'm Edward. I'm 49 years old. I'm a person who loves to run and feel like I'm physically active and have good health"
  Achieved goal and how they achieved it:
  "I increased the number of daily kilometers I walk on a given week.
  I achieved this because I started with shorter distances, when I achieved 10 km I sped up."

3.5. Data collection part B: Ratings of examples

The goal of this part is to find out how people rate the collected examples of people who achieved a running or walking goal. Based on these ratings, a prediction model can be created to decide which example should be shown to a user who has a conversation with the virtual coach.

Participants were asked to give two types of ratings:

1. The introductions of the people of part A were rated based on similarity. The participant was asked to read the introduction and answer the question *How **similar** do you consider this person to yourself?* by filling out a scale from -3 to 3, where -3 was labeled as not similar, 0 was labeled as neutral and 3 was labeled as very similar.

2. The goals of the people of part A and how they achieved the goal were rated based on motivation. The participant was asked to read the goal and how they achieved it, and answer the question *How **motivating** do you consider this goal and how they achieved it?* by filling out a scale from -3 to 3, where -3 was labeled as not motivating, 0 was labeled as neutral and 3 was labeled as very motivating.

3.5.1. Participants

The participants in this experiment were fluent English-speaking adults (18 years and older). People who participated in experiment A were excluded from participation. The same 18 user variable filter groups that were used for recruitment in part A were also used for part B (Table D.1 in Appendix D). For part B, two people were recruited per group, resulting in a total of 36 participants. This number was chosen to ensure that we have two raters per group to reduce bias and was within the budget assigned for this part of the experiment.

Table B.2 in Appendix B shows the characteristics of the participants in part B.

3.5.2. Materials

Qualtrics was used to host the online questionnaires.

3.5.3. Measures

The user data collected in this experiment is the same as in experiment A, as the participants in both experiments were compared to each other. They were compared to each other because we wanted to measure how similar they are to each other, and we did this by calculating the difference of the independent variables between them. Questions asking for an introduction and running or walking goal and how they achieved it were not asked to the participants in part B, as these data were only needed from the participants of part A.

3.5.4. Procedure

The participants were asked to fill out a Qualtrics questionnaire. It started with the same questions to obtain user information as in part A, except for the questions about introducing themselves and describing a running or walking goal they achieved. Afterwards, participants were asked to rate the perceived similarity of the person and the motivational impact of the corresponding running or walking goal for 18 people). The 18 people were randomly chosen such that there was one person per group (see Table D.1 in Appendix D).

When a participant failed half or more of the attention check questions or did not complete the questionnaire, they were excluded and a new participant from the same group was recruited to replace them. The participants were paid according to the minimum payment rules on Prolific (i.e., min. 5 GBP/hour).

3.5.5. Data preparation

After parts A and B, we were left with 72 examples and approximately 9 ratings per example. 36 Participants from part B rated 18 examples each (on similarity and motivation), resulting in 648 data points for both similarity and motivation ratings. The data obtained from parts A and B needed to be cleaned and combined to have the appropriate data format for the data analysis. The following data indices were created:

- The self-efficacy scale index was obtained by summing the confidence ratings and dividing by the total number of items on the scale.
- The Godin-Leisure time score was calculated and classified into one of the three groups: active, moderately active, or insufficiently active [38].

- The Big-5 personality (10-item TIPI questionnaire) contained 2 items for each of the 5 personality dimensions. The answers given for the 2 items for a personality dimension were averaged (accounting for direction) to obtain the score for the given dimension.
- For the need for cognition and the physical activity self-identity, indices were computed by averaging the respective items (accounting for direction). To measure the internal consistency of these two variables, Chronbach's Alpha was calculated. The need for cognition questionnaire consisted of 3 items (α = .70) which is considered acceptable. The physical activity self-identity questionnaire consisted of 9 items (α = .91), which is considered excellent.

3.6. Model performance

In the data collection parts A and B, we collected examples from people in part A and let the people of part B rate the examples on how similar they perceived the example introductions, how motivating they considered the example goals and how the goal was achieved. These ratings were needed to create input for a prediction model. We need to predict which examples the virtual coach should show to a person having the goal-setting conversation. With the data obtained, a model could be fit to predict the perceived similarity and perceived motivation of the examples based on the user variables gathered in parts A and B.

3.6.1. Linear model

Linear regression was used to analyze the data. Two regressions were calculated:

- **Similarity model:** The first linear regression model was calculated to predict the dependent variable *similarity rating* based on the independent variables, which are the difference in user variables between the rater (person of part B) and the example person (person of part A) that the rater rated. *age, agreeableness, conscientiousness, education level, emotional stability, extraversion, gender, godin leisure time activity, household income, household size, need for cognition, openness to experiences, personal income, physical activity self-identity, running or walking self-efficacy, sitting hours weekday, sitting hours weekend day, smoking status, smoking frequency, socioeconomic status, ttm phase physical activity, and weekly exercise*.
- **Motivation model:** The second multiple linear regression model was calculated to predict the dependent variable *motivation rating* based on the same independent variables as in the similarity model.

These models were both analyzed to see which one would be better to use for the prediction of which example to show to a new user. In both cases, the multiple R^2 statistic was low (0.14 and 0.17 for the similarity and motivation model respectively), indicating that the models are not fitting the data very well, as only roughly 14% and 17% of the variance found in the dependent variables can be explained by the predictor variables. Since the model that predicts motivation ratings performed better, we decided to continue using that model. Details of the similarity and motivation models can be found in Appendix F in Tables F.1 and Table F.2.

3.6.2. Final prediction model

For the final model, we tried to reduce the number of variables necessary for the prediction without reducing the accuracy of the model. This was done to avoid overfitting and to decrease the number of variables required to predict the motivation rating. If fewer variables are required, users must answer fewer questions, which reduces the entry barrier to participate [52] and ultimately to make use of the virtual coach. Furthermore, we tried to use the similarity ratings data as input for the motivation rating model. All of these analyses are explained in Appendix G.

Final model

The final model (Table 3.2) contains the following variables: *age, extraversion, godin leisure time activity, household income, household size, openness to experiences, physical activity self-identity, running or walking self-efficacy, sitting hours weekend day, ttm phase physical activity, cluster 1, cluster 2, and cluster 3.*

Table 3.2: Final model used to predict motivation rating.

	Estimate	Std. Error	t value	Pr(>|t|)
age	-0.65	0.28	-2.33	0.02 *
household income	0.77	0.36	2.11	0.04 *
household size	0.60	0.33	1.84	0.07 .
extraversion	-0.73	0.33	-2.25	0.03 *
openness to experiences	-0.79	0.31	-2.58	0.01 *
ttm physical activity level	-0.53	0.27	-2.00	0.05 *
physical activity self-identity	-0.44	0.39	-1.13	0.26
running or walking self-efficacy	-0.82	0.30	-2.72	<0.01 **
sitting hours weekend day	0.57	0.37	1.55	0.12
godin activity	-0.89	0.18	-5.03	<0.001 ***
cluster1	0.13	0.05	2.43	0.02 *
cluster2	-0.01	0.05	-0.20	0.84
cluster3	0.40	0.06	6.79	<0.001 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The final model has a Multiple R^2 of 0.23 and Adjusted R^2 value of 0.21, which is slightly higher than the inital step model (0.22 and 0.21 respectively), and comparable to the full model (0.23 and 0.20) respectively). Thus in the final model, around 23% of the observed variation in the dependent variable can be explained by the independent variables.

This model is used to decide which examples the virtual coach shows to the user. For the new user, the motivation ratings of the examples are predicted and the two examples with the highest motivation ratings are chosen.

4

Evaluation

The aim of this chapter is to answer the following research question:

How effective is the designed goal-setting dialogue with regards to:

- *Running or walking self-efficacy*
- *Motivation*
- *Establishing a positive attitude towards the virtual coach*

In previous chapters, we described the design of a goal-setting dialogue with a virtual coach for physical activity. To evaluate the goal-setting dialogue, an experiment was set up. This chapter starts with describing the experimental setup, followed by the data analysis and results. The chapter ends with a discussion of the findings.

As we explained before, self-efficacy impacts the goals we choose, how we approach the goal, and goal-commitment. The virtual coach that we created was designed to increase the user's self-efficacy with regards to running and walking by showing examples of other people who successfully achieved a goal and by encouraging the user. To examine this, the following hypothesis was formulated:

- **H1:** The user's self-efficacy is higher after the conversation with the virtual coach than before the conversation with the virtual coach.

To predict which examples to show the user, a prediction model was fitted. This model predicts which example the user would consider most motivating. We wanted to analyze whether these examples chosen by the prediction model (personalized examples) are more effective in enhancing the participant's self-efficacy than when the virtual coach shows general examples. To evaluate this, we formulated the following hypothesis:

- **H2:** The self-efficacy is higher when people receive personalized examples than when they receive general examples.

Besides the indirect effect of the examples that might be observed in an increase of self-efficacy, we were also interested to directly analyze whether the personalized examples are considered more motivating than the general examples. The following hypothesis emerged from this:

- **H3:** The personalized examples are perceived as more motivating than the general examples.

Finally, we wanted to analyze how the people perceive the virtual coach, as this is important for behavior change in general. A better relationship with the virtual coach makes people like, trust and respect the virtual coach more, which leads to more positive behavior changes [103, 12]. A positive attitude towards the virtual coach has a positive impact on goal-setting as well, because a better relation with the coach can increase the goal commitment and shared goal feeling [11]. The following hypothesis was formulated:

- **H4:** People have a positive attitude towards the virtual coach.

In addition to analyzing these four hypothesis, we conducted a thematic analysis to find out what people find motivating about the example goals. We try to answer the following question: *What do people find motivating about the running or walking goals from other people?*

4.1. Methods

The experiment to evaluate the goal-setting dialogue was run in March 2022. Before running the experiment, we registered the design of this study with the Open Science Framework (OSF) [51]. Furthermore, the study design was approved by the TU Delft Human Research Ethics committee (HREC reference number: 1707).

4.1.1. Experimental Design

Study design

This study adopts a mixed study design with:

- 1 Between-subjects factor (the type of examples the virtual coach provides) and 2 levels (general examples or personalized examples), and
- 1 Within-subjects factor (time) measuring the running or walking self-efficacy twice, before the conversation with the virtual coach and after (referred to as pre and post respectively).

Randomization

Randomization was applied to the between-participants component of this study. Participants were randomly assigned to one of the two groups. The three variables of our prediction model that were able to best predict motivation ratings were considered for randomization. These were the baseline self-efficacy and two of the three cluster similarity rating variables.

To account for these three continuous variables in the randomization, we used adaptive covariate randomization [63]. For both groups, the means were calculated and covariance matrices were estimated. We tried 100 ways of assigning people to the two groups. For each of the 100 samples, two multivariate Gaussian distributions were estimated. To choose one of these 100 samples, Jeffrey's distance was computed between the two Gaussians of each sample. The sample with the smallest Jeffrey's distance (i.e. the pair of groups with the most similar multivariate Gaussians) was chosen. If

there were multiple samples with the smallest Jeffrey's distance, one of these samples was randomly chosen.

To ensure diversity between participants, 36 groups with different user variables were created beforehand and participants were recruited based on these groups. The groups filter on gender (male/female), age (age ranges 18-35, 36-55, and 55+), activity level (less than 60 minutes active per week, 60-120 minutes active per week, and more than 120 minutes active per week) and smoking status (smoking at least once a day or not). Prolific offers the ability to prescreen participants based on user variable criteria. The prescreening ensured that the complete sample was represented equally by all different groups. The combinations of the four variables gender (2 levels), age (3 levels), activity level (3 levels), and smoking status (2 levels) resulted in 36 (2x3x3x2) groups. The groups and their corresponding user variables can be found in Table D.2 in Appendix D.

4.1.2. Materials

Prolific was used to recruit participants. The questionnaires were hosted on Qualtrics. Furthermore, data from the previous data collection phase was used, including the examples of people that achieved a running or walking goal and the prediciton model. The chatbot was implemented using Rasa version 2.8.0[1]. The code for the chatbot can be found online [49]. The chatbot was running on a Google Cloud server[2].

The examples that were shown by the virtual coach during the goal-setting conversation were obtained in the data collection procedure that we described in the previous chapter (Section 3.3). We collected 72 examples in total. The three examples that received the overall highest motivation rating were used as the general examples. Each person received a random selection of two of these three examples. The three most motivating examples can be found in Appendix H.

Algorithm input variables

The same variables that we collected in experiments A and B are also gathered in this experiment as input for the prediction model. In addition, we asked participants to rate six introductions of example people based on how similar they consider that person. This is used as a predictor of motivation ratings. A more detailed description of this can be found in Appendix G. Note that the six examples they rated similarity were not used in the dialogue with the virtual coach, as the participants had seen part of the examples beforehand.

4.1.3. Measures

Primary variables

We identified the primary variables as variables that are relevant to answer the hypotheses. These are the following variables:

- Running- or walking self-efficacy, an adaptation of the exercise self-efficacy scale by McAuley [68], instead of exercise to specifically walking or running. The whole scale can be found in Appendix C.
- Acceptance of the virtual coach, an adaption of the questions asked in the paper by Bickmore et al. [13] that is adapted in the paper by Provoost et al. [81]. In the post-questionnaire, participants

[1] https://rasa.com/
[2] https://cloud.google.com/

were asked to fill out questions about the acceptance of the coach to assess their attitude towards and satisfaction with the virtual coach. For each of the six items, participants were asked to provide a rating (on a scale from -3 to 3) and then to elaborate by means of a free-text response for the question ”Why do you think so?”. The 6 items of the questionnaire assess the satisfaction with the conversational agent, ease/difficulty of talking to the conversational agent, preference regarding continuing or stopping to work with the conversational agent, relationship with the conversational agent, preference regarding working with the conversational agent or a questionnaire, and intended use of the conversational agent's advice in the future.

- Motivation ratings of examples. Participants were asked to rate the two examples they were given and the two examples that would have been given to them if they were in the other group on how motivating they perceived them (on a scale from -3 to 3 where -3 is not motivating and 3 is very motivating). This allows us to see whether personalized examples are considered more motivating. If one of the predicted personalized examples was the same as the general examples, they only rated that example once.

Secondary variables

The secondary variables concern the variables that were gathered for future research. Additionally, data from the dialogue between the participant and the virtual coach were saved for exploratory purposes. This includes the following:

- Type of goal the participant sets (running or walking).
- What the participants take away from the examples (free text response).

The other dialogue variables that were measured but not used in the analysis can be found in the OSF pre-registration form.

Finally, participants were asked an open question about the examples. After rating the example goals in the post-questionnaire, participants were asked the following question: *What do you find motivating about the running or walking goals that other people achieved?* This allows us to understand which elements of the examples are considered motivating so that they can be used positively in the future.

4.1.4. Participants

The participants for this experiment were fluently English speaking adults (age 18 and older). The people from whom we had collected example running or walking goals were excluded from participation.

To determine the sample size required for this experiment, we performed a power analysis using G*Power [33]. To obtain a medium effect size f of 0.25 [24], we used a power of 0.8 and 0.05 alpha error probability. The resulting required sample size was 34. We created 36 groups to balance participant recruitment, thus we decided to aim for two more participants for a total of 36 participants.

The experiment consisted of two parts: a pre-questionnaire and a post-questionnaire. Participants were excluded from the analysis after the first part of the experiment if they did not give informed consent, did not indicate they speak English fluently or failed more than 25% of the attention check questions of the pre-questionnaire. Participants were invited for part 2 if they successfully passed part

1. In part 2, participants were excluded from the analysis if they did not complete the conversation with the virtual coach, did not pass at least 50% of the attention checks of the post-questionnaire, or did not give sensible answers to free-text questions of the post-questionnaire.

Based on these criteria, eight participants were excluded: One participant did not give informed consent, two participants failed the pre-screening question about fluency in English, two participants did not finish the post-questionnaire, two participants did not return for the second part of the study within 1 week, and one participant did not complete the conversation with the virtual coach. We ended up with more than 36 participants because of two reasons. First, we did a pilot study beforehand to make sure that everything was working. The pilot study was successful and no changes were made afterwards, thus we added the pilot data to the data collected in the actual experiment. Second, one extra person was recruited by accident. In total, 47 participants were recruited of which 39 participants were included in the analysis. The participant characteristics can be found in Table 4.1. Participants were nationals of various countries, including Portugal, the United Kingdom, and South Africa.

Table 4.1: Participant characteristics of the two groups (General and Personalized) and both groups together (Total).

	General	**Personalized**	**Total**
	N = 20	N = 19	N = 39
Age *mean (SD), range*	44 (18), 19-72	39 (15), 22-66	42 (16), 19-72
Running or walking self-efficacy *mean, range*	86 (46, 100)	76 (60, 98)	84 (55, 100)
Godin Leisure-Time score *mean, range*	28 (8, 52)	29 (10, 44)	29 (9, 45)
Gender *n (%)*			
Female	10 (50%)	10 (53%)	20 (51%)
Male	10 (50%)	9 (47%)	19 (49%)
Weekly exercise (per week) *n (%)*			
Less than 60 minutes	8 (40%)	6 (32%)	14 (36%)
60-120 minutes	6 (30%)	6 (32%)	12 (31%)
More than 120 minutes	6 (30%)	7 (37%)	13 (33%)
TTM-phase for physical activity *n (%)*			
Maintenance phase	6 (30%)	5 (26%)	11 (28%)
Action phase	2 (10%)	3 (16%)	5 (13%)
Preparation phase	6 (30%)	3 (16%)	9 (23%)
Contemplation phase	3 (15%)	8 (42%)	11 (28%)
Precontemplation phase	3 (15%)	0	3 (7.7%)

Abbreviations: SD, Standard deviation; TTM, Transtheoretical model.

We checked whether there were differences between the two groups of the independent variable 'example type' (general and personalized). To do this, we used Bayesian t-tests and Bayesian test of

proportions. None of the variables had a certainly high probability of being different between the two groups according to Chechile's guidelines [21].

Table 4.2: Differences between the two groups (General and Personalized group). The *probability difference in means between the two groups > 0* shows the probability that the difference in means between the two groups is greater than 0. The *estimated probability to be a participant of the General group* shows the probability that the participant belongs to the 'General' group. For example, we see that participants who exercise less than 60 minutes per week have a probability of 56% to belong to the 'General' group, which means that there is a probability of 44% that the participant belongs to the 'Personalized' group.

Variable	**Probability difference in means between the two groups > 0 [95% CI]**
Running or walking self-efficacy	0.84
Age	0.85
Godin-leisure time activity	0.68
	Estimated probability to be in the General group [95% CI]
Smoking status	
Smoker	0.55
Non-smoker	0.33
Weekly exercise:	
Less than 60 minutes	0.56
60-120 minutes	0.50
More than 120 mintues	0.46
TTM-phase for physical activity	
Maintenance phase	0.54
Action phase	0.42
Preparation phase	0.64
Contemplation phase	0.30
Precontemplation phase	0.84

Abbreviations: CI, Credible interval, TTM: Transtheoretical model.

4.1.5. Procedure

Participants were recruited from Prolific. They received monetary compensation based on the payment rules on Prolific (i.e., min. 5 GBP/hour).

The experiment was divided into two parts:

1. The participants were first briefed about the nature of the experiment. They filled out the pre-questionnaire, which started with an informed consent form. Participants continued only if they provided informed consent. The pre-questionnaire gathered the user variables necessary to fit the prediction model, asked for ratings on examples of other people that achieved a running or walking goal, and measured running and walking self-efficacy.

   Participants were invited for the second part of the experiment after one week. Participants had a week to respond to the invitation and were excluded and replaced otherwise.

2. The second part was the conversation with the virtual coach (approximately 7 minutes), followed by a post-questionnaire to measure self-efficacy (either running or walking self-efficacy based on

whether the participant has set a running or walking goal during the conversation with the virtual coach), to measure the acceptance of the virtual coach, to get example ratings, and to gather feedback on what participants consider motivating about the example goals.

To give an overview, the experiment procedure is illustrated in Figure 4.1.

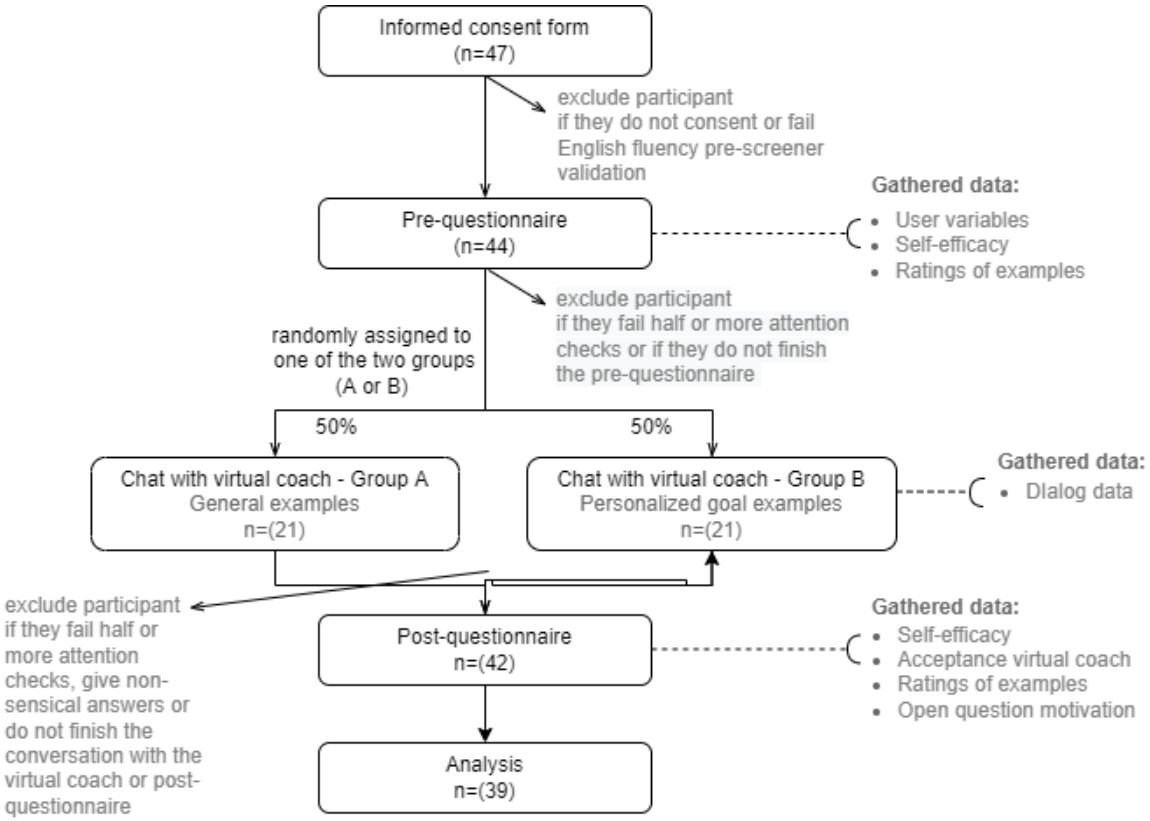


Figure 4.1: Experiment design showing the experiment process, the data gathered and exclusion criteria.

4.1.6. Data preparation and analysis

The collected data were cleaned in Python and analysed in R. All data and analysis code can be found in the 4TU ResearchData Repository [50] and the results can be reproduced with Docker. Data cleaning entailed anonymizing the data, removing data from excluded participants, and transforming and restructuring the data.

The transformations for the measures that were used for experiments A and B are explained in Section 3.5.5. Indices for the other measures were created as follows:

- For the similarity ratings given for the clusters, the participants rated two examples per cluster. The average of the two examples was taken to create one index value.
- For the motivation ratings given for the examples, the participant rated two examples per group (personalized examples and general examples). One index was created for this per group by taking the average value of the two ratings.
- For the acceptance questionnaire, we computed the indices by averaging the corresponding six items. To measure the internal consistency of this variable, Cronbach's α was computed. The acceptance questionnaire consisted of 6 items (α = .87) which is considered very reliable [26].

To analyze the four hypotheses, Bayesian hypothesis testing was used. This is in contrast to what was written in the OSF pre-registration form, where we described a frequentist analysis. We initially did the frequentist analysis, and the results were inconclusive for some hypotheses. Therefore, we reanalyzed the data using a Bayesian approach, which allowed us to draw conclusions about the hypotheses. Compared to the commonly used frequentist approach, Bayesian testing assigns a probability to a hypothesis [48]. We used Bayesian testing because the frequentist analysis results were inconclusive, and using the Bayesian approach gives us more insight about the data. Bayesian hypothesis testing shows the probability of whether a hypothesis holds, instead of only rejecting the hypothesis if the p-value is smaller than alpha.

The four hypotheses were analyzed as follows:

- To analyze the first hypothesis (H1), we performed a Bayesian paired t-test. The two measurements compared were the pre self-efficacy measurement and post self-efficacy measurement. For the self-efficacy measure, we compared the pre- and post-measurement of the participants' running or walking self-efficacy depending on which activity the participant set a goal for. Both running and walking self-efficacy were measured beforehand, because we did not yet know which activity the participants would choose. Only the self-efficacy of the activity that was chosen during the conversation with the conversational agent was measured afterward.
- To check H2, a Bayesian two-sample t-test was conducted. The two independent groups were the general and personalized examples groups, and the dependent variable was the change in self-efficacy between the pre and post measurement.
- To check whether personalized examples were perceived as more motivating than general examples (H3), a Bayesian two-sample t-test was performed for the motivation rating measure.
- For the last hypothesis (H4), we conducted a Bayesian one-sample t-test of the participants' responses to the acceptance questionnaire.

In addition to the Bayesian analysis to answer the four hypotheses, we conducted a thematic analysis to further analyze what participants considered motivating about the example goals they read. The themes were coded by two raters with a background in computer science. To measure inter-rater reliability, we calculated Cohen's Kappa which resulted in a value of 0.79, which means substantial agreement [69]. To reduce bias in the findings, we used triangulation [80]. We asked the participants two different questions through two different platforms to gather data for the thematic analysis.

4.2. Results

H1: The user's self-efficacy is higher after the conversation with the virtual coach than before the conversation with the virtual coach.

Figure 4.2 shows the differences in running or walking self-efficacy before (pre) and after (post) the conversation with the virtual coach. The difference is calculated by subtracting the pre measurement from the post measurement. On average, the participants had a lower self-efficacy after the conversation with the virtual coach (M = 56.68, SD = 28.11) than before the conversation (M = 71.61 , SD = 25.55). We looked at the posterior probability distribution of the self-efficacy change between pre and post. We found that the mean difference is less than 0 by a probability of >0.99 and more than 0 by a probability of <0.01. A mean paired difference of -15 is found with a 95% credibility interval of [-6, -24].

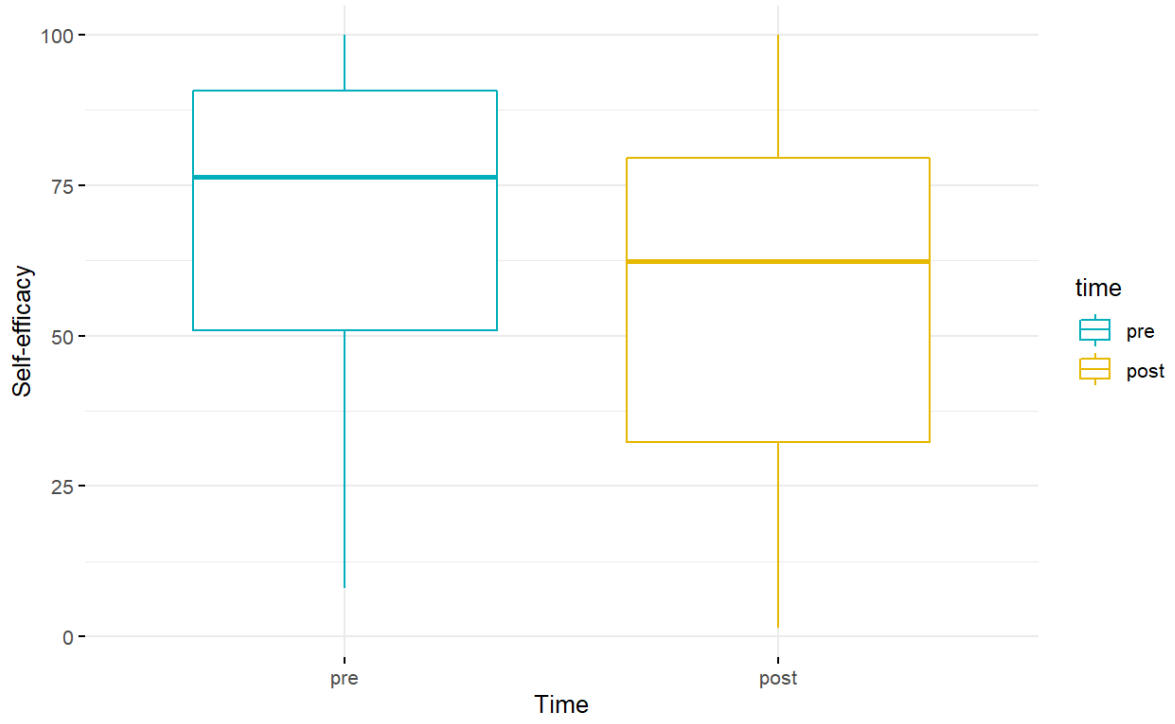


Figure 4.2: Comparison self-efficacy before (pre) and after (post) the conversation with the virtual coach.

H2: The self-efficacy is higher when people receive personalized examples than when they receive general examples.

The box plots in Figure 4.3 illustrate the differences in running or walking self-efficacy between the two groups (general and personalized examples). The difference is calculated by taking the self-efficacy measurements of the general group minus personalized group. There is a probability of 0.43 that the difference in means is greater than 0, and a probability of 0.57 that the difference in means is less than 0. The estimated mean difference in self-efficacy change between the two groups is -1.6 with a credibility interval of [-20, 17]. We observe a larger mean difference between the pre and post self-efficacy measurement for the general group, so the self-efficacy drops more for the general group. Table 4.3 gives an overview of the means, standard deviations, and difference in means of the general and personalized group.

Table 4.3: Mean, standard deviation, and mean difference of the self-efficacy of the two groups (general examples and personalized examples).

	Mean change pre- and post [95% CI]	**SD [95% CI]**	**Mean between-group diff [95% CI]**
General	-16 [-30, -1.6]	29 [20, 41]	1.6 [-17, 20]
Personalized	-14 [-27, -1.3]	26 [17, 37]	

Abbreviations: SD, Standard deviation; CI, Credible interval.

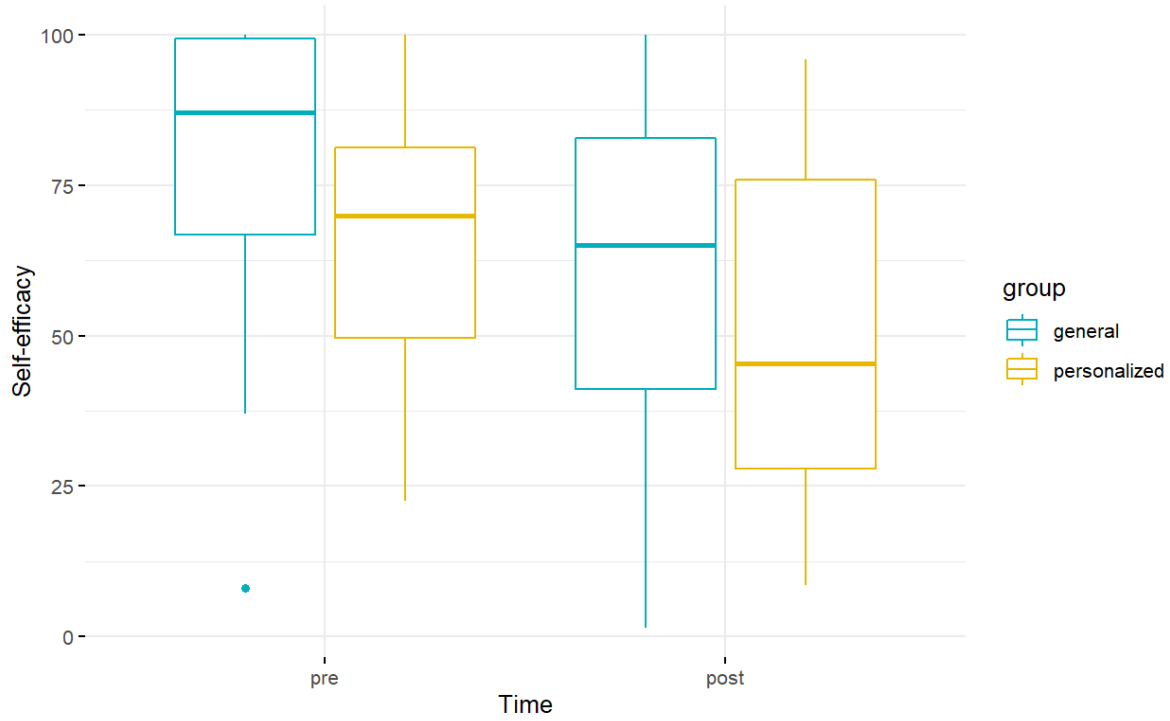


Figure 4.3: Comparison self-efficacy before (pre) and after (post) the conversation with the virtual coach per group (general or personalized examples).

H3: The personalized examples are perceived as more motivating than general examples.
The difference between means is greater than 0 by a probability of 0.80 and less than 0 by a probability of 0.20. We looked at the difference of given motivation ratings between the general examples minus the personalized examples. The estimated mean difference between the two groups is -0.3 with a credibility interval of [-0.9, 0.4]. Table 4.4 shows the means, standard deviations and the difference in means of the general and personalized examples. Figure 4.4 gives an illustration of the values of the table.

Table 4.4: Mean, standard deviation, and mean difference of the motivation ratings for general and personalized examples.

	Mean [95% CI]	**SD [95% CI]**	**Mean between-group diff [95% CI]**
General examples	1.2 [0.7, 1.6]	1.4 [1.1, 1.8]	-0.3 [-0.9, 0.4]
Personalized examples	1.5 [1.0, 1.9]	1.3 [1.0, 1.7]	

Abbreviations: SD, Standard deviation; CI, Credible interval.

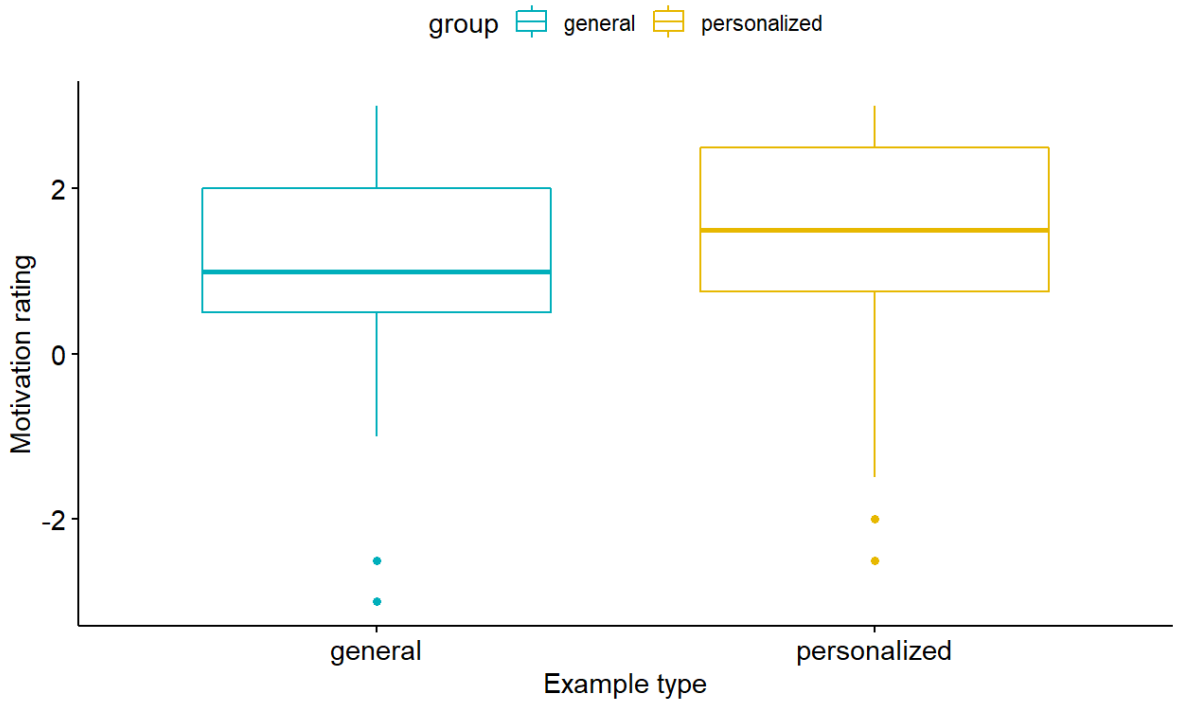


Figure 4.4: Comparison of motivation ratings between the two example types: General and Personalized.

H4: People have a positive attitude towards the virtual coach.

Looking at the posterior probabilities, we observe that the mean is more than 0 by a probability of >0.99 and less than 0 by a probability of <0.01 (M = 1.5, SD = 1.1) with a 95% credibility interval of [1.2, 1.9].

Table 4.5 shows the mean and standard deviation of each question of the acceptance questionnaire. Figure 4.5 illustrates the values of the table with box plots.

Table 4.5: Means and standard deviations for each item of the acceptance questionnaire. 'Average' shows the mean and standard deviation of all 6 questions combined (overall acceptance score).

	Mean [95% CI]	**SD [95% CI]**
Q1 - Satisfaction	1.9 [1.5, 2.3]	1.1 [0.9, 1.5]
Q2 - Ease of use	3.0 [3.0, 3.0]	0.0 [0.0, 0.0]
Q3 - Preference continuing or stopping	1.7 [1.2, 2.2]	1.4 [1.1, 1.8]
Q4 - Relationship with virtual coach	0.4 [0.0, 0.9]	1.3 [0.9, 1.7]
Q5 - Preference virtual coach or questionnaire	1.4 [0.8, 2.1]	1.7 [1.0, 2.3]
Q6 - Following advice	1.4 [0.8, 1.9]	1.6 [1.2, 2.0]
Average	1.5 [1.2, 1.9]	1.1 [0.8, 1.4]

Abbreviations: SD, Standard deviation; CI, Credible interval

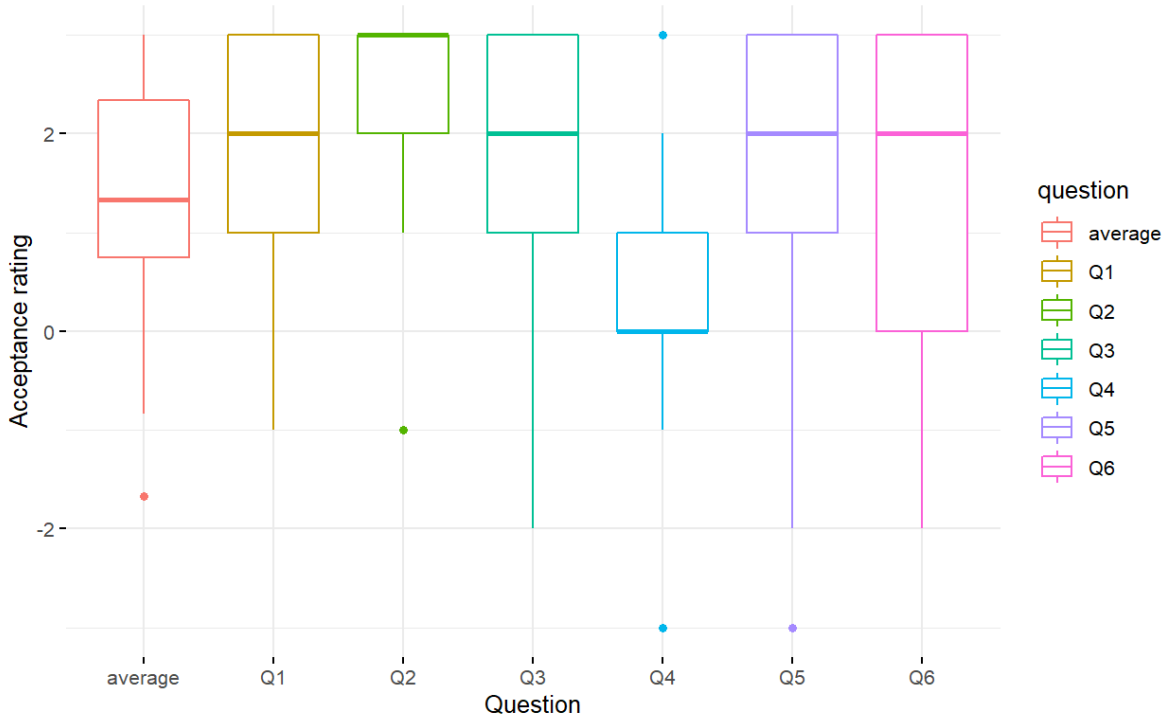


Figure 4.5: Box plots of the overall score (average) and each question (Q1-Q6) of the acceptance questionnaire. Questions: Q1 - Satisfaction, Q2 - Ease of use, Q3 - Preference continuing or stopping, Q4 - Relationship with virtual coach, Q5 - Preference virtual coach or questionnaire, Q6 - Following advice.

4.2.1. Qualitative results

Figure 4.6 shows the themes that we derived from the responses to the two open questions about what the participants found motivating about the example goals and what they could take away from the example goals. Four main themes were identified in the responses to these questions:

- **Goal-related:** The 'Goal-related' theme concerns all themes that are related to the contents of the example goal. Participants found it motivating that the goals were achievable and challenging. They also mention that they found it useful to see that the people from the examples wrote specific goals.
- **Path to goal:** The 'Path to goal' theme concerns everything that occurs in the process of reaching the goal. Participants liked reading that the people from the examples enjoyed the process of achieving their goal. They found it motivating that they did not give up and stayed consistent. Further, they reported to like the fact that the example people achieved their goal step by step.
- **Person-related:** Participants reported that they found it motivating to read that the example people that they related to were able to achieve certain goals.
- **Goal completion:** The fact that the examples represented people who successfully achieved a goal motivating. Participants believed that if the example people were able to achieve certain goals, they would be able to do the same.

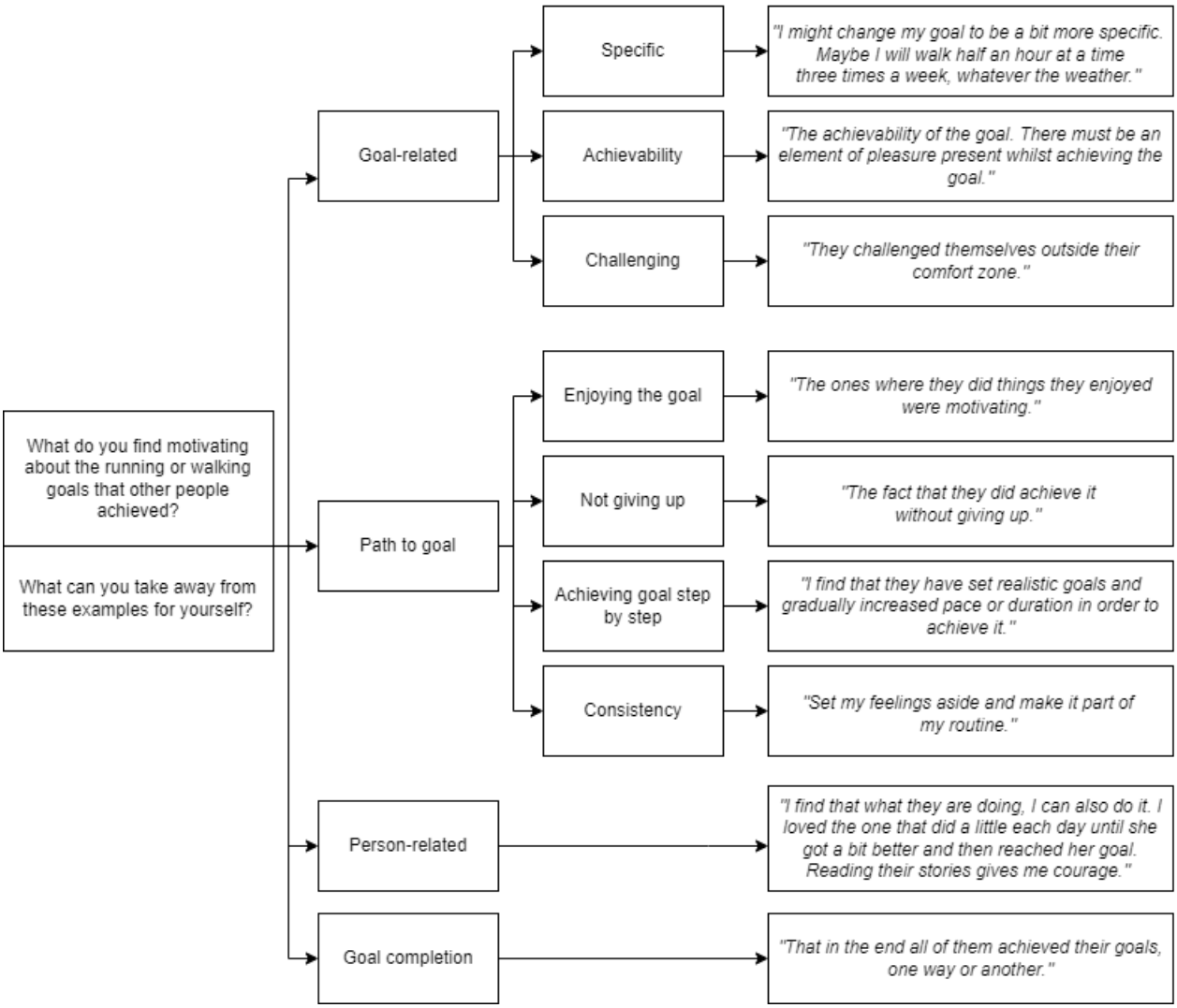


Figure 4.6: Thematic analysis about what participants found motivating about the example goals, including quotes from participants.

4.3. Discussion of results

H1: The user's self-efficacy is higher after the conversation with the virtual coach than before the conversation with the virtual coach.

The participants' running or walking self-efficacy decreased after the conversation with the virtual coach with a posterior probability of 99%, which is according to Chechile's criteria for posterior probabilities [21] a virtually certain probability. Our reasoning for this effect is as follows. Dunning and Kruger explain that people tend to overestimate themselves when people have little experience or knowledge regarding a certain task [58]. In our experiment, the pre-measurement of self-efficacy was measured one week before the conversation with the virtual coach. The post-measurement occurred directly after the conversation with the virtual coach. During the dialogue with the virtual coach, the participant is stimulated to think about multiple aspects of their goal, including their ability to achieve their goal. A possible explanation could be that the interaction with the virtual coach caused the participants to actively think about their abilities to achieve a running or walking goal and possibly judge their abilities

more realistically. It could be that they overestimated themselves before the conversation with the virtual coach due to the absence of knowledge and experience with running or walking goals [58]. During the conversation, the participants read examples of others who achieved a running or walking goal, giving the participants a better idea of what kind of running or walking goals could be successfully achieved.

A similar effect has been observed in the study by Kang et al. [55]. The authors explain that participants might have lower self-efficacy after their virtual reality experience because the participants realized that the activity they performed was more difficult than what they initially anticipated.

Another idea that is important to keep in mind is that other factors could decrease the participant's self-efficacy. It could be that the examples were considered motivating, but that the motivation effect was not sufficient to increase their self-efficacy. There are other factors that could affect their self-efficacy, such as the type of activity, the experience they have, or the difficulty of the goal they set.

H2: The self-efficacy is higher when people receive personalized examples than when they receive general examples.

When we look at the posterior probabilities of the Bayesian test we performed, we observed a probability of 57% that the personalized group has a higher increase in self-efficacy than the general group after the conversation with the virtual coach. Since in our case the self-efficacy decreases after the conversation with the virtual coach, this means that the personalized group has a probability of 57% that the decrease in self-efficacy is lower in the personalized group than in the general group after the conversation with the virtual coach. As the probability is close to 50%, we can draw the conclusion that these data do not suggest a difference between the two groups. According to Chechile [21], this probability is not worth betting on. We also see that the 95% credibility interval of the difference between groups has a wide range from negative to positive values [-17, 20], showing that there is no clear difference between the two groups. It could be that the effect of the example type was not a prominent part of the interaction with the virtual coach and therefore does not affect overall self-efficacy.

In Figure 4.3, we see a difference in means between the general and personalized group for the pre-measurement of self-efficacy. However, the two groups were balanced on their baseline self-efficacy to avoid bias. The groups were balanced based on self-efficacy for the activity they preferred. However, some participants set a goal for a different activity than the activity they indicated they prefer. This happened with 8 out of 39 participants. Because of that, the two groups were not as balanced as they initially were. We performed a Bayesian t-test to measure this difference (see Table 4.2). We found that there is a probability of 84% that the baseline self-efficacy of the two groups is different. According to Chechile's criteria for posterior probabilities, we cannot say that the difference is certain, but we could do a casual bet that there is a difference in self-efficacy between the two groups. This could also have affected the results of the self-efficacy measurements, as the personalized group started with a lower self-efficacy on average. In addition to the difference in baseline self-efficacy, we can see that the variables age and precontemplation phase have a similar probability of being different between the two groups. Although these variables probably do not affect self-efficacy as much as the running or walking self-efficacy measure, and we did not balance for these variables, there is a chance that it has an effect on self-efficacy.

H3: The personalized examples are perceived as more motivating than the general examples.

For the third hypothesis, we found that there is a probability of 79.4% that the personalized examples are rated higher than the general examples. According to Chechile's criteria for posterior probabilities

[21], we could say that we can do a casual bet that the personalized examples are more motivating than the general examples. We considered reasons why there is no pronounced difference in means between personalized and general examples. An explanation for this could be that the personalized examples are not personalized well enough. As we noticed in the previous chapter, the prediction model that was used to predict which personalized examples to show to the participant had a small R-squared value, namely 0.23. This indicates that the independent variables could only explain about 23% of the variation of the dependent variable, and thus the model could not predict the dependent variable very well. Another explanation could be that general examples were rated relatively high because these were the examples that were considered the most motivating overall. This means that the general examples that were shown can be considered to be examples that were motivating. For most people, the personalized examples they received were different from the general examples, which means that the predicted personalized examples were not the same as the general examples.

Finally, both types of examples have a positive estimated mean and a 95% credibility interval of [0.7-1.6] for the general examples and [1.0-1.9] for the personalized examples. Thus, it appears that the examples are generally perceived as motivating, as these values are all above zero.

H4: People have a positive attitude towards the virtual coach.

The results of the acceptance questionnaire indicate that participants have a positive attitude towards the virtual coach. The probability that the acceptance rating of the virtual coach is higher than 0 is 0.99, so we can assume that H4 holds according to the guidelines listed by Chechile [21] with a virtually certain probability. We cannot directly compare the questions with each other as they all measure different aspects of acceptance. However, we noticed that the second acceptance question about the ease of use of the virtual coach had a high score, indicating that participants had no difficulty talking to the virtual coach. A relatively low but still positive score was given to the question about the relationship with the virtual coach (Q4), indicating that the participants did not necessarily consider the virtual a very close friend.

Thematic analysis

With the thematic analysis, we tried to answer the question *What do people find motivating about the running or walking goals that other people achieved?* Participants considered it motivating when the goals were achievable and challenging. Moreover, they found it useful when the example goals were specific. These aspects related to the content of the goals are related to the principles of the SMART goal-setting framework [36]. It seems that when goals are in accordance with these principles, they are considered motivating.

Participants also indicated that they liked that people they can relate to are able to achieve running or walking goals. This is an important finding because it is in line with the purpose of the examples. The goal was to motivate people by showing them that other people could achieve running or walking goals, so people believed that they could also do it. This was based on Bandura's theory that vicarious experiences can increase people's self-efficacy. Many participants considered it motivating that the example people were able to complete running or walking goals, which supports the purpose of the examples.

4.3.1. Limitations

We identified the limitations of this experiment. First, we are not sure whether we can properly generalize for the appropriate population: the virtual coach is meant for people who want to set a goal for running or walking and are willing to improve their physical activity. The participants in our experiment were people who did not necessarily want to set a goal for running or walking, and they were not necessarily motivated to be more physically active. A reason for this is that the participants were in various stages of physical activity change. Some of the participants were already in the maintenance state, which means that they were already consistently active and most probably satisfied with their progress and therefore not necessarily motivated to run or walk more. It could be interesting to test the virtual coach only with people who intend to increase their physical activity by running or walking, so that the participants of the study are actually the target audience.

Second, our experiment is limited to only setting a goal, independent of the actual behavior. Therefore, in our study, we cannot see the effect of the dialogue with the virtual coach on actual physical activity behavior; we are limited to observing only their first perception of motivation. It might be interesting to study the effects of motivation over a longer period of time [45], combined with actual behavior. We only measured their self-efficacy and motivation once after the conversation with the virtual coach, but after multiple interactions with the virtual coach, these results might change.

5

Discussion and Conclusion

In this chapter, we discuss and conclude the work that has been done in this thesis. First, the research questions are presented and answered. Thereafter, the contributions of this thesis are discussed. The contributions are followed by a discussion on the limitations, and finally we give suggestions for future research.

5.1. Findings

The main research question we answered in this study is:

How can a goal-setting dialogue for a virtual coach be designed to motivate people in the context of physical activity?

To answer this main research question, we broke the question down into three subquestions that we have seen in the previous chapters. The answers to these subquestions are described below.

- *What are the requirements for a goal-setting dialogue with a virtual coach in the context of physical activity?*

After a literature study and expert consultation, we were able to formulate what was required to design a goal-setting dialogue with a virtual coach. We found that self-efficacy is important in goal-setting because it increases goal commitment [75] and motivation to change behavior. Self-efficacy can be enhanced by encouragement [101] and vicarious experiences [5]. R1 in Table 5.1 summarizes these findings. Furthermore, we found that setting SMART goals [36] is effective [77] and a good way to set complete goals (R2). The effects of personalization appeared to be useful to motivate the user, increase the user's attention, and ultimately enhance the chance of successful behavior change [83] (R3). Asking reflective questions increases the commitment to a goal [62] and helps the user understand why their goal is important to them [96] (R4). From the expert consultation, we found that explicitly asking the user if they believe their goal is achievable helps the user understand and consider whether their goal is realistic (R5). We found that this aspect of goal-setting was lacking in current approaches. Finally, we

agreed that the virtual coach should use easy to understand language with short and simple sentences [81] (R6).

Table 5.1: Requirements for the goal-setting dialogue.

R1	The virtual coach should motivate the patient to enhance their self-efficacy.
R2	The dialogue with the virtual coach should support the patient in creating specific, measurable, attainable, relevant and time-bound (SMART) goals.
R3	The dialogue with the virtual coach should be personalized.
R4	The dialogue with the virtual coach should make the user realize why the goal they are setting is important to them; the user should become committed to the goal.
R5	The dialogue with the virtual coach should make the patient consider whether the goal they are setting is achievable and realistic.
R6	The language used by the virtual coach should be simple, short and easy to understand.

- *How can a goal-setting dialogue with a virtual coach be designed in the context of physical activity?*

With the established requirements in mind, a goal setting dialogue was designed with a virtual coach named Jody. The purpose of this dialogue was to guide the user in setting a long-term running or walking goal. We found that self-efficacy impacts the goals we choose, how we approach the goal, and goal-commitment [65][6]. Thus, the virtual coach was designed to increase the user's self-efficacy by showing examples of other people that successfully achieved a goal and by encouraging the user. The examples that were shown consisted of two parts: an introduction of a person that achieved a goal, and the running or walking goal they achieved, including how they achieved it. Since we did not have access to such examples, we conducted two data-gathering experiments. To choose which examples Jody should present to the user, we fit a model that predicts which examples the user would consider most motivating. This is one way of personalization that is implemented in the dialogue, which was one of the requirements. Additionally, the direction of the dialogue changed based on the user responses. For example, if the user indicated that they consider their goal too difficult, Jody responded accordingly.

Setting SMART goals makes the goal more effective and result-oriented [77]. To support the user in setting SMART goals, the virtual coach asked specific questions to make the user consider each element of the SMART goal-setting theory. Jody asked reflective questions to make the user realize why their goal is important to them to increase relevance and consequently the commitment to the goal. Moreover, Jody asked the user whether they believe their goal is achievable, and supported them with changing their goal if they want to.

Finally, we considered the language of the virtual coach and made it easy to understand, as this was shown to be beneficial [81]. The language used by the virtual coach was simple, and long sentences were divided to make the messages more readable.

- *How effective is the designed goal-setting dialogue with regards to:*
  - Running or walking self-efficacy
  - Motivation

- Establishing a positive attitude towards the virtual coach

The implemented virtual coach was evaluated in an online experiment.

The results of this experiment indicate that there is no increase in self-efficacy after the conversation with the virtual coach. Instead, we observed a decrease in self-efficacy with a probability of 99%. An explanation that we considered for this effect is that the presence of a cognitive bias might have led the participants to overestimate their abilities. This could happen because during the conversation with the virtual coach, the user is asked to think about the running or walking goal they want to achieve and whether their goal is realistic or not, making them consider their abilities more seriously. With the pre self-efficacy measurement, the participants of the experiment were not inspired to think about their goal in detail, they were only asked whether they thought they would be able to achieve specific amounts of running or walking.

We observed that the personalized group has a probability of 57% that the decrease in self-efficacy is lower in the personalized group than in the general group after the conversation with the virtual coach, indicating that there is no certain difference in self-efficacy between the two groups.

With regard to how motivating the examples are considered, we found that there is a 80% chance that people will rate the personalized examples higher than the general examples. Both example types have a 95% credibility interval in a positive range, suggesting that the examples are generally considered motivating. The findings of the thematic analysis also showed that participants consider the examples motivating, as they indicated to like to see others achieving running or walking goals.

The results of the acceptance questionnaire indicate that the participants had a positive attitude towards the virtual coach. Jody especially scored high on ease of use, and the participants were satisfied with Jody. The lowest but still positive score was given to the question asking whether the participants considered Jody a close friend, which suggests that they consider Jody somewhat close, but not very.

5.2. Contributions

The first contribution of this project is the use of literature and the expert consultation to give an overview of the goal-setting theory and how it can be used by a virtual coach for physical activity. This information can be used in future research regarding conversational agents and goal-setting.

The next contribution is the dialogue with the virtual coach to set long-term running or walking goals. The conversation structure, as well as the goal-related questions that are asked by the virtual coach, can be used for other types of goal-setting interventions because it is not dependent on the goal activity type. Additionally, the virtual coach itself is a contribution, which can be used as a basis for further expansions to potentially allow for different types of goals or short-term goals.

Furthermore, a contribution of this project is the set of examples of people that have achieved a running or walking goal, and ratings on these examples on how motivating their goals are perceived. These examples can also be used for other research related to running or walking goals. In addition to this, the model that predicts motivation ratings can be useful for researchers to see how the independent variables affected motivation rating.

The findings of this study are a contribution as well. We analyzed and gained insight on how the dialogue with the virtual coach affects self-efficacy, the motivation of the examples, and the attitude towards the virtual coach. The results indicate that people have a positive attitude towards the virtual

coach, and that the gathered examples are perceived as motivating.

The virtual coach, example data, analysis, analysis data, and results can be consulted online.

5.3. Limitations

We recognize several limitations in our work. The first limitation is the prediction model that we fit. This model was used to decide which examples the virtual coach shows the user. For a new user, the motivation ratings of the examples were predicted and the two examples with the highest motivation ratings were chosen. In the final prediction model, about 23% of the observed variation in the dependent variable could be explained by the independent variables, which is not very high. Therefore, the accuracy of the predicted motivation ratings is limited. Consequently, the performance of the personalized examples is also limited, as these are the examples that are chosen based on the predicted motivation rating.

Second, self-reporting bias might be present in all experiments, which is a deviation between the true values of a measure and the self-reported value [8]. The variables we collected from the participants and the responses to the survey were all self-reported data. It could be that the participants responded differently due to cognitive processes, social desirability, and survey conditions [8]. For example, we asked questions about how often people exercise per week. The participants may have given an answer that is more accepted in society, although they might not have meant to do so.

Finally, with regard to motivation, we can distinguish two types of motivation: reflective motivation (including evaluation and plans) and automatic motivation (including emotions and impulses) [71]. We only consider reflective motivation when asking people to rate how motivating they consider the examples, so we are limited to this type of motivation, while there could have been effects on automatic motivation as well.

5.4. Future work

There are several aspects of this study that can be explored or expanded in future research. We gathered many additional variables that could impact perceived motivation or self-efficacy. These variables can still be explored to see whether they are covariates, so that we can find out how they affect perceived motivation or self-efficacy.

Future work could expand the virtual coach by for example including more types of physical activity to make it applicable in more situations. An even better expansion might be the ability to set short-term goals to be able to reach the long-term goal, to make the goal-setting experience more complete [7]. Long-term goals provide the end goal that people want to achieve, and short-term goals are useful to provide feedback on progress toward the long-term goal [7]. In addition to using the designed goal-setting dialogue for physical activity goals, it can be interesting to use it for different types of goals as well, as the goal-setting elements of the dialogue are not limited to physical activity.

The prediction model could be improved. The results indicated that the personalized examples have a higher chance of being considered more motivating, and we should not discard that probability. To improve the model, it could be useful to test other user variables to see their effect on perceived motivation. This could, for example, be variables related to people's culture or religion, as these characteristics were identified to influence how users perceive health recommendations [52], or variables that are not directly related to the user, such as the content of the message.

Moreover, different ways of choosing which example to show to the user could be tested. We have used a regression model to predict motivation ratings, but there are recommendation algorithms, such as collaborative filtering or hybrid filtering, [91] that could be tried out as well.

As personalization is considered to be useful in conversational agents [83, 57], other ways to personalize the dialogue could also be considered, such as incorporating past running or walking experiences in the dialogue flow and responses of the virtual coach.

5.5. Final remarks

In this thesis, we presented a virtual coach that supports users in setting effective long-term running or walking goals. The dialogue with the virtual coach was designed and evaluated on self-efficacy, motivation, and acceptance. The results indicate that people have a positive attitude towards the virtual coach, and that the gathered examples are perceived as motivating. Suggestions are made for future work to further explore the effectiveness of the virtual coach and the set of examples.

References

[1] *Adult stage of change (short form)*. URL: https://web.uri.edu/cprc/measures/smoking/adult-stage-of-change-short-form/.

[2] D. F. Anderson and C. M. Cychosz. “Development of an exercise identity scale.” In: *Perceptual and motor skills* 78 (3 Pt 1 1994). ISSN: 00315125. DOI: 10.1177/003151259407800313.

[3] *App Store - Apple (NL)*. https://www.apple.com/nl/app-store/. (Accessed on 03/01/2021).

[4] Stefanie Ashford, Jemma Edmunds, and David P. French. “What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis”. In: *British Journal of Health Psychology* 15 (2 2010). ISSN: 1359107X. DOI: 10.1348/135910709X461752.

[5] Albert Bandura. *Self-efficacy: The exercise of control.* New York, NY, US: W H Freeman Times Book Henry Holt and Co, 1997, pp. ix, 604–ix, 604. ISBN: 0-7167-2626-2 (Hardcover); 0-7167-2850-8 (Paperback).

[6] Albert Bandura. “Self-efficacy: Toward a unifying theory of behavioral change”. In: *Advances in Behaviour Research and Therapy* 1.4 (1978). Perceived Self-Efficacy: Analyses of Bandura’s Theory of Behavioural Change, pp. 139–161. ISSN: 0146-6402. DOI: https://doi.org/10.1016/0146-6402(78)90002-4. URL: https://www.sciencedirect.com/science/article/pii/0146640278900024.

[7] Dario Baretta et al. “Implementation of the goal-setting components in popular physical activity apps: Review and content analysis”. In: *DIGITAL HEALTH* 5 (2019), p. 2055207619862706. DOI: 10.1177/2055207619862706.

[8] Sebastian Bauhoff. “Self-Report Bias in Estimating Cross-Sectional and Treatment Effects”. In: *Encyclopedia of Quality of Life and Well-Being Research*. Ed. by Alex C. Michalos. Dordrecht: Springer Netherlands, 2014, pp. 5798–5800. ISBN: 978-94-007-0753-5. DOI: 10.1007/978-94-007-0753-5_4046. URL: https://doi.org/10.1007/978-94-007-0753-5_4046.

[9] Gökben Bayramoğlu et al. “Self-Efficacy in the workplace : Implications for motivation and performance”. In: *Journal of Management* 3 (2 2013). ISSN: 02788209.

[10] Tessa Beinema et al. “Tailoring coaching strategies to users’ motivation in a multi-agent health coaching application”. In: *Computers in Human Behavior* 121 (2021), p. 106787. ISSN: 0747-5632. DOI: https://doi.org/10.1016/j.chb.2021.106787. URL: https://www.sciencedirect.com/science/article/pii/S0747563221001102.

[11] Timothy Bickmore. “Relational Agents: Effecting Change through Human-Computer Relationships”. In: (Mar. 2003).

[12] Timothy W. Bickmore et al. “It’s just like you talk to a friend’ relational agents for older adults”. In: *Interacting with Computers* 17.6 (2005). HCI and the Older Population, pp. 711–735. ISSN:

0953-5438. DOI: https://doi.org/10.1016/j.intcom.2005.09.002. URL: https://www.sciencedirect.com/science/article/pii/S0953543805000755.

[13] Timothy W. Bickmore et al. “Response to a relational agent by hospital patients with depressive symptoms”. In: *Interacting with computers* 22 4 (2010), pp. 289–298.

[14] Craig S. Brinkman, Robert S. Weinberg, and Wrose Marie Ward. “The big five personality model and self-determined motivation in sport”. In: *International Journal of Sport Psychology* 47 (5 2016). ISSN: 00470767. DOI: 10.7352/IJSP2016.47.389.

[15] John T. Cacioppo and Richard E. Petty. “The need for cognition”. In: *Journal of Personality and Social Psychology* 42 (1 1982). ISSN: 00223514. DOI: 10.1037/0022-3514.42.1.116.

[16] John T. Cacioppo, Richard E. Petty, and Chuan Feng Kao. “The Efficient Assessment of Need for Cognition”. In: *Journal of Personality Assessment* 48.3 (1984), pp. 306–307. DOI: 10.1207/s15327752jpa4803_13.

[17] James E Cameron. “A Three-Factor Model of Social Identity”. In: *Self and Identity* 3 (3 2004). ISSN: 1529-8868. DOI: 10.1080/13576500444000047.

[18] Lorainne Tudor Car et al. “Conversational agents in health care: Scoping review and conceptual analysis”. In: *Journal of Medical Internet Research* 22 (8 2020). ISSN: 14388871. DOI: 10.2196/17158.

[19] *Cardiovascular diseases*. (Accessed on 02/01/2021). URL: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.

[20] J M Carroll. *Making use: scenario-based design of human-computer interactions*. 2000. ISBN: 0262032791.

[21] Richard A Chechile. *Bayesian Statistics for Experimental Scientists: A General Introduction Using Distribution-Free Methods*. MIT Press, 2020. ISBN: 0262044587; 9780262044585. URL: libgen.li/file.php?md5=de6fd45b9cc9332a6b52d4341cf1d4.

[22] Sung Hyeon Cheon, Johnmarshall Reeve, and Yong Gwan Song. “Recommending goals and supporting needs: An intervention to help physical education teachers communicate their expectations while supporting students’ psychological needs”. In: *Psychology of Sport and Exercise* (2019). ISSN: 14690292. DOI: 10.1016/j.psychsport.2018.12.008.

[23] Neel P. Chokshi et al. “Loss-framed financial incentives and personalized goal-setting to increase physical activity among ischemic heart disease patients using wearable devices: The ACTIVE REWARD randomized trial”. In: *Journal of the American Heart Association* (2018). ISSN: 20479980. DOI: 10.1161/JAHA.118.009173.

[24] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New York: Routledge, 1988.

[25] Terry L. Conway and Terry A. Cronan. “Smoking, exercise, and physical fitness”. In: *Preventive Medicine* 21 (6 1992). ISSN: 00917435. DOI: 10.1016/0091-7435(92)90079-W.

[26] Jose M. Cortina. “What Is Coefficient Alpha? An Examination of Theory and Applications”. In: *Journal of Applied Psychology* 78 (1993), pp. 98–104.

[27] Henry M Cothran and Allen F Wysocki. “Developing SMART Goals for Your Organization”. In: *Document No FE577, Florida Cooperative Extension Service, University of Florida* (2012).

[28] Björn Dahlöf. “Cardiovascular Disease Risk Factors: Epidemiology and Risk Assessment”. In: *American Journal of Cardiology* (2010). ISSN: 00029149. DOI: 10.1016/j.amjcard.2009.10.007.

[29] Tracy Epton, Sinead Currie, and Christopher J. Armitage. “Unique Effects of Setting Goals on Behavior Change: Systematic Review and Meta-Analysis”. In: *Journal of Consulting and Clinical Psychology* (2017). ISSN: 19392117. DOI: 10.1037/ccp0000260.

[30] “Examining the association between education level and physical activity changes during early old age”. In: *Journal of Aging and Health* 20 (7 2008). ISSN: 08982643. DOI: 10.1177/0898264308321081.

[31] Ahmed Fadhil and Silvia Gabrielli. “Addressing challenges in promoting healthy lifestyles”. In: *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare - PervasiveHealth ’17*. 2017. ISBN: 9781450363631.

[32] Haiyan Fan and Marshall Scott Poole. “What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems”. In: *Journal of Organizational Computing and Electronic Commerce* 16.3-4 (2006), pp. 179–202. DOI: 10.1080/10919392.2006.9681199. eprint: https://doi.org/10.1080/10919392.2006.9681199. URL: https://doi.org/10.1080/10919392.2006.9681199.

[33] Franz Faul et al. “G*Power 3.1.7: A flexible statistical power analysis program for the social, Behavioral and Biomedical sciences, Beh”. In: *Res. Meth.s* 39 (Jan. 2013), pp. 175–191.

[34] Jasper Feine et al. “Gender Bias in Chatbot Design”. In: vol. 11970 LNCS. 2020. DOI: 10.1007/978-3-030-39540-7_6.

[35] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. “Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial”. In: *JMIR Mental Health* (2017). ISSN: 2368-7959. DOI: 10.2196/mental.7785.

[36] Doran T George. “There’s a S.M.A.R.T. way to write management’s goals and objectives.” In: *Management Review* 70 (11 1981).

[37] Suparna Ghanvatkar, Atreyi Kankanhalli, and Vaibhav Rajan. “User Models for Personalized Physical Activity Interventions: A Scoping Review”. In: *JMIR mhealth and uhealth* 21 (May 2018). DOI: 10.2196/11098.

[38] Gaston Godin. “The Godin-Shephard Leisure-Time Physical Activity Questionnaire”. In: *Health and Fitness Journal of Canada* 4 (1 2011). ISSN: 19206216.

[39] Peter M. Gollwitzer, Heinz Heckhausen, and Heike Ratajczak. “From weighing to willing: Approaching a change decision through pre- or postdecisional mentation”. In: *Organizational Behavior and Human Decision Processes* (1990). ISSN: 07495978. DOI: 10.1016/0749-5978(90)90004-S.

[40] *Google Play Store*. https://play.google.com/store. (Accessed on 03/01/2021).

[41] Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann. “A very brief measure of the Big-Five personality domains”. In: *Journal of Research in Personality* 37 (6 2003). ISSN: 00926566. DOI: 10.1016/S0092-6566(03)00046-1.

[42] Jorne Grolleman et al. “Break the habit! designing an e-therapy intervention using a virtual coach in aid of smoking cessation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2006. ISBN: 3540342915. DOI: 10.1007/11755494_19.

[43] Jason W. Hart et al. “The big five and achievement motivation: Exploring the relationship between personality and a two-factor model of motivation”. In: *Individual Differences Research* 5 (4 2007). ISSN: 1541745X.

[44] *Health & Fitness Tracker with Calorie Counter - Apps on Google Play*. https://play.google.com/store/apps/details?id=com.droidinfinity.healthplus&hl=en_US&gl=US. (Accessed on 03/10/2021).

[45] Cameron Hecht, Stacy Priniski, and Judith Harackiewicz. “Understanding Long-Term Effects of Motivation Interventions in a Changing World”. In: vol. 20. Mar. 2019, pp. 81–98. ISBN: 978-1-78754-614-1. DOI: 10.1108/S0749-742320190000020005.

[46] Olivier A. Blanson Henkemans et al. “An online lifestyle diary with a persuasive computer assistant providing feedback on self-management”. In: *Technology and Health Care* 17 (3 2009), pp. 253–267. ISSN: 09287329. DOI: 10.3233/THC-2009-0545.

[47] Gholamreza Heydari et al. “Smoking and physical activity in healthy adults: A cross-sectional study in Tehran”. In: *Tanaffos* 14 (4 2015). ISSN: 17350344.

[48] Roger Higdon. “Hypothesis Testing, Bayesian vs Frequentist”. In: *Encyclopedia of Systems Biology*. Ed. by Werner Dubitzky et al. New York, NY: Springer New York, 2013, pp. 933–934. ISBN: 978-1-4419-9863-7. DOI: 10.1007/978-1-4419-9863-7_1183. URL: https://doi.org/10.1007/978-1-4419-9863-7_1183.

[49] Beyza Hizli. *Goal-setting dialogue for physical activity with a virtual coach: code.* Version 1.0.0. June 2022. DOI: 10.5281/zenodo.6647381. URL: https://github.com/PerfectFit-project/goal_setting_virtual_coach.

[50] Beyza Hizli, Nele Albers, and Willem-Paul Brinkman. “Data and code underlying the master thesis: Goal-setting dialogue for physical activity with a virtual coach”. In: (June 2022). DOI: 10.4121/20047328. URL: https://doi.org/10.4121/20047328.

[51] Beyza Hizli, Nele Albers, and Willem-Paul Brinkman. *Goal-setting dialogue for physical activity with a virtual coach*. Feb. 2022. DOI: 10.17605/OSF.IO/4DUWH. URL: osf.io/4duwh.

[52] Santiago Hors-Fraile et al. “Opening the Black Box: Explaining the Process of Basing a Health Recommender System on the I-Change Behavioral Change Model”. In: *IEEE Access* 7 (2019), pp. 176525–176540. DOI: 10.1109/ACCESS.2019.2957696.

[53] “International physical activity questionnaire: 12-Country reliability and validity”. In: *Medicine and Science in Sports and Exercise* 35 (8 2003). ISSN: 01959131. DOI: 10.1249/01.MSS.0000078924.61453.FB.

[54] Mira El Kamali et al. *Virtual Coaches for Older Adults’ Wellbeing: A Systematic Review*. 2020. DOI: 10.1109/ACCESS.2020.2996404.

[55] Ni Kang et al. “Self-identification with a Virtual Experience and Its Moderating Effect on Self-efficacy and Presence”. In: *International Journal of Human–Computer Interaction* 37.2 (2021),

pp. 181–196. DOI: 10.1080/10447318.2020.1812909. eprint: https://doi.org/10.1080/10447318.2020.1812909. URL: https://doi.org/10.1080/10447318.2020.1812909.

[56] S. Kiesler and L. Sproull. “Human values and the design of technology”. In: CLSI Publications, 1997. Chap. Social responses to “social” computers.

[57] Ahmet Baki Kocaballi et al. *The personalization of conversational agents in health care: Systematic review*. 2019. DOI: 10.2196/15360.

[58] Justin Kruger and David Dunning. “Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments”. In: *Journal of Personality and Social Psychology* 77 (Jan. 2000), pp. 1121–34. DOI: 10.1037//0022-3514.77.6.1121.

[59] Liliana Laranjo et al. *Conversational agents in healthcare: A systematic review*. 2018. DOI: 10.1093/jamia/ocy072.

[60] Gary P. Latham. “The motivational benefits of goal-setting”. In: *Academy of Management Executive* 18 (4 2004). ISSN: 10795545. DOI: 10.5465/AME.2004.15268727.

[61] K. Blaine Lawlor. “Smart Goals: How the Application of Smart Goals can Contribute to Achievement of Student Learning Outcomes”. In: *Journal of Developments in Business Simulation and Experiential Learning* (2012).

[62] Min Kyung Lee et al. “Personalization revisited: A reflective approach helps people better personalize health services and motivates them to increase physical activity”. In: *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2015. ISBN: 9781450335744. DOI: 10.1145/2750858.2807552.

[63] Yunzhi Lin and Zheng Su. “Balancing continuous and categorical baseline covariates in sequential clinical trials using the area between empirical cumulative distribution functions”. In: *Statistics in Medicine* 31.18 (2012), pp. 1961–1971. DOI: https://doi.org/10.1002/sim.5363. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.5363. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.5363.

[64] Edwin A Locke and Gary P Latham. “Breaking the Rules: A Historical Overview of Goal-Setting Theory”. In: *Advances in Motivation Science*. 2015.

[65] Edwin A. Locke. “Motivation through conscious goal setting”. In: *Applied and Preventive Psychology* (1996). ISSN: 09621849. DOI: 10.1016/S0962-1849(96)80005-9.

[66] Edwin A. Locke and Gary P. Latham. “Building a practically useful theory of goal setting and task motivation: A 35-year odyssey”. In: *American Psychologist* (2002). ISSN: 0003066X. DOI: 10.1037/0003-066X.57.9.705.

[67] P. Lockwood, C. H. Jordan, and Z. Kunda. “Motivation by positive or negative role models: Regulatory focus determines who will best inspire us.” In: *Journal of Personality and Social Psychology* 83 (4 2002), pp. 854–864. DOI: doi:10.1037/0022-3514.83.4.854.

[68] Edward McAuley. “Self-efficacy and the maintenance of exercise participation in older adults”. In: *Journal of Behavioral Medicine* 16 (1 1993). ISSN: 01607715. DOI: 10.1007/BF00844757.

[69] Mary L. McHugh. “Interrater reliability: The kappa statistic”. In: *Biochemia Medica* 22 (3 2012). ISSN: 13300962. DOI: 10.11613/bm.2012.031.

[70] Eline Meijer et al. “Socio-economic status in relation to smoking: The role of (expected and desired) social support and quitter identity”. In: *Social Science and Medicine* 162 (2016). ISSN: 18735347. DOI: 10.1016/j.socscimed.2016.06.022.

[71] Susan Michie, Maartje van Stralen, and Robert West. “The Behaviour Change Wheel: a new method for characterising and designing behaviour change interventions”. In: *Implementation science : IS* 6 (Apr. 2011), p. 42. DOI: 10.1186/1748-5908-6-42.

[72] Anouk Middelweerd et al. *Apps to promote physical activity among adults: A review and content analysis*. 2014. DOI: 10.1186/s12966-014-0097-9.

[73] Mayke Mol et al. “Behind the scenes of online therapeutic feedback in blended therapy for depression: Mixed-methods observational study”. In: *Journal of Medical Internet Research* (2018). ISSN: 14388871. DOI: 10.2196/jmir.9890.

[74] “New directions in goal-setting theory”. In: *Current Directions in Psychological Science* (2006). ISSN: 14678721. DOI: 10.1111/j.1467-8721.2006.00449.x.

[75] H.F. O’Neil and M. Drillings. *Motivation: Theory and Research*. 1994. URL: https://doi.org/10.4324/9780203052686.

[76] Jan O’Neill et al. “The SMART Goals Process”. In: *The Power of SMART Goals: Using Goals to Improve Student Learning*. 2006. ISBN: 193400992X.

[77] Osahon Ogbeiwi. “Why written objectives need to be really SMART”. In: *British Journal of Health Care Management* (2017). ISSN: 13580574. DOI: 10.12968/bjhc.2017.23.7.324.

[78] Ashley H. Oiknine et al. “Need for Cognition Is Positively Related to Promotion Focus and Negatively Related to Prevention Focus”. In: *Frontiers in Psychology* 12 (2021). ISSN: 16641078. DOI: 10.3389/fpsyg.2021.606847.

[79] George Papathanasiou et al. “Smoking and physical activity interrelations in health science students. Is smoking associated with physical inactivity in young adults?” In: *Hellenic Journal of Cardiology* 53 (1 2012). ISSN: 10117970.

[80] Michael Patton. “Enhancing the Quality and Credibilty of Qualitative Analysis”. In: *Health services research* 34 (Jan. 2000), pp. 1189–208.

[81] Simon Provoost et al. “Improving adherence to an online intervention for low mood with a virtual coach: study protocol of a pilot randomized controlled trial”. In: *Trials* (2020). ISSN: 17456215. DOI: 10.1186/s13063-020-04777-2.

[82] A. Rand. *Introduction to Objectivist epistemology.* New York, 1990.

[83] Barbara K. Rimer and Matthew W. Kreuter. “Advancing tailored health communication: A persuasion and message effects perspective”. In: *Journal of Communication* (2006). ISSN: 00219916. DOI: 10.1111/j.1460-2466.2006.00289.x.

[84] Amelia Romeo et al. *Can smartphone apps increase physical activity? systematic review and meta-analysis*. 2019. DOI: 10.2196/12053.

[85] Preethi R Sama et al. “An Evaluation of Mobile Health Application Tools”. In: *JMIR mHealth and uHealth* (2014). ISSN: 2291-5222. DOI: 10.2196/mhealth.3088.

[86] *Samsung Health | Apps & Services | Samsung NL*. https://www.samsung.com/nl/apps/samsung-health/. (Accessed on 03/10/2021).

[87] Badrul Sarwar et al. “Item-based collaborative filtering recommendation algorithms”. In: 2001. DOI: 10.1145/371920.372071.

[88] Robert A. Schnoll et al. “Increased self-efficacy to quit and perceived control over withdrawal symptoms predict smoking cessation following nicotine dependence treatment”. In: *Addictive Behaviors* 36 (1-2 2011). ISSN: 03064603. DOI: 10.1016/j.addbeh.2010.08.024.

[89] Dale Schunk and Maria Dibenedetto. “Self-efficacy and human motivation”. In: Nov. 2020. ISBN: 9780128226841. DOI: 10.1016/bs.adms.2020.10.001.

[90] Dale H. Schunk and Maria K. DiBenedetto. *Self-efficacy and human motivation*. 2021. DOI: 10.1016/bs.adms.2020.10.001.

[91] George Sielis, Aimilia Tzanavari, and George Papadopoulos. “A Review of Recommender Systems: Types, Techniques and Applications”. In: July 2014, pp. 329–339. ISBN: 9781466658882.

[92] “Social acceptance of negotiation support systems: Scenario-based exploration with focus groups and online survey”. In: *Cognition, Technology and Work* (2012). ISSN: 14355566. DOI: 10.1007/s10111-011-0181-8.

[93] *Stages of change (short form)*. URL: https://web.uri.edu/cprc/exercise-stages-of-change-short-form/.

[94] Wayne T. Steward et al. “Need for Cognition Moderates Responses to Framed Smoking-Cessation Messages”. In: *Journal of Applied Social Psychology* 33 (12 2003). ISSN: 00219029. DOI: 10.1111/j.1559-1816.2003.tb02775.x.

[95] *Virtuagym Fitness Tracker - Home & Gym - Apps on Google Play*. https://play.google.com/store/apps/details?id=digifit.virtuagym.client.android&hl=en_US&gl=US. (Accessed on 03/15/2021).

[96] *Wat er toe doet*. https://watertoedoet.info/. (Accessed on 03/03/2021).

[97] Alice Watson et al. “An internet-based virtual coach to promote physical activity adherence in overweight adults: Randomized controlled trial”. In: *Journal of Medical Internet Research* (2012). ISSN: 14388871. DOI: 10.2196/jmir.1629.

[98] *Weight Loss Coach - Reduce Body Fat & Lose Weight - Apps on Google Play*. https://play.google.com/store/apps/details?id=com.droidinfinity.weightlosscoach&hl=en_US&gl=US. (Accessed on 03/10/2021).

[99] Robert West and Susan Michie. “A brief introduction to the COM-B Model of behaviour and the PRIME Theory of motivation”. In: *Qeios* (2020). DOI: 10.32388/ww04e6.2.

[100] Jessica Fitts Willoughby and Shuang Liu. “Do pictures help tell the story? An experimental test of narrative and emojis in a health text message intervention”. In: *Computers in Human Behavior* (2018). ISSN: 07475632. DOI: 10.1016/j.chb.2017.10.031.

[101] Y. Joel Wong et al. “I Believe in You! Measuring the experience of encouragement using the academic encouragement scale”. In: *Journal of Positive Psychology* (2019). ISSN: 17439779. DOI: 10.1080/17439760.2019.1579357.

[102] E. H. Wu et al. “Advantages and Constraints of a Hybrid Model K-12 E-Learning Assistant Chatbot”. In: *IEEE Access* 8 (2020), pp. 77788–77801. DOI: 10.1109/ACCESS.2020.2988252.

[103] Jingwen Zhang et al. “Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint”. In: *J Med Internet Res* 22.9 (Sept. 2020), e22845. ISSN: 1438-8871. DOI: 10.2196/22845. URL: http://www.ncbi.nlm.nih.gov/pubmed/32996892.

A

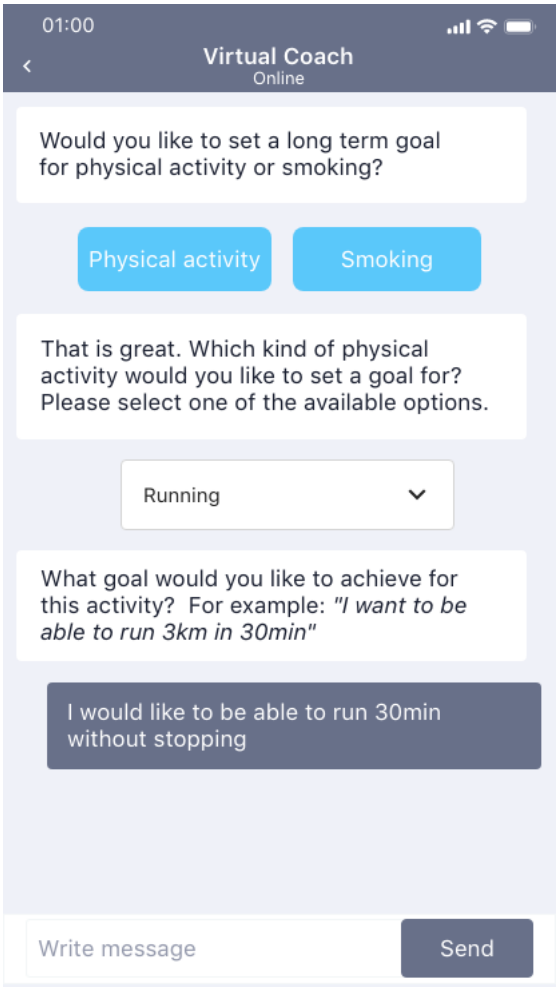
Scenarios

In this appendix, the scenarios that were used for the expert consultation can be found. The aim of the expert consultation was to find out and understand why experts would choose certain design choices over others. The scenarios were presented together with claims to emphasize the differences in the scenarios and to raise a discussion.

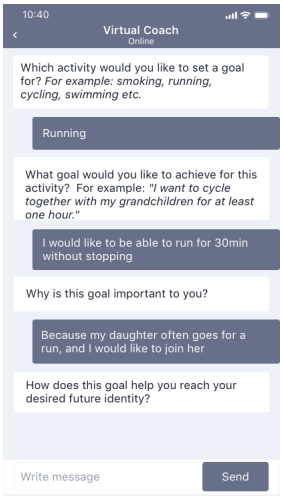
A.1. Scenario 1 - Long-term goal-setting

The first scenario is a use case example in which the user is talking with the virtual coach to set a long-term goal. Two options are given:

- In scenario 1A, the user has to choose whether they want to set a long-term goal for physical activity or smoking. The user selects physical activity and chooses the activity 'running' from the drop-down menu. Next, the virtual coach asks the user to write their goal down and gives an example.
- In scenario 1B, the coach starts with asking similar questions as in scenario 1A, but instead of selecting one of the given options, the user needs to write their answers down. In addition, the coach asks reflective questions about why the user wants to achieve a goal and how the user thinks this would help them reach their future identity.



Long-term goal-setting option A



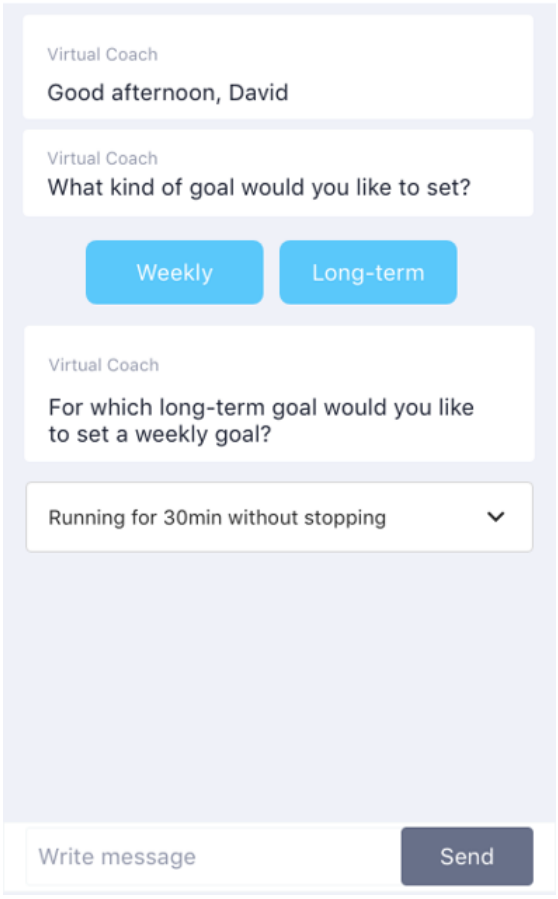
Long-term goal-setting option B

Figure A.1: In this scenario, two options are given to approach the long-term goal-setting part. In option A, the user needs to select the type of activity they are setting a goal for, and write down the long-term goal they would like to achieve. In option B, the user writes down the type of activity instead of selecting it. In addition, extra reflective questions are asked about why the goal is important to the user.

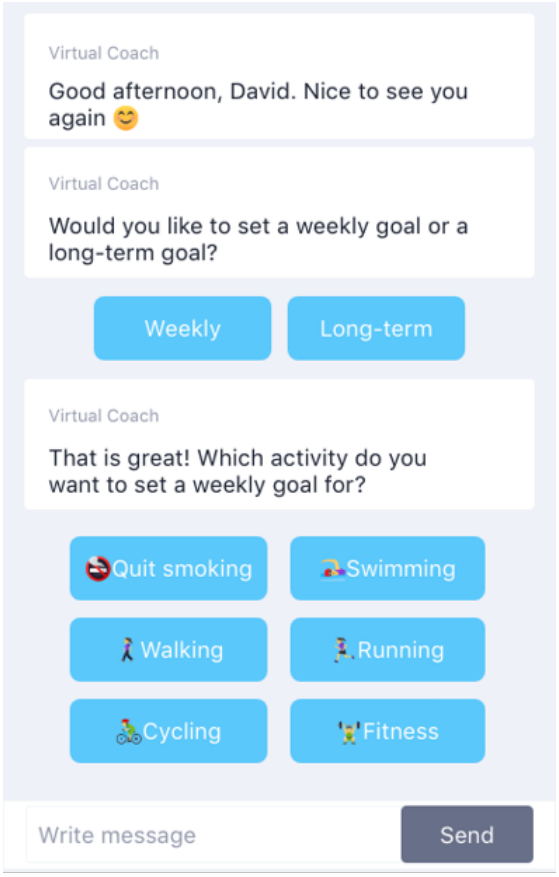
A.2. Scenario 2 - Opening weekly goal

The second scenario shows examples of opening dialogues for setting a weekly (short-term) goal. The user always sets a long-term goal before setting a weekly goal. Two examples are presented with two main differences:

- In scenario 2A, the coach is formally greeting the user and asking them how they want to proceed. Important to note here is that the virtual coach asks the user to set a weekly (short-term) goal that is part of a long-term goal. Thus, we are assuming that a long-term goal is part of a long-term goal.
- In scenario 2B, the coach asks the same questions but approaches the user in a more friendly, informal manner, also using emojis. Moreover, weekly goals can be set independently of the long-term goals.



Opening weekly goal option A



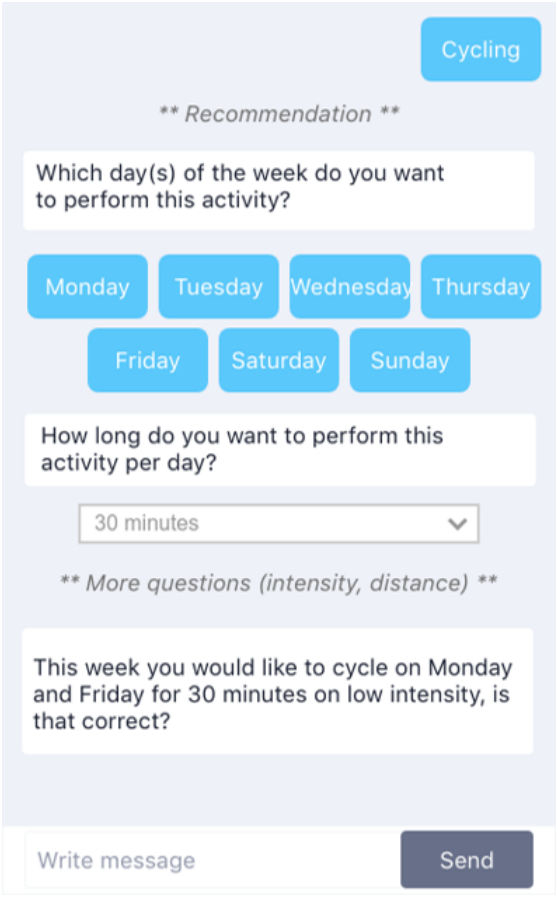
Opening weekly goal option B

Figure A.2: This figure shows the second scenario. In this scenario, two options are given to approach the weekly goal-setting part of the dialogue. In option A, the virtual coach is more formal and weekly goals are bound to a long-term goal. In option B, the virtual coach is more informal and weekly goals do not have to be part of a long-term goal.

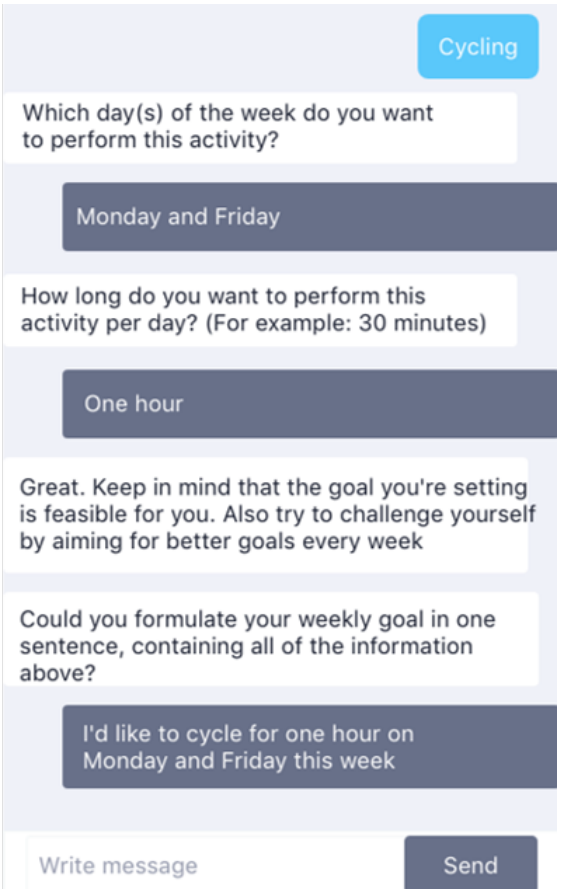
A.3. Scenario 3 - Weekly goal details

The third scenario is focused on the details of the weekly goal-setting.

- In scenario 3A, part of the weekly goal-setting process is shown. The user needs to select the days they want to perform the activity, and provide other relevant information such as the duration of their activity and the desired intensity. All questions are being answered by selecting one of the provided options. In the end, the virtual coach formulates a weekly goal based on the answers that are given by the user.
- In scenario 3B, the same questions are asked, but the user is required to answer the questions by writing their responses. Additionally, the coach is encouraging the user to keep in mind if the goal is feasible for them and to aim for higher goals every week.



Weekly goal details option A



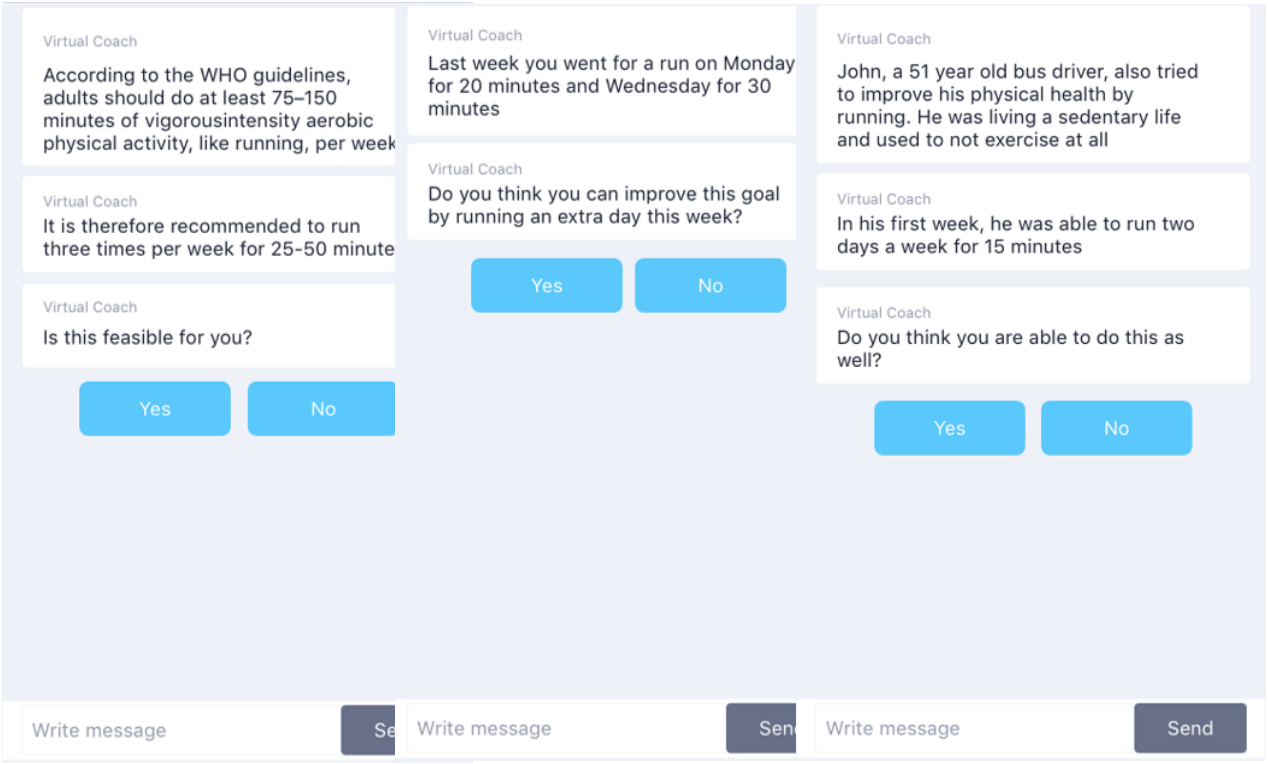
Weekly goal details option B

Figure A.3: In this scenario, two ways of setting the weekly goal details is shown. In option A, the user gets button options to set their goal. In option B, the user has to write their answers, and the virtual coach encourages the user to challenge themselves.

A.4. Scenario 4 - Weekly goal recommendations

The last scenario presents three different ways the virtual coach could recommend goals to the user. Recommending a goal might help the user to set reasonable goals and motivate them to aim for better goals [22].

- In scenario 4A, the weekly goal that is recommended is based on what health experts think is good for the user. This could be based on guidelines from the World Health Organization, but also what doctors say.
- In scenario 4B, the weekly goal recommendation is based on the user's performance in the previous week.
- In scenario 4C, the weekly goal is proposed to the user in the form of a narrative, describing a person in a similar position in terms of demographics, and physical activity and smoking.



Goal recommendation option A | Goal recommendation option B | Goal recommendation option B

Figure A.4: This figure shows three different ways of recommending weekly goals that the user could set. Option A uses WHO guidelines, option B uses previous progress and option C uses example achievements of other people.

B

Participants experiment A and B

		Number	%
Gender			
	Male	36	50
	Female	36	50
Weekly exercise			
	Less than 60 minutes active per week	24	33
	60-120 minutes active per week	24	33
	More than 60 minutes active per week	24	33
Smoking status			
	Smoker	51	71
	Non-Smoker	21	29
Household income range			
	Less than £10,000	12	17
	£10,000 - £15,999	8	11
	£16,000 - £19,999	6	8.3
	£20,000 - £29,999	19	26
	£30,000 - £39,999	8	11
	£40,000 - £49,999	4	5.6
	£50,000 - £59,999	5	6.9
	£60,000 - £69,999	2	2.8
	£70,000 - £79,999	4	5.6
	£80,000 - £89,999	1	1.4
	£90,000 - £99,999	1	1.4
	£100,000 - £149,999	1	1.4
TTM-phase for physical activity			
	Maintenance phase	25	35
	Action phase	18	25
	Preparation phase	11	15
	Contemplation phase	16	22
	Precontemplation phase	2	2.8
Household size			
	1	10	14
	2	11	15
	3	20	28
	4	21	29
	5	5	6.9
	6	4	5.6
	7	1	1.4
		Mean	**Range**
Age in years		43	(20, 74)
Hours spent sitting weekend day		5	(4, 8)
Godin Leisure-Time physical activity score		42	(28, 63)
Running or walking self-efficacy		77	(61, 92)
Personality			
	Extraversion	4	(3, 5.5)
	Openness to experiences	5	(4.5, 6)

Table B.1: Participant characteristics part A.
Abbreviations: TTM, Transtheoretical model.

		Number	%
Gender			
	Male	18	50
	Female	18	50
Weekly exercise			
	Less than 60 minutes active per week	12	33
	60-120 minutes active per week	12	33
	More than 60 minutes active per week	12	33
Smoking status			
	Smoker	29	81
	Non-Smoker	7	19
Household income range			
	Less than £10,000	6	17
	£10,000 - £15,999	10	28
	£16,000 - £19,999	4	11
	£20,000 - £29,999	7	19
	£30,000 - £39,999	5	14
	£40,000 - £49,999	2	5.6
	£50,000 - £59,999	1	2.8
	£70,000 - £79,999	1	2.8
TTM-phase for physical activity			
	Maintenance phase	10	28
	Action phase	6	17
	Preparation phase	10	28
	Contemplation phase	3	8.3
	Precontemplation phase	7	19
Household size			
	1	5	14
	2	15	42
	3	7	19
	4	4	11
	5	3	8.3
	6	1	2.8
	7	1	2.8

	Mean	Range
Age in years	42	(20, 71)
Hours spent sitting weekend day	8	(6, 10)
Running or walking self-efficacy	79	(45, 97)
Personality		
Extraversion	4.5	(3.5, 5.5)
Openness to experiences	5.0	(3.0, 6.5)

Table B.2: Participant characteristics part B.
Abbreviations: TTM, Transtheoretical model.

C

Self-efficacy questionnaire

Self-efficacy questionnaire for running:

Using the scales listed below please indicate how confident you are that you will be able to run at a moderate intensity for the given amount of minutes per week. When you're running at a moderate intensity, you breathe heavily but you can still hold a short conversation.

- How confident are you to run for 15 minutes per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to run for 30 minutes per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to run for 45 minutes per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to run for 60 minutes per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to run for 75 minutes per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to run for 90 minutes per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to run for 105 minutes per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to run for 120 minutes per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to run for 135 minutes per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to run for 150 minutes or more per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Self-efficacy questionnaire for walking:

Using the scales listed below please indicate how confident you are that you will be able to walk at a moderate intensity for the given amount of minutes per week. When you're walking at a moderate intensity, you breathe heavily but you can still hold a short conversation.

- How confident are you to walk for 30 minutes (0.5 hour) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to walk for 60 minutes (1 hour) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to walk for 90 minutes (1.5 hours) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to walk for 120 minutes (2 hours) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to walk for 150 minutes (2.5 hours) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to walk for 180 minutes (3 hours) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to walk for 210 minutes (3.5 hours) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to walk for 240 minutes (4 hours) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to walk for 270 minutes (4.5 hours) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

- How confident are you to walk for 300 minutes (5 hours) per week at a moderate intensity?

  Not at all confident — Moderately confident — Highly confident

  0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

D

Recruitment groups

Groups for the pre-screening of participants for the data collection experiments (experiment A and B) and the final experiment. Low activity indicates less than 60 minutes of physical activity per week, moderate activity indicates 60-120 minutes of physical activity per week and high activity indicates more than 120 minutes of physical activity per week. A smoker is defined as a person who smokes at least one tobacco product a day. A non-smoker is considered a person who does not smoke at least one tobacco product a day.

D.1. Recruitment groups experiment A and B

Table D.1: Groups that filter on user variables for balanced recruitment of participants. For experiment A, 72 participants were recruited (4 per group). For experiment B, 36 participants were recruited (2 per group).

ID	Female	Male	Age 18-35	Age 36-55	Age 55+	Low activity	Moderate activity	High activity
1	x		x			x		
2	x			x		x		
3	x				x	x		
4	x		x				x	
5	x			x			x	
6	x				x		x	
7	x		x					x
8	x			x				x
9	x				x			x
10		x	x			x		
11		x		x		x		
12		x			x	x		
13		x	x				x	
14		x		x			x	
15		x			x		x	
16		x	x					x
17		x		x				x
18		x			x			x

D.2. Recruitment groups final experiment.

Table D.2: Groups that filter on user variables for balanced recruitment of participants for the final experiment. We recruited 36 participants (1 per group) and 3 additional participants.

ID	Smoker	Non-smoker	Female	Male	Age 18-35	Age 36-55	Age 55+	Low activity	Moderate acitvity	High activity
1	x		x		x			x		
2	x		x			x		x		
3	x		x				x	x		
4	x		x		x				x	
5	x		x			x			x	
6	x		x				x		x	
7	x		x		x					x
8	x		x			x				x
9	x		x				x			x
10	x			x	x			x		
11	x			x		x		x		
12	x			x			x	x		
13	x			x	x				x	
14	x			x		x			x	
15	x			x			x		x	
16	x			x	x					x
17	x			x		x				x
18	x			x			x			x
19		x	x		x			x		
20		x	x			x		x		
21		x	x				x	x		
22		x	x		x				x	
23		x	x			x			x	
24		x	x				x		x	
25		x	x		x					x
26		x	x			x				x
27		x	x				x			x
28		x		x	x			x		
29		x		x		x		x		
30		x		x			x	x		
31		x		x	x				x	
32		x		x		x			x	
33		x		x			x		x	
34		x		x	x					x
35		x		x		x				x
36		x		x			x			x

E

Data cleaning criteria

Rewriting criteria:

These rewriting criteria were used for the goal data that we gathered in experiment A. The goal data are the introductions of the participants, the running or walking goal they achieved and how they achieved it.

- **Correct spelling mistakes.**
  Examples:
  acheive -> achieve
  when ever -> whenever
- **Correct grammar mistakes.**
  Examples:
  I has a friend. -> I have a friend.
  I achieved this with a passion. -> I achieved this with passion.
- **Correct punctuation mistakes.**
  Example:
  Hi I'm Lisa -> Hi, I'm Lisa.
- **Finish uncompleted sentences.**
  Example:
  10 km in under 4 hours. -> I walked 10 km in under 4 hours.
- **Replace participant names with common English names.**
  Example:
  Hi, my name is Elsa. -> Hi, my name is Mary.
- **Rewrite abbreviations and numbers for consistency.**
  Examples:

5 kilometers -> 5 km
6k -> 6.000
10000 -> 10.000
hrs -> hours

- **Write out uncommon abbreviations.**
  Example: NF1 -> Neurofibromatosis type 1

- **Add words that are assumed to be known for clarification.**
  Example:
  Walking 10.000 >steps< every day

- **Remove sentences directed to reader.**
  Example:
  Hi, my name is Mark. Pleased to meet you. How do you know (friends name)? -> Hi, my name is Mark.

Additionally for the ’How the goals are achieved’ data:

- **Adjust the sentence to make a full independent sentence clarifying that the sentence indicates how they have achieved their goal.** Add ’I achieved …’ if this is unclear.
  Example:
  By walking every day for 30 minutes. -> I achieved this by walking every day for 30 minutes.

F

Similarity and Motivation models

Similarity model

Table F.1: Summary of model predicting similarity rating with all independent variables. All the independent variables represent the difference of these variables, e.g. 'age' represents the difference in age.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.16	0.37	3.18	<0.01 **
age	-1.54	0.32	-4.90	<0.001 ***
education	0.03	0.34	0.10	0.92
household income	0.08	0.49	0.16	0.87
personal income	-0.76	0.56	-1.36	0.17
household size	0.16	0.36	0.44	0.66
gender	-0.49	0.15	-3.28	0.001 **
smoking frequency	-0.30	0.36	-0.85	0.40
socioeconomic status	-0.40	0.39	-1.01	0.31
weekly exercise	-0.32	0.21	-1.53	0.13
extraversion	0.07	0.36	0.21	0.84
agreeableness	0.25	0.37	0.67	0.50
conscientiousness	0.47	0.42	1.13	0.26
emotional stability	-0.27	0.39	-0.70	0.48
openness to experiences	-0.66	0.34	-1.94	0.05 .
need for cognition	0.25	0.36	0.71	0.48
ttm phase physical activity	-0.08	0.30	-0.25	0.80
physical activity self-identity	-2.10	0.44	-4.73	<0.001 ***
running or walking self-efficacy	-0.18	0.34	-0.53	0.59
sitting hours weekday	1.36	0.47	2.89	<0.01 **
sitting hours weekend day	-0.72	0.43	-1.68	0.09 .
godin activity	-0.34	0.20	-1.73	0.08 .
smoking status	0.49	0.16	3.08	<0.01 **
—				
Residual standard error:	1.889 on 625	degrees of freedom		
Multiple R-squared:	0.1422,	Adjusted R-squared:	0.112	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abbreviations: TTM, Transtheoretical model.

Motivation model

Table F.2: Summary of model predicting motivation rating with all independent variables. All the independent variables represent the difference in these variables, e.g. 'age' represents the difference in age.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.77	0.34	5.14	<0.001 ***
age	-0.85	0.30	-2.88	<0.01 **
education	0.13	0.32	0.40	0.69
household income	0.54	0.46	1.18	0.24
personal income	0.07	0.53	0.13	0.90
household size	0.53	0.34	1.55	0.12
gender	-0.06	0.14	-0.40	0.69
smoking frequency	0.11	0.34	0.32	0.75
socioeconomic status	-0.04	0.37	-0.11	0.91
weekly exercise	0.11	0.20	0.56	0.58
extraversion	-1.02	0.34	-3.00	<0.01 **
agreeableness	0.54	0.35	1.55	0.12
conscientiousness	0.19	0.39	0.48	0.63
emotional stability	-0.07	0.37	-0.20	0.84
openness to experiences	-0.96	0.32	-2.96	<0.01 **
need for cognition	0.15	0.34	0.45	0.65
ttm phase physical activity	-0.71	0.28	-2.49	0.01 *
physical activity self-identity	-0.69	0.42	-1.66	0.10 .
running or walking self-efficacy	-0.97	0.32	-3.04	<0.01 **
sitting hours weekday	0.47	0.44	1.07	0.29
sitting hours weekend day	0.61	0.40	1.51	0.13
godin activity	-0.97	0.18	-5.28	<0.001 ***
smoking status	0.27	0.15	1.80	0.07 .
—				
Residual standard error:	1.776 on 625	degrees of freedom		
Multiple R-squared:	0.1699,	Adjusted R-squared:	0.1407	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abbreviations: TTM, Transtheoretical model.

G

Model analysis

G.1. Cluster analysis

We decided to use the model predicting motivation ratings, however, we wanted to check whether the similarity ratings could be used to potentially improve the accuracy of the model. The correlation between motivation rating and similarity rating was analyzed. Similarity rating and motivation rating were found to be moderately positively correlated $r(648) = .33, p =< 0.001$. As we were aware of the positive correlation between the two variables, a way of using the similarity ratings as input for the motivation model was considered.

The idea was to put similar people of part A in groups, and get the average similarity rating for each group. If the people of part A are clustered based on similarity, and a mean rating for each cluster is found, these mean ratings per cluster can be used as predictors for the motivation model. The clusters are created based on similarity ratings, thus people of part A that are similarly rated belong to the same cluster. To analyze this, the following steps were taken:

1. The clusters were created based on similarity ratings, however, not all raters (participants of part B) rated all example persons (people of part A). This means that there is missing data, which is a problem, as it is necessary for the raters to have a rating for every person of part A. Item-based collaborative filtering was used to predict the missing ratings, which is a recommendation method looking for similar items (in this case items are examples) based on the items users have already rated [87]. So the missing values were predicted based on ratings that were already given to other examples.

2. The example persons of part A were clustered based on their similarity ratings using k-means clustering. This resulted into 3 clusters, as a larger number caused too much overlap between the clusters.

3. The two most centered people are looked up per cluster, and the average rating is calculated per cluster for all raters. Thus every rater has a rating for each cluster.

After these steps, three new predictors variables were created: *cluster 1, cluster 2* and *cluster 3*. These new variables were used as predictors in the motivation model, slightly increasing the Multiple R^2 value to 0.23 (from 0.17 before) and Adjusted R^2 to 0.20 (from 0.14 before). Although those values are still low, there is improvement.

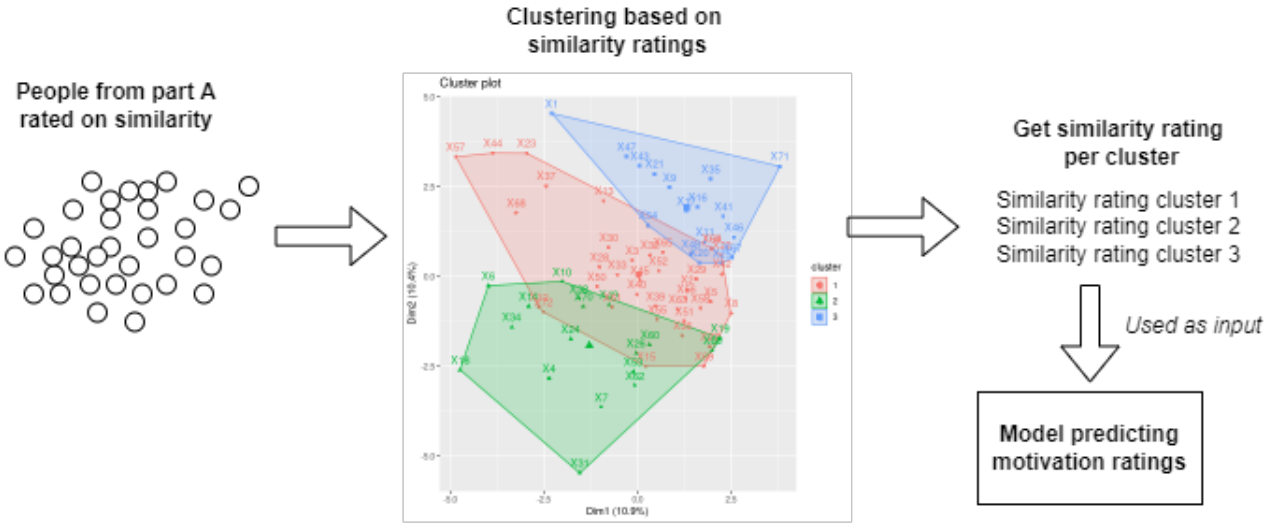


Figure G.1: Illustration of the clustering procedure. The people from part A were clustered based on similarity, resulting into the three clusters shown in the figure. For each cluster, one similarity rating is computed per cluster for the people of part B by taking the average rating of the two most centered people of that cluster. These similarity ratings for the clusters are used as input for the prediction model that predicts motivation ratings for the participants. This way, we made use of the similarity ratings in the prediction of motivation ratings.

To reproduce these average similarity ratings per cluster for the new participants in the last experiment, the two most centered examples per cluster were taken out of the total example set and presented to these participants in the pre-questionnaire. Participants were asked to rate these examples on similarity (2 examples per cluster, 6 in total), and the average of the two similarity ratings they gave per cluster was used as the similarity rating for that cluster. Note that these 6 examples were not used in the dialogue with the virtual coach, as the participants had seen the examples beforehand.

Table G.1: Model using all available variables to predict motivation rating. All the independent variables represent the difference in these variables, e.g. 'age' represents the difference in age.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.07	0.45	-0.14	0.89
age	-0.70	0.29	-2.42	0.02 *
education	0.13	0.31	0.42	0.68
household income	0.81	0.45	1.80	0.07 .
personal income	-0.22	0.51	-0.44	0.66
household size	0.60	0.34	1.78	0.08 .
gender	-0.05	0.14	-0.38	0.71
smoking frequency	-0.02	0.33	-0.05	0.96
socioeconomic status	0.10	0.36	0.27	0.79
weekly exercise	0.05	0.19	0.24	0.81
extraversion	-0.74	0.33	-2.24	0.03 *
agreeableness	0.31	0.34	0.90	0.37
conscientiousness	0.01	0.38	0.02	0.99
emotional stability	0.09	0.36	0.24	0.81
openness to experiences	-0.78	0.31	-2.47 *	0.01 **
need for cognition	-0.002	0.33	-0.008	0.99
ttm phase physical activity	-0.52	0.28	-1.87	0.06 .
physical activity self-identity	-0.55	0.41	-1.34	0.18
running or walking self-efficacy	-0.75	0.31	-2.42	0.02 *
sitting hours weekday	0.38	0.43	0.88	0.38
sitting hours weekend day	0.44	0.39	1.11	0.27
godin activity	-0.89	0.18	-4.98	<0.001 ***
smoking status	0.14	0.15	0.96	0.34
cluster1	0.13	0.05	2.36	0.02 *
cluster2	-0.008	0.05	-0.16	0.88
cluster3	0.39	0.06	6.32	<0.001 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abbreviations: TTM, Transtheoretical model.

G.2. Step model

Stepwise regression was applied as a first attempt to reduce the number of required variables. The predictors were iteratively added and/or removed in the predictive model to find the subset of variables in the data set resulting in the best performing model, that is a model that lowers prediction error. Stepwise selection (sequential replacement) was used, which is a combination of forward and backward selection. See Table G.2 for the selected variables.

Table G.2: Stepwise regression model predicting motivation rating. All the independent variables represent the difference in these variables, e.g. 'age' represents the difference in age.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.06	0.37	0.16	0.88
age	-0.66	0.28	-2.36	0.02 *
household income	0.70	0.36	1.94	0.05 .
household size	0.55	0.32	1.70	0.09 .
extraversion	-0.75	0.32	-2.33	0.02 *
openness to experiences	-0.79	0.31	-2.59	0.01 **
ttm phase physical activity	-0.64	0.25	-2.54	0.01 *
running or walking self-efficacy	-0.83	0.30	-2.75	0.01 **
sitting hours weekend day	0.55	0.36	1.53	0.13
godin activity	-0.90	0.17	-5.18	<0.001 ***
cluster1	0.14	0.05	2.68	0.01 **
cluster3	0.40	0.06	6.83	<0.001 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abbreviations: TTM, Transtheoretical model.

The variables that were selected by the step model were similar to the significant predictors from the full motivation model, except for the *household income* and *number of hours spent sitting in the weekend* variables. This model has a Multiple R^2 of 0.22 which is a slight decrease compared to the full motivation model model, and a slightly higher Adjusted R^2 value of 0.21, which was expected as there are less variables than in the full motivation model.

G.3. Correlation analysis

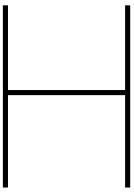
One problem with the step model is that when two independent variables are highly correlated with each other, the step model randomly picks one of the two. This was not desired, as we wanted to use all variables that are highly correlated with the dependent variable (motivation rating). We did not want to discard potential effects that would be lost if one of the two variables was discarded. Therefore, we looked at the correlations between all independent variables and the dependent variable, and included the variables that were significantly correlated to the variables selected by the step model. The correlations can be found in Table G.3.

Table G.3: Correlation between the independent variables and dependent variable (motivation rating). All the independent variables represent the difference in these variables, e.g. 'age' represents the difference in age.

variable	correlation	p-value	
age	-0.11	<0.01	**
agreeableness	0.09	0.03	
conscientiousness	-0.02	0.70	
education	-0.004	0.91	
emotional stability	0.02	0.60	
extraversion	-0.13	<0.001	***
gender	-0.02	0.62	
godin leisure time activity	-0.26	<0.001	***
household income	0.05	0.19	
household size	0.05	0.18	
need for cognition	0.004	0.91	
openness to experiences	-0.09	0.017	.
personal income	0.02	0.65	
physical activity self-identity	-0.13	<0.001	***
running or walking self-efficacy	-0.22	<0.001	***
sitting hours weekday	0.03	0.46	
sitting hours weekend day	0.01	0.87	
smoking status	0.09	0.03	.
smoking frequency	0.03	0.45	
socioeconomic status	-0.02	0.62	
ttm phase physical activity	-0.20	<0.001	***
weekly exercise	-0.03	0.51	
cluster1	0.14	<0.001	***
cluster2	0.04	0.32	***
cluster3	0.25	<0.001	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abbreviations: TTM, Transtheoretical model



Best rated examples

The three examples that received the highest motivation rating are the following examples (in no particular order):

- Here is how the person introduces themselves:
  ”I enjoy the outdoors. My hobbies are building puzzles and riding horses. I’m easy to talk to and I listen well.”
  This person achieved the following goal:
  ”I have walked more than 10.000 steps in half a day. I achieved this by walking everywhere I needed to be.”

- Here is how the person introduces themselves:
  ”Hello, I like to meet new people. One of my hobbies is watching tv. I am trying to stay in shape by walking.”
  This person achieved the following goal:
  ”I set a goal to walk every day regardless of weather factors for 1 straight month, and I achieved the goal. I achieved this by setting all my feelings aside and forced myself to ensure I met this goal.”

- Here is how the person introduces themselves:
  ”I started having walks around country lanes which is nicer than walking alone on a road. I prefer to look at sheep and cows rather than cars and buses.”
  This person achieved the following goal:
  ”I increased the time that I am walking by walking further. I achieved this by building up my walks by doing a little bit more each month.”