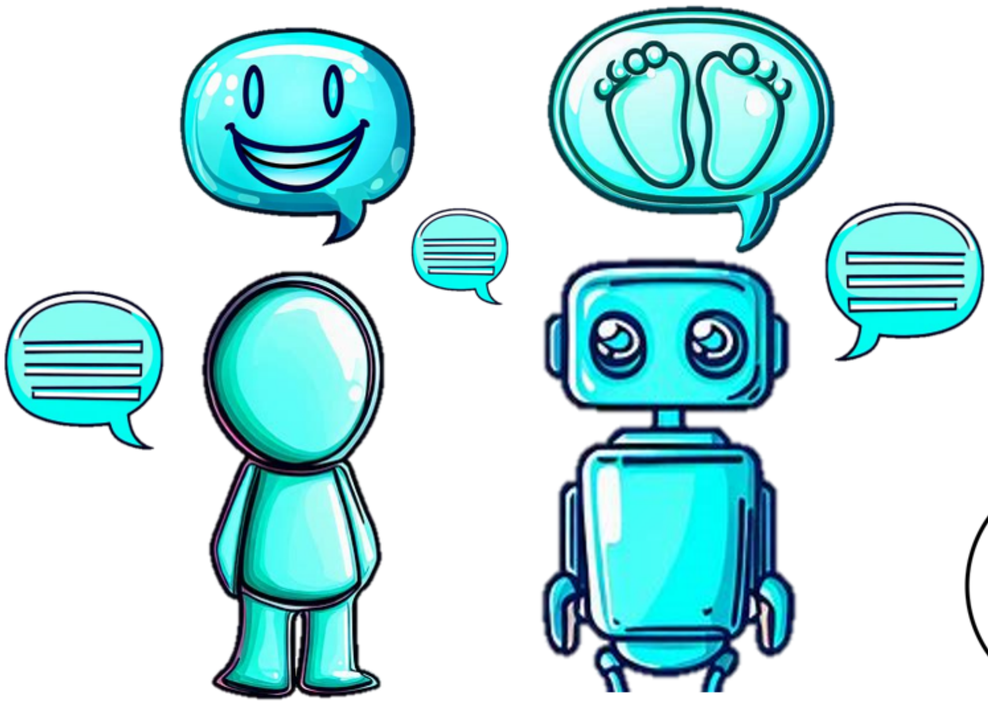
Using Reinforcement Learning to Personalize Daily Step Goals for a Collaborative Dialogue with a Virtual Coach

Martin Dierikx



Perfect Fit

TU Delft

Using Reinforcement Learning to Personalize Daily Step Goals for a Collaborative Dialogue with a Virtual Coach

by

Martin Dierikx

Institution:	Delft University of Technology
Place:	Faculty of Electrical Engineering, Mathematics and Computer Science, Delft
Project duration:	December 16, 2022 – October 6, 2023
Thesis committee:	Dr. ir. W.P. Brinkman, TU Delft, supervisor
	Ir. N. Albers, TU Delft, daily supervisor
	Dr. Ir. M.S. Pera, TU Delft, committee member

Acknowledgements

This work is part of the multidisciplinary research project Perfect Fit, which is supported by several funders organized by the Netherlands Organization for Scientific Research (NWO), program Commit2Data - Big Data & Health (project number 628.011.211). Besides NWO, the funders include the Netherlands Organisation for Health Research and Development (ZonMw), Hartstichting, the Ministry of Health, Welfare, and Sport (VWS), Health Holland, and the Netherlands eScience Center.

I would like to thank my supervisors, Nele Albers and Willem-Paul Brinkman, for their support. You helped me from the first weeks of my thesis when I had no idea what needed to be done, to the final weeks where the last changes had to be made. The whole process of refining my thesis topic, setting up the expert consultation, setting up the experiment, and analysing my findings, I could not have done it without you two. I wish you both the best in your future research.

I would also like to thank Andrei and Ayrton for the fun (and productive of course) times on campus. It really helps to have others struggle with their thesis as well and to have others who you can ask for help with anything you need. Good luck finishing the final part of your thesis. We should definitely grab another beer sometime soon.

Whether or not I stay on the TU Delft campus to work or do research, I do not know yet. What I do know is that I really enjoyed my time studying here. Getting to know new people, gathering knowledge about all kinds of interesting topics, and finishing off with doing actual research, I will never forget those nice experiences.

Martin Dierikx
Pijnacker, September 2023

Abstract

To reduce morbidity and mortality caused by multiple chronic conditions, the number of steps people take each day should be gradually increased. For this, a recommended step goal can be created that is based on an individual's previous walking behaviour. However, for a person, the achievability of this recommended goal can change daily because of that person's state, such as their mood or self-motivation. It could be, for example, that if someone's self-motivation is low, proposing a lower goal than the recommended one, increases their self-motivation and allows them to achieve the recommended goal the next day. Therefore, we investigated the use of a person's state to personalize daily step goal proposals. To do so, we designed and implemented a virtual coach, named Steph, to propose daily step goals to people during an observational study. We used people's states collected in that study to train a reinforcement learning model to optimally personalize the step goal proposals. Based on simulations of our model, we found that people in high states (e.g. who were very motivated and had a positive mood) were more likely to achieve their recommended goals, while people in low states (e.g. who were not motivated and had a negative mood) were less likely to achieve their goals. We also found that proposing higher goals to people in certain states was better than for people in other states. This was because, for some people, a higher goal improved their state while for others, it worsened it. This suggests that personalizing people's step goal proposals optimally could change people's states to where they are more likely to achieve their recommended step goals. So, this thesis provides a model for personalizing daily step goal proposals which can be used as part of behaviour change support systems. It can also serve as a basis for different approaches to predict and change people's walking behaviour to make them more active and less susceptible to chronic diseases.

Contents

1 Introduction **1**
1.1 Research Question . . . 2
1.2 Approach . . . 3

2 Foundation **4**
2.1 Method for the expert consultation. . . . 4
2.2 Goal-setting . . . 5
2.3 Goal personalization . . . 6
2.4 Dialogue with a virtual coach . . . 7
2.5 Overview of factors and concerns for daily personalized goal-setting . . . 8

3 Design **9**
3.1 Messages of the virtual coach . . . 10
3.1.1 Start of the conversation . . . 10
3.1.2 Goal-setting loop . . . 11
3.1.3 Dialogue style. . . . 13
3.2 Initial step goal options creation . . . 13
3.3 Step goal options bounding . . . 15
3.4 Reinforcement learning . . . 15
3.4.1 States . . . 16
3.4.2 Actions . . . 17
3.4.3 Reward function . . . 17
3.4.4 Discount factor . . . 17

4 Evaluation **18**
4.1 Methods . . . 19
4.1.1 Materials . . . 19
4.1.2 Measures . . . 19
4.1.3 Participants . . . 21
4.1.4 Procedure . . . 21
4.1.5 Data preparation and analysis . . . 23
4.2 Results . . . 25
4.3 Discussion of the results . . . 33
4.4 Limitations . . . 37

5 Discussion and Conclusion **38**
5.1 Conclusion . . . 38
5.2 Contributions . . . 39
5.3 Limitations . . . 39
5.4 Future work . . . 40
5.5 Final remarks . . . 40

References **47**

A Expert consultation scenarios **48**
A.1 Psychology experts discussion . . . 48
A.1.1 Scenario 1 . . . 48
A.1.2 Scenario 2 . . . 49
A.1.3 Scenario 3 . . . 49
A.1.4 Scenario 4 . . . 50

A.2 Physical activity expert discussion 51
A.2.1 Scenario 1 51
A.2.2 Scenario 2 51
A.2.3 Scenario 3 52
A.2.4 Scenario 4 53

B Measure details 54

C Mood to valence score mappings 55

D Samples details 56

E Distributions of rewards per state 57

F Policies of the model 59

1 Introduction

In 2022, according to the World Health Organisation (WHO) [87], more than 80% of adults worldwide were not physically active enough. This is a major problem since insufficient physical activity can increase the risk of death by 20-30%. Therefore, the WHO recommends adults to be moderately physically active for at least 150 minutes throughout the week. To support people in increasing their personal activity and changing their behaviour regarding physical activity, more than 17 thousand health and fitness apps are available in the Google Play Store, and the Apple App Store even has more than 23 thousand available apps for this purpose [51]. In 2020, these health and fitness apps had more than 87 million users in the United States [12]. Having around 258 million adults at that time, this meant that in the United States, more than a third of the adults were using a health or fitness app. Next to providing feedback and self-monitoring, the most prominent feature in these apps is goal-setting [51].

According to goal-setting theory, founded by Locke and Latham [47], people's motivation and performance increase when they are working towards a goal. As motivation is one of the factors affecting behaviour change [50], goal-setting can be an effective strategy to change people's behaviour. Although aimed for organizational or work-related tasks, goal-setting has also been shown to be effective for physical activity related interventions and sports [48, 82]. For goals to be effective, it is recommended that they adhere to the SMART principle introduced by Peter Drucker [23]. The five principles of SMART goals are Specific, Measurable, Attainable, Relevant, and Timely. Despite the importance of the SMART principles, most of the health- and fitness apps in the Google Play Store and Apple App Store that use goal-setting do not adhere to all of them. This is because many fitness trackers, for example, just give a step goal of 10,000 steps per day [15], which might not be attainable for everyone.

Goal-setting in health- and fitness apps in the Google Play Store and Apple App Store is typically done following one of two approaches. The first approach is that the app sets the goal for the user depending on some data of the user such as, for example, user experience and ability level. The other approach is that the user can set the goal themselves. Both of these approaches have their advantages and disadvantages. For the first one, when the app sets the goal for the user, the goal will most likely be specific and measurable as it can be based on a lot of data available to the app. However, in this case, the user has little to no say in whether the goal is relevant or attainable for them. But goals are generally more effective when self-set and important to the user [53]. Especially since personal factors (e.g., self-efficacy) can have an impact on attainability [8] but are often not taken into account in this approach. For the other approach, when the goal is self-set by the user, the goal will almost always be personal and relevant which makes it more effective, as the user might be more committed to achieving it because of the feeling of ownership of the goal [37]. However, they might not be specific or measurable enough if the user is inexperienced, as goal-setting is generally difficult [10]. On top of that, the specificity of the goal is crucial since too simple or too complex goals are less effective [61]. Hence, both approaches often use goal-setting which is not done in the SMART way and are, therefore, not exploiting the full behaviour change potential of the goals.

An example of an app that does have SMART goals, is from a study by Resnick et al. [66] where the authors created an app for cancer prevention facilitation. The app would suggest some goals based on the data from the American Cancer Society which the user could then choose from or adjust if needed.

So, although in a simple form with only one interaction, the goals are set collaboratively between the app and its user. In health care, this notion of collaborative goal-setting is becoming increasingly popular [7]. Collaborative goal-setting for health behaviour change can be defined as the process in which the patient and healthcare personnel agree on a health-related goal together [7, 66]. Thus, combining the expertise of the caregiver and the personal factors of the patient to set a goal for the patient. In this process, the caregivers assist the patient in the goal-setting, providing the necessary information to the patient but having the patient make the final decision.

However, caregivers are not always available to collaboratively set a goal. Luckily, a virtual coach can create an interactive relationship with its users without the need for actual humans to support them [85] and is, therefore, always available. As virtual coaches can support goal-setting [90], a virtual coach could also be used to do collaborative goal-setting. The virtual coach can be an effective tool to increase physical activity [83] and one of its benefits over just a health- or fitness app is the social interaction with the coach which increases trust and commitment to the set goal [5].

To make a virtual coach comprehend people's personal factors, an approach is needed to personalize the goals based on those factors. This is important since the lack of personalization can lead to people discarding their goals and not persisting with the behavioural change [41]. For personalization, differently tailored interventions are needed [29, 70]. This is because people can have different levels of fitness, for example. In those cases, it would be less effective to give everyone the same goal because for some that would be too easy and for others too difficult. Therefore, having adaptive goals can make the intervention more effective [8, 41]. This adaptation can be based on several factors that influence the physical activity of a person. Numerous studies point to different types of factors to increase physical activity such as psychological (e.g. self-motivation [62]) and behavioural factors (e.g. previous activity [60]), which can together be seen as the state of a person. As these factors, and with that the state of people, can change on a daily basis, adaptive goal-setting based on these factors should also be done daily. Additionally, the goal for one day could influence a person's state of the next day, for instance, an unattained goal could lower one's self-motivation. Because of this, adaptive goal-setting should also consider the future state of an individual. Therefore, using Reinforcement Learning might be effective in creating adapted goals. A study by Zhou et al. [92] used the Behavioral Analytics Algorithm to predict daily step goals one week in advance using the previous activity and self-efficacy of people. However, because they predicted for a whole week in advance, they did not take personal factors that can change on a daily basis into account, such as self-motivation.

In literature, there is little research done on collaborative goal-setting with virtual coaches, while it is shown that it is important to have a say in the goal-setting [55, 75], because it creates a feeling of personal importance of the goal which enhances the commitment to the goal. Also, little work is done on a reinforcement learning model to adapt goals based on people's current and future personal factors. Therefore, this research aimed to create a reinforcement learning model for personalizing daily goals to be used by a virtual coach to collaboratively set physical activity goals with a person.

1.1. Research Question

Walking, taking steps, is an easy way to reduce morbidity and mortality caused by multiple chronic conditions [59]. This is because taking steps is something that almost everyone already does each day and it does not require any specific equipment. The only problem is that people generally do not take enough steps per day to get to their recommended level of physical activity. Therefore, we focused on creating step goals to increase people's physical activity. With that, the main research question was as follows:

How can reinforcement learning be used to personalize daily step goals for a collaborative dialogue with a virtual coach?

This research question was divided into three sub-questions:

- *What relevant factors and concerns arise when using reinforcement learning to personalize daily step goals for a collaborative dialogue with a virtual coach?*
- *How can a reinforcement learning model and virtual coach be designed for daily collaborative personalized step goal-setting?*

- *How effective is reinforcement learning for personalizing daily step goals for a collaborative dialogue with a virtual coach?*

1.2. Approach

The first step to answering the research questions was to explore the research on goal-setting theory in the context of physical activity. Therefore, we explored what factors could be used to personalize a daily step goal with reinforcement learning and what concerns could arise in a dialogue with a virtual coach. In addition to that, to make design choices for the collaborative dialogue and step goal creation, the advice of experts in the field of psychology and physical activity was obtained. From the literature research and the expert consultations, we obtained factors and concerns for daily collaborative step goal-setting and personalized step goal creation that supported the final design decisions (Chapter 2). After the literature research and expert consultation, we designed the collaborative goal-setting dialogue and the reinforcement learning model for personalized step goal creation (Chapter 3). Subsequently, an observational study was set up and run to gather data on goal-setting with the virtual coach to fit a reinforcement learning model for optimizing the personalized step goal creation. We also ran some simulations to investigate the effectiveness of the model. (Chapter 4). Finally, we discussed the thesis results, identified limitations, proposed future work, and gave some final remarks (Chapter 5).

2
Foundation

This chapter will answer the first sub-question:

What relevant factors and concerns arise when using reinforcement learning to personalize daily step goals for a collaborative dialogue with a virtual coach?

To identify these factors and concerns, we looked for information on the topics of *goal-setting*, *goal personalization*, and *dialogue with a virtual coach* since these form the core of our research. To do so, we used two approaches. We explored previous work and applications to understand the state-of-the-art. Additionally, we consulted experts in the field of health psychology and physical activity for insights into current practices. We report our findings per explored topic, combining the insights from the literature and the experts. Then, at the end of this chapter, a list of identified factors and concerns is presented for the daily personalized goal-setting.

2.1. Method for the expert consultation

We wanted to get further insights into the factors and concerns that arise when personalizing daily step goals for a virtual coach, next to the findings from the literature. Therefore, we consulted experts for a discussion on design choices.

Participants For the expert consultation, we invited three experts:

- A PhD student in the field of biomedical signals and systems, expert in physical activity.
- A senior researcher at the National eHealth Living Lab (NeLL) and assistant professor in the field of health psychology, expert in changing health behaviour.
- A medical psychologist and PhD student in the field of health psychology, expert psychosocial support.

Materials During the expert consultations, we presented the experts with a few scenarios that contained a small use case of the goal-setting dialogue. Each scenario we created, had two or three alternatives displaying different possible designs. An example scenario is shown in Figure 2.1. All the scenarios used in the expert consultations can be found in Appendix A. We used scenarios in the consultation because scenarios are generally effective for discussing different use cases [11], as they are easily made and help people understand the context of the technology.

Procedure In total, we held two discussion sessions. In the first one, we talked with the expert in the field of physical activity to get a better understanding of, for example, the amount of activity that is healthy for an individual. In the second consultation, we spoke with the experts in the field of health psychology to get a better understanding of, for example, how to create the interaction with a person for the virtual coach. For each of the two sessions, we created different scenarios. For each of the scenarios, the experts were asked to indicate their preferences for the alternatives we presented. We also asked them the reasons why they preferred it over the other options. This way, we got insights into why experts would choose certain designs over others.

Figure 2.1: A scenario for one of the expert consultations. In this scenario, the virtual coach proposes a step goal and the person has three options to give their own thoughts on the goal. In option A, the person gets four goal choices to pick from. In option B, they are free to fill in any goal they want. In option C, the person can accept the proposed goal or indicate that they want more or less.

2.2. Goal-setting

In a study, Locke and Latham [47] went through 35 years of empirical research on goal-setting theory to summarize the core findings regarding goal-setting theory. One of the findings they mention in this study is the relationship between the difficulty of a goal and performance. The authors state that performance and effort increase as the difficulty of the goal increases. However, this only works as long as the limit of the ability of a person is not reached and the person is committed to the goal. This goal commitment can be referred to as the determination of a person to achieve a goal [5, 39]. Since goals have no motivational effect when there is no commitment [39], goal commitment is a crucial factor in goal-setting theory. Part of this goal commitment comes from self-efficacy [47], the belief that one can attain the goal. Without this belief, the commitment to the goal drops and so does the motivation to take action. The other part of facilitating goal commitment is the importance of the goal. When people believe that attaining the goal is important for them, they are more determined to take action and achieve the goal, hence increasing the goal commitment.

Besides goal commitment, another important aspect of goal-setting is goal specificity [47]. Goals that only want people to do their best do not work most of the time, because these goals do not refer to anything and can therefore be defined freely. This makes it hard to track progress and does not motivate as much as goals that are defined specifically [61]. Goal *specificity* is the first of five elements of SMART goal-setting, which stands for specific, measurable, attainable, relevant, and timely [23]. *Measurable* is closely related to specific and means that the goal should be measurable, for example, by collecting data [48]. *Attainable* means that the difficulty of the goal is not too high or too low as pointed out by the study of Locke and Latham [47]. *Relevant* points to the relevance of the goal for the person who sets it and *timely* means that there is a deadline specified for the goal [48]. Making goals adhere to this SMART framework increases the effectiveness and makes it more focused on the result, increasing the likeliness of attaining the goal. During the discussion with the psychology experts, they too mentioned that the goals should be SMART and that personalizing a goal could help with this.

In a study by Nelis et al. [55], the authors identify the importance of the SMART principle in their goal-setting intervention for increasing cognitive and physical activity. However, they also mention the co-production of goals, since the participants could work together with the interviewer to set a goal. In the study, the participants were positive about this collaboration as it created a sense of personal importance to the goal. During both expert consultations, the experts all valued the notion of involving people in the goal-setting process and letting them have a say. This is, according to the experts, important to increase motivation and commitment towards the goal. Locke and Latham [47] also mention in their study that participating in goal-setting leads to higher performance. In healthcare, patients generally also participate in setting goals. They do so in collaboration with the caregivers, which is more effective than just assigning a goal to the patients [7]. For this collaborative goal-setting to be effective, the goals should be specific and proximal, which is in line with the SMART principle. From the patients' view, collaborative goal-setting should have a few more elements to be effective according to a study by Morris et al. [52]. In that study, the authors mention that patients want the goal-setting not to be one-sided but that both parties listen and learn from each other. They also say that patients appreciate a measurable goal and want the goal to be feasible. These last two elements again point to the SMART principles, indicating the importance of it for both the theory and practice of goal-setting.

Thus, we identified three factors and concerns for daily personalized goal-setting from the goal-setting topic:

- FC1: People's commitment to the goal that is set.
- FC2: The goals' specificity, measurability, attainability, relevance and timebound (SMART).
- FC3: People's involvement in the goal-setting process.

2.3. Goal personalization

In health behaviour change, a problem that often occurs is the lack of compliance [41], where the person does not stick to the behaviour change. The lack of personalization is one of the reasons for these compliance failures [8]. To tackle this, one way to personalize goals is to adjust the difficulty based on an individual rather than a national recommendation. In a study, Konrad et al. [41] used a person's previous physical exercise performance as a source to base the next goal on. As long as the person achieved the goal, the next exercise goal was more challenging. But when they did not perform well enough, the next goal was immediately easier. This way of adapting the goals created more compliance among the participants in the study. In a study by Cabrita et al. [8], the authors adapted the goals that were set based on people as well. Their adaptation was done by analysing the daily physical activity routine, by looking at factors such as energy expenditure per minute and the difference between the goal and the actual physical activity. This resulted in a goal recommendation for each day, which was averaged with the data of other weeks to set the goal for the next week. In one of the discussions with the experts, the physical activity expert indicated that looking at an individual's previous activity can be an effective way to set daily step goals. By using the previous activity, the new goals will most likely be at the capability of that person, which could increase both motivation and commitment. The physical activity expert also said that step goals should gradually increase in difficulty and not be set too high too quickly as that could increase the chance of injuries and possibly scare people off. Conveniently, by setting goals based on previous activity, this gradual increase is automatically incorporated [2], making this way of adapting very effective.

Next to looking at the previous activity of a person, other factors could also be used to personalize a goal. Several studies point to different factors that influence the physical activity of a person, such as self-motivation and age [62, 79]. Part of these factors can change on a daily basis, meaning that they could influence the amount of physical activity a person does on a certain day. This group of factors can also be split up into two, namely the uninfluenceable ones and the influenceable ones. The uninfluenceable ones are factors that we cannot influence by setting a goal, for example, time, rest, injuries, sickness, and mood [60, 70, 79]. The influenceable factors, on the other hand, are factors that we can change during the goal-setting. These include, for example, self-efficacy, and self-motivation [60, 70, 79]. Self-efficacy is, besides in physical activity literature, also mentioned a lot in the goal-setting for health behaviour change literature [53, 75], indicating the importance of it for goal-setting in the context of physical activity.

However, the physical activity expert did point out that basing a step goal on an individual without any restrictions could lead to goals that are too low or too high. This thought is shared by literature, including studies by Munson & Consolvo [53] and Strecher et al. [75], in which the authors mention that goals that are too easy will not give satisfaction when completed and will not be taken seriously. On the other hand, the authors also recommend that the goals should not be too difficult, since they will not be performed when considered impossible. For health benefits regarding morbidity and mortality due to multiple chronic conditions, adults younger than 60 years old should take at least 4,500-5,000 steps per day [59, 69], so ideally this number would be a minimum for the number of steps. Furthermore, when looking at the increase in health benefits, after roughly 8,000 to 10,000 steps per day the increase flattens [59]. This indicates that setting a step goal of more than 10,000 steps per day does not benefit one's health much more than setting a goal of just 10,000 steps. So in that sense, the 10,000 steps per day could be seen as an upper bound on the recommended step goal. Unfortunately, as was pointed out by the physical activity expert, when a person's previous activity is already above 12,000 steps per day for example, it would not be good to lower the goal to just 10,000 steps. This is because that would be too easy for that person, lowering motivation and commitment. Another example the expert gave, was when someone's previous activity is below 2,000 steps per day. In that case, immediately setting a goal of 4,500 steps per day could be too overwhelming for that person, causing them to lose motivation or even drop out of the intervention. Instead, the step goal should be slightly higher than their current activity to gradually increase people's physical activity, according to the physical activity expert.

One way of dealing with a possible motivation loss is to account for upcoming days. By just proposing a step goal for one day that is challenging because of a mediocre level of self-efficacy, for example, and not considering the next few days, it could be that the challenging step goal is not attained. This decreases the self-efficacy, making the next day have a much lower goal which results in a low total number of steps over those two days. When the next few days are considered, it might be more beneficial to have a slightly lower goal for the first day. This goal then is attained, increasing the self-efficacy, which allows the second day to have a higher goal. Because of that, it could be beneficial to consider future days when setting step goals. Reinforcement learning is an approach that uses current data but also takes the future into account. An example of a study using such an approach to predict step goals is by Zhou et al. [92]. In their study, they use people's previous activity and self-efficacy to predict daily step goals for the upcoming week. Another study regarding reinforcement learning for health behaviour change is by Liao et al. [46]. In that study, the authors use reinforcement learning to predict the most effective moments to send reminders to people to do physical activity. Both studies indicate that considering upcoming days helps with effectively changing peoples' behaviour.

So, we identified three more factors and concerns for daily personalized goal-setting from the personalization topic:

- FC4: People's personal factors that could change on a daily basis such as their previous activity and self-efficacy.
- FC5: The boundaries of the goals.
- FC6: The future days of the intervention.

2.4. Dialogue with a virtual coach

In a study by Wlasak et al. [90], a virtual coach in the form of a chatbot was designed to increase motivation for physical activity. The authors indicated that a chatbot could be an effective way to enhance people's autonomous motivation for physical activity. The design of the chatbot in the study is based on the motivation behaviour change techniques, described by Teixeira et al. [76]. In that study, the authors classified different motivation and behaviour change techniques according to the factors from self-determination theory: autonomy, relatedness and competence. One of the techniques they described is to provide choices, which is also what the psychology experts recommended in the consultation. The experts indicated that people should be able to choose what goal they want to set to make it realistic, but that leaving it completely open could cause people to set goals too high, making them unattainable. According to the psychology experts, not guiding people in choosing a goal could also confuse them, as people expect the virtual coach to know what is best for them and not that they have to come up with a goal themselves. So giving them multiple options to pick from, for example, was said to be effective by the experts. Both the psychology experts and the physical activity expert also indicated that, along with these options, information on what the options are based on is important to provide. By explaining the reasoning or source of the options, the virtual coach becomes more credible and people are more confident that the coach can help them become more physically active. This is in line with a study by Wilkinson et al. [88], in which the authors showed that explaining the why impacts the trust and willingness to depend on the advice of the system.

Other techniques described in the overview of Teixeira et al. [76] are the use of empathic listening and the demonstration of interest in people. In the expert consultation, this came forward when discussing what to do in case a person declines to do any physical activity. The experts said that asking why is the best way to handle this as sometimes it is because of something outside of the person's control, for example, when they are injured. Understanding the person and showing empathy can be an effective way for a virtual coach to respond, according to the experts, creating a bond with the coach which positively influences the motivation and commitment of the person. This is in line with a study by Shum et al. [73], in which the authors said that a bond can be created with a person by having capabilities including empathy and social skills. Lee et al. [45] mentioned in their study that people like to be understood to be able to bond with a chatbot. Another reason why understanding the person is important, as indicated by the experts, is because it enables the coach to react adequately by informing and motivating the person. Reminding a person of their long-term goal, for example, could increase motivation as also found in a study by O'Leary et al. [57]. Finally, simple interactions, such as asking how someone is doing, are appreciated by people improving the bond between people and the virtual coach, according to a study by Albers et al. [4].

Thus, we identified another two factors and concerns for daily personalized goal-setting from the dialogue with a virtual coach topic:

- FC7: People's trust and willingness to depend on the advice of the virtual coach.
- FC8: The bond between a person and the virtual coach.

2.5. Overview of factors and concerns for daily personalized goal-setting

Based on the literature and expert consultation, we identified eight factors and concerns for a virtual coach for collaboratively setting daily step goals that are personalized using reinforcement learning. We found that people's commitment (FC1) and SMART goals (FC2) are important for effective goal-setting. We also found that the involvement of people in the goal-setting process (FC3) increases their commitment to the goal because it feels more important to them. Next to this, we found that three factors and concerns need to be considered when personalizing a goal. The first of these is the consideration of people's personal factors that could change on a daily basis such as previous activity and self-efficacy (FC4), as the number of steps people take on a day could change depending on these factors. The second one is goal boundaries (FC5), as goals need to be of the right difficulty. Third, we found future goals of the intervention (FC6), as setting a goal on one day may impact the personal factors and physical activity of the coming days. Finally, we found that people's trust and willingness to depend on the advice of the virtual coach (FC7) is important for an effective interaction between a person and the virtual coach and that the bond between them (FC8) influences the motivation and commitment of a person.

3

Design

This chapter describes the proposed solution to the research question:

How can a reinforcement learning model and virtual coach be designed for daily collaborative personalized step goal-setting?

Our design idea is to create a goal-setting collaboration where the virtual coach proposes daily step goal options which are personalized using a reinforcement learning model and the person is able to indicate their thoughts on the proposed goals, involving them in the goal-setting process. The proposed design is depicted in Figure 3.1. In our design, first, the coach interacts with the person, using the *Messages* component, to get some data from them (e.g. their number of steps of the previous day). This is done by sending different kinds of messages such as informative messages (e.g. “Let me explain ...”), questions (e.g. “Can you tell me ...”), and confirmative messages (e.g. “Thank you for sharing ...”). Then, three step goal options are created, using the *Initial Step Goal Options Creation* component, based on the person’s previous activity. Next, the *Step Goal Options Bounding* component ensures that the step goal options do not get too high or too low by adjusting them to fall into a certain range. After that, the *Reinforcement Learning* component, which is the core component of our design, uses the bounded step goal options and the data on the person’s personal factors (e.g. self-efficacy) to personalize those options using a reinforcement learning model. This reinforcement learning model optimizes the goal options to make people perceive them in such a way that those people are more likely to do exactly their recommended number of steps each day. Here, people’s recommended number of steps each day is the lowest of the three initial step goal options from the *Initial Step Goal Options Creation* component. Finally, the personalized step goal options are returned to the *Messages* component to be proposed to the person.

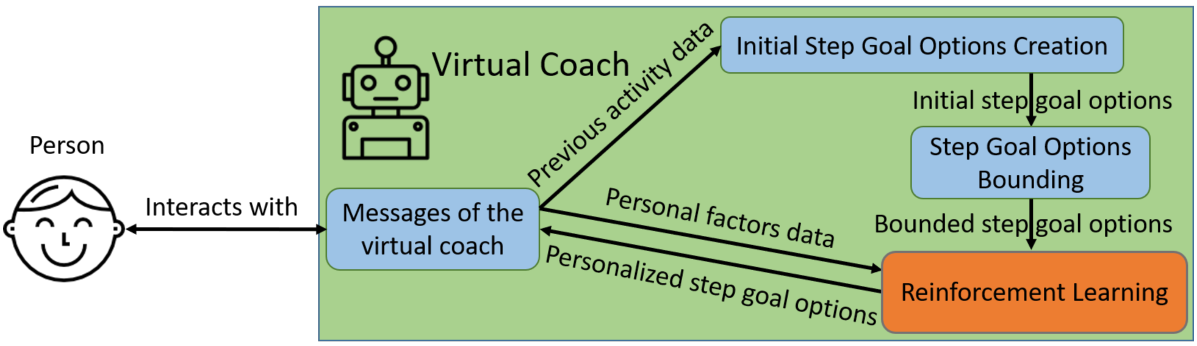


Figure 3.1: The proposed design for a virtual coach that collaboratively sets daily step goals that are personalized using reinforcement learning with a person.

As described in the previous chapter, several studies have shown that setting goals increases the motivation to change one’s behaviour. We also saw that commitment to those goals (FC1) is essential for the actual performance and effort towards a certain behaviour. Part of the commitment to a goal comes from self-efficacy, the belief that one can attain the goal. The other part of creating goal commitment we have seen in the previous chapter is the importance of the goal. So, to design a system that increases people's goal commitment, we wanted to increase and maintain their self-efficacy and make the goal feel important to them. To do this, we used two approaches: let people have a say in the goal-setting process (FC3) and use a reinforcement learning algorithm to personalize goals. Letting people have a say in the goal-setting process creates a feeling of more ownership of the goal [37], increasing the importance of the goal to people [55, 75]. Next to that, using reinforcement learning to personalize the goals allows for the consideration of future days of the intervention (FC6) [92]. This means that the goal of one day can be made lower to increase people's personal factors (FC4), such as self-efficacy, with the idea that focusing on lower goals now could lead to higher personal factors and a higher likelihood that people take their recommended number of steps in the future. Therefore, the reinforcement learning model optimizes how high a goal is for each day to make people do their recommended number of steps each day.

To increase the effectiveness of the goals proposed by the virtual coach, we made them adhere to the SMART framework (FC2) that we explained in the previous chapter. We made the goals *specific* by having the *Initial Step Goal Options Creation* component come up with a number of steps to reach. By using step goals, we also ensured that the goals are *measurable* by monitoring the number of steps people take each day using, for example, a smartwatch or an app on a mobile phone. Next, we tried to make the goals as *attainable* as possible by letting the *Initial Step Goal Options Creation* component tailor the goals towards a person by using, for example, the previous physical activity of the person. This way, the goals would not be far from the normal behaviour of the person, making them attainable. On top of that, we set a limit to how high and low the goals could be in the *Step Goal Options Bounding* component to prevent goals from being too easy or too difficult (FC5). To involve people in the goal-setting process, we had them indicate whether a goal felt attainable so the virtual coach could adjust it if needed. Since people interacting with the virtual coach want to become more physically active and walking is one way to do so, setting daily step goals is *relevant* to them. Finally, the goals were *timely* since we set daily goals. Hence, with our design we adhered to the SMART framework, increasing the effectiveness of the goals.

The final part of our design is that of the virtual coach. To make the behaviour change successful, we also wanted the virtual coach to increase people's motivation and commitment to the goals. Therefore, we wanted to have people feel a bond between them and the virtual coach (FC8). This could be done, for example, by having the coach use a social communication style, making it empathetic and social [73]. These features would also support creating trust in the virtual coach (FC7) [32].

3.1. Messages of the virtual coach

In this section, we describe the dialogue flow of the goal-setting collaboration used to interact with a person. The high-level overview of the dialogue is given in Figure 3.2. The dialogue was designed to collaboratively set daily step goals with a person. The messages from the virtual coach were pre-defined and checked by the psychology experts. Furthermore, most of the responses that people could give also consisted of pre-defined options to simplify the interaction. We decided to give the virtual coach the name Steph because it is short and gender-neutral, avoiding possible gender bias [27]. So from now on, we will use Steph to refer to the virtual coach.

3.1.1. Start of the conversation

As can be seen in Figure 3.2, the start of the conversation consists of a greeting and a check of the previous activity. The exact contents of these steps depend on whether it is the first day of the intervention or not. After that, the remaining steps are the same, no matter the day of the intervention.

If it is the first day of the intervention for the person, meaning it is the first interaction with the person, Steph greets the person, introduces itself, and explains the purpose of the conversation, as can be seen in Figure 3.3. Steph explains why it is important to be physically active and how setting goals could help change the behaviour of the person. After this, Steph asks for the number of daily steps the person did in the past five days. But Steph also gives the person the option to fill in their steps of the past nine days while explaining that that would be even better since it would give a more accurate representation of their normal walking behaviour. So the person can fill in their steps of the past five to nine days. In

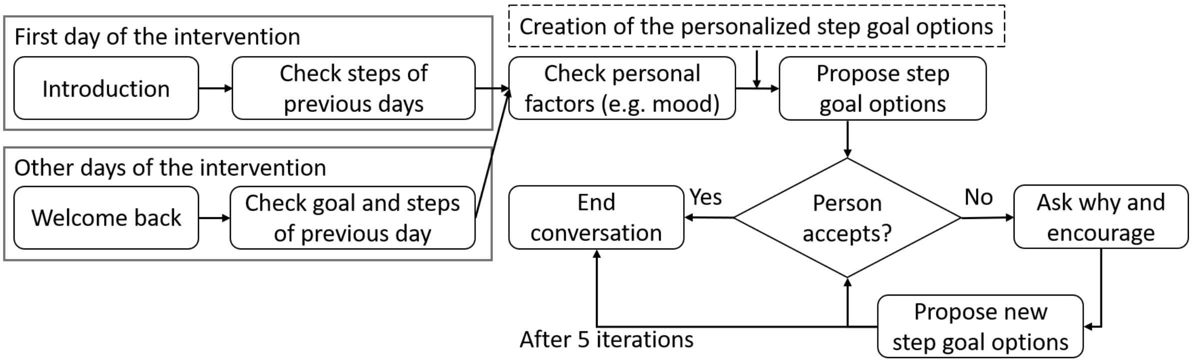


Figure 3.2: A high-level overview of the collaborative goal-setting dialogue flow.

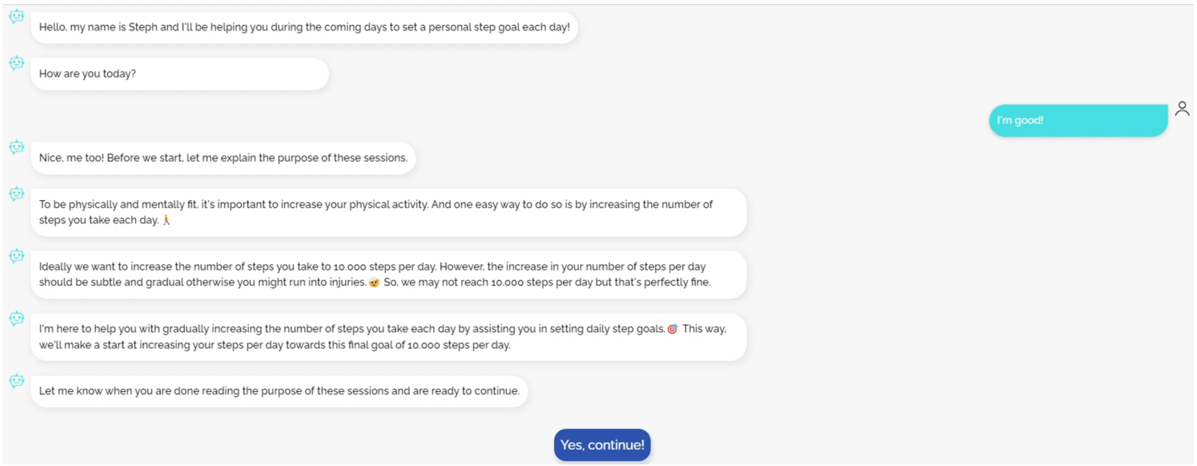


Figure 3.3: Steph introducing itself and explaining the purpose of the sessions on the left and on the right the response of someone using the blue buttons such as the one at the bottom.

case the person gives their steps for less than nine days, the mean of those inputs is added to end up with nine entries of steps per day. This data is needed for the *Initial Step Goal Options Creation* component.

In case it is not the first interaction with someone, Steph simply greets them with a “welcome back” message. After the greeting, Steph asks the person about the number of steps they took on the previous day, as shown in Figure 3.4. Then, Steph either motivates the person if they did not reach the goal of the previous day or gives a compliment to them if they did reach the goal. These empathic and encouraging messages are intended to increase the person’s trust in Steph [32]. The psychology experts also indicated that positive feedback is important to keep someone motivated. After the feedback, Steph asks the person to indicate how achievable they think the goal of the previous day was, looking back at it. Steph also asks whether the person thinks the self-efficacy they indicated on the previous day was accurate looking back at it. These reflections are done to make people think about the goal and their capabilities, which could help them learn to make better decisions in the future [13]. Next to this, it also gives data to be used in the *Reinforcement Learning* component.

After checking the previous activity, Steph asks the person some questions about their personal factors such as self-efficacy. To keep it short and because people tend to be more positive towards filling in multiple-choice questions compared to open questions [49], Steph asks multiple-choice questions to gather the data. Using this approach, the person is asked to indicate their personal factors to be used in the *Reinforcement Learning* component.

3.1.2. Goal-setting loop

When all the questions regarding the person’s personal factors are answered, Steph proposes a set of three personalized step goal options, as can be seen in Figure 3.5. During the expert consultations, the

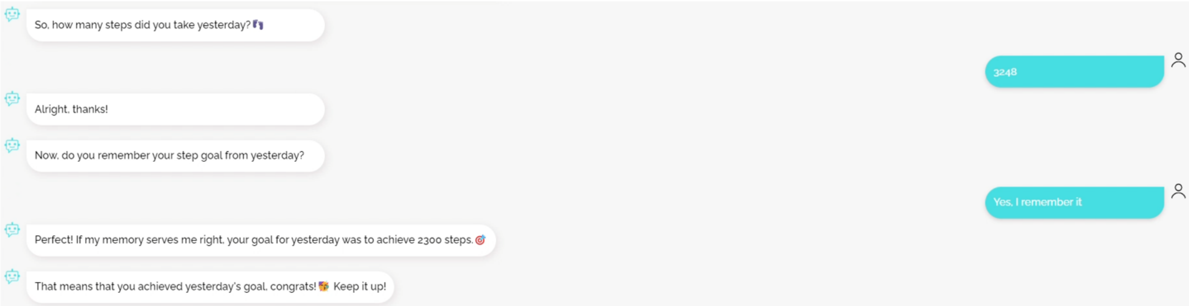


Figure 3.4: Steph asking for the number of steps someone took on the previous day, asking whether they remembered the goal that they set on the previous day, and congratulating the person for achieving that goal. On the right is the response of a person with the number of steps they took by typing the number and the person indicating that they remembered the goal of the previous day by clicking a button.

experts indicated that giving the person the opportunity to choose a step goal themselves increases their commitment, but giving them a completely free choice could overwhelm the person and might make them pick a step goal that is too high or too low for them. So, we chose to give them some options in a range based on their previous activity, to guide the person towards a step goal of the correct difficulty, while still enabling the person to make the final decision themselves. On top of that, Steph explains what the proposed step goal options are based on, increasing Steph's credibility [72]. As explained in the previous chapter, both literature and experts indicate that credibility is essential for building a bond between a person and Steph.

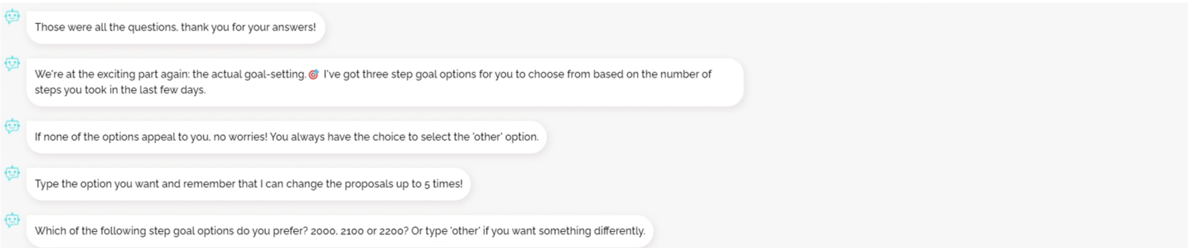


Figure 3.5: Steph thanking the person for answering the questions about their personal factors and explaining how the goal-setting will work. Then, Steph proposes three goal options and the "other" option to adjust the goals. After that, the person can type the option they want to communicate to Steph.

When choosing one of the proposed step goal options, a person is also given the option to choose "other". With that, a person can indicate that they want something different than the proposed set of goals. By giving the option "other", we nudged people towards picking one of the three proposals without restricting their choice [77]. This was done because the proposals are personalized goals and might, therefore, be more effective. So, we wanted people to be more likely to pick one of those instead of adjusting the goals a lot. If the person chooses one of the proposed goals, Steph will congratulate them on the goal and end the conversation with some motivational words. In case the person picks the "other" option, Steph asks whether they think the goal options are too high or too low, as depicted in Figure 3.6. When the person indicates that they are too high, Steph asks them why, which is important for the bonding between the person and Steph since people like to be understood. This is done by giving the person some options to explain themselves (e.g. the weather is not good for walking). Then, Steph explains that, even though the person might not want to, it is important to do at least a certain amount of steps. Steph encourages the person and proposes a new set of step goal options for the person to choose from which are lower than the previous ones. When the person indicates that the proposed step goals are too low, Steph replies that it is happy that the person is up for a challenge but Steph also warns them that the goal should not be too high. Then, Steph proposes a new set of step goal options that are higher than the previous ones to satisfy the person. This loop of asking the person's opinion and changing the step goal options can be done up to four times. This is to prevent the dialogue from getting too long, which could make the person lose interest and drop out, and to make sure that the goal options are not getting too low or too high.

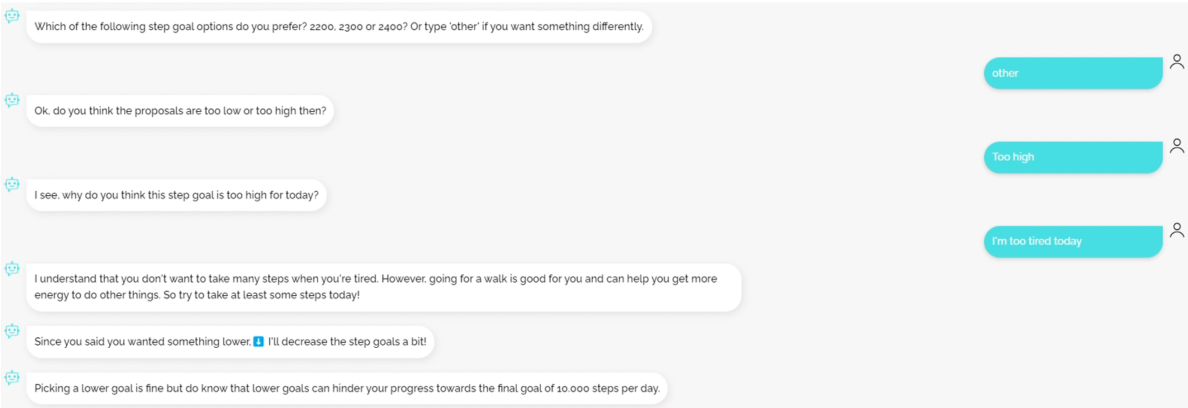


Figure 3.6: Steph proposing three goal options and asking the person why they want something lower. It then explains to the person that, despite being tired, it is important to take at least some steps. On the right are the responses from the person, picking the option that the goals are too high because they feel too tired to walk.

When the maximum number of iterations is reached, Steph informs the person that this is the final proposal and that the following choice of the person will be used as the goal for that day. Steph explains why there is a limit to the number of times it proposes goal options and encourages the person to attain the goal and come back the next day to set a new one.

3.1.3. Dialogue style

In studies by Völkel et al. [81] and De Lannoy [43], the authors said that people prefer an extroverted chatbot over an average or introverted one because chatting with the extroverted chatbot felt more as talking a real person and the extroverted chatbot felt more enthusiastic. Those studies were held in the context of stress tracking and e-commerce respectively, which are two completely different domains. Since, despite the different contexts, both studies found extroverted chatbots to be preferable, we chose to have Steph be extroverted as well. This meant that we let Steph use emojis, which was also recommended by the psychology experts, and have it react enthusiastically and delightfully to the person's input. On top of that, we made Steph use more informal phrases such as "Aww, that's annoying" and "Thanks for answering". Also, Steph uses relatively many exclamation marks and positively valenced words such as "great" and "cool". This informal language, social praise and well-wishing are all part of a social communication style, which often creates more satisfaction for people [91], which can again support in enhancing trust in Steph.

Next to the extroverted personality, we made sure to follow the principles of motivational interviewing, as this could help increase the motivation for the behaviour change [78]. To do so, Steph expresses empathy and uses reflective listening (e.g. by asking the person why they want to set a lower goal and acknowledging answers given by the person). Steph also tries to increase the self-efficacy of the person (e.g. by explaining that the next goal is attainable for someone as it is based on their previous activity).

3.2. Initial step goal options creation

For the creation of the initial step goal options, a starting point was needed. To get that, we used the approach from the study by Adams et al. [2] in which the authors created daily adapted step goals for their participants. They did so for 180 consecutive days and found that participants took more steps when given the adapted goals than when they were given a default goal of 10,000 steps. For the algorithm used in that study, the authors took the number of steps of a person from the previous day and a list of the number of steps of the nine days before that, as shown in Algorithm 1. Then, the oldest step count in the list is replaced by the step count of the previous day to update the list. After that, the list is sorted in ascending order and the 60th percentile gives the step goal for the day.

We adapted the algorithm of Adams et al. by adding the rounding of the goal and adding the creation of two more step goals, as shown in Algorithm 2. To use the goal created by looking at the previous steps of a person, we decided to round it up to the nearest 100 steps (e.g. 1,200 steps instead of 1,176 steps). We did so because people generally need more time to process non-rounded numbers

Algorithm 1 Step goal creation algorithm from Adams et al. [2]

input: List containing the number of steps from the nine days before yesterday of the person S, Number of steps taken by the person on the previous day c.
output: Step goal for the person for the day.

```
if c is not Null then
    Remove least recent entry from S
    Add c to S
end if
Sort S in ascending order
return 60th percentile of S
```

[36], which might mean that getting a non-rounded goal decreases motivation as it takes more effort to process. Also, in a study by Pope and Simonsohn [63] the authors found that people tend to put in extra effort to complete a rounded goal if they are close to it. On top of that, in another study by Gunasti and Ozcan [31] the authors said that reaching a rounded goal gives more satisfaction as people tend to prefer rounded numbers. So by using a rounded step goal, people might push themselves a bit further if they are close to achieving it, increasing their performance, and people might feel more satisfied by completing a rounded step goal. We used the rounded version of the created goal as the initial step goal.

Since we wanted to give people a choice in which step goal they wanted, we converted the initial step goal into three initial step goal options. During the expert consultations, the psychology experts pointed out that there is no need to use a goal below the initial step goal in the initial options, as it does not contribute that much to the behaviour change of becoming more physically active. Next to that, the literature indicated that higher goals lead to higher performance as long as the goals are achievable [47]. Since the initial step goal is based on the previous activity of the person, they are most likely able to achieve this goal. Also, a goal option that is below the initial goal would have a lower impact on the behaviour change of the person, which was not desired. Therefore, we used the initial step goal as the lowest step goal in the set of initial step goal options. To not overwhelm the person with options that might not mean much to them, we decided to only propose a total of three options at once. This was also recommended by the psychology experts as it would otherwise become too difficult to make a choice according to them. Each of the initial options is incremented with 100 steps compared to the previous one. We used increases of 100 steps because 100 steps is not too little where a person would not feel like they have a choice, as pointed out by the psychology experts. But 100 steps is also not too much where the options would deviate a lot from the initial step goal. So, if the initial step goal was 2,000, for example, our algorithm would create the options 2,000, 2,100 and 2,200 steps. The final algorithm for the creation of the initial step goal options is given in Algorithm 2.

Algorithm 2 Initial step goal options creation algorithm, adapted from Adams et al. [2]

input: List containing the number of steps from the nine days before yesterday of the person S (possibly containing multiple entries with the average of the past days as explained in Section 3.1.1), Number of steps taken by the person on the previous day s.
output: Initial step goal options for the person for the day $o1$, $o2$ and $o3$.

```
if s is not Null then
    Remove least recent entry from S
    Add s to S
end if
Sort S in ascending order
Step goal g ← 60th percentile of S
g ← round up g to nearest 100
o1 ← g
o2 ← g + 100
o3 ← g + 200
return o1, o2 and o3
```

3.3. Step goal options bounding

To make the initial step goal options not too low or too high in general, we bounded the range of the step goal options. We did not want the step goal options to get too low as they would then lose their impact on the behaviour change and health benefits of the person. We also did not want the options to get too high as the person would then run the risk of getting injured because of taking too many steps [74]. Since numerous studies [25, 34, 86] showed that only less than 1% of their participants, who represented the general population, took 2,000 steps or less, we assumed that a goal of 2,000 steps would be the minimum attainable goal in general with a very low risk of pushing a person too far in terms of physical activity. Therefore, for the general lower bound, we used 2,000 steps. As for the general upper bound, we used 10,000 steps. This was because a goal of 10,000 steps is generally seen as the number to aim for in daily physical activity [84]. On top of that, during the expert consultations, the physical activity expert pointed out that doing more than 10,000 steps per day does not have any additional health benefits. This is in line with studies by Paluch et al. [59] and Lee et al. [44], which plotted the hazard ratio for all-cause mortality compared to the average number of steps per day. The curves in these plots flatten around 10,000 steps per day, indicating that doing more than that does not significantly lower the risk of all-cause mortality. Hence, we used 10,000 steps as the general upper bound of the created step goal options.

So, the bounding of the initial step goal options involves adjusting the options to fall within the range of 2,000 to 10,000 steps. The pseudocode of the algorithm for the bounding of the step goal options is given in Algorithm 3. As shown, when the lowest initial step goal option is below the lower bound, the lowest option is set to the lower bound and the other options are set to increments of 100 and 200. This ensured that the step goal options were not too low. At the same time, to make sure the options were not too high, we capped them at 10,000. To do so, we set the highest option to 10,000 when the highest initial step goal is above that and set the other options to be decrements of 100 and 200. Finally, when the initial step goal options already fall within the bounds, we do not change them. We will refer to the bounded initial step goal as the “recommended step goal” from now on.

Algorithm 3 Step goal options bounding

input: Initial step goal options $o1$, $o2$ and $o3$ from Algorithm 2.
output: Three bounded step goal options $g1$, $g2$ and $g3$.

```
Lower bound lb ← 2,000
Upper bound ub ← 10,000
if o1 < lb then                          ▷ Use the lower bound as the lowest option
    g1 ← lb
    g2 ← lb + 100
    g3 ← lb + 200
else if o3 > ub then                     ▷ Use the upper bound as the highest option
    g1 ← ub − 200
    g2 ← ub − 100
    g3 ← ub
else                                     ▷ Use the initial step goal options
    g1 ← o1
    g2 ← o2
    g3 ← o3
end if
return g1, g2 and g3
```

3.4. Reinforcement learning

The final part of creating step goal proposals is the personalization of the bounded step goal options, which involves reinforcement learning. By basing the initial step goals options on the previous activity of the person, they are adapted in a simple way for the general state of the person. This means that, on an average day, the person is able to attain those goals. However, as seen in other studies [60, 70, 79], there are several personal factors, such as someone’s mood, that influence the daily physical activity of a person. So, it could be that the recommended goal is not attainable for the person on that specific day because they might not be feeling well, for example. It might then be better to propose lower goal options to make the person not lose motivation or confidence to allow them to achieve their recommended goals in the future. On the other hand, when people are really motivated, proposing higher goals might be the only way to keep them motivated and make them achieve their recommended goal. Therefore, we opted to consider the person's daily personal factors to further personalize the step goal options. For that, we used reinforcement learning to learn when someone would be more likely to achieve their recommended goal given lower or higher step goal options. With this approach, we could also consider the future personal factors of the person in the decision of which step goal options to propose.

We defined our approach as a Markov Decision Process (MDP) $\langle S, A, R, T, \gamma \rangle$ where:

- S is the state space,
- A is the set of possible actions in each state,
- R is the reward function for each action taken,
- T is the transition function between each pair of states,
- and γ is the discount factor.

In an MDP, the agent aims to learn an optimal policy π^* (a set of optimal actions to take in each state), which maximizes the expected cumulative sum of discounted rewards over time $E[\sum_t^\infty \gamma^t r_t]$, where t represents time steps, γ^t is the discount factor in a timestep, and r_t is the reward in a timestep. Each time step corresponds to taking an action in a state and transitioning to a next state while getting a reward. To compute the optimal policy, we used value iteration [33]. The algorithm for that is given in Algorithm 4, where Q_t represents the Q-table at timestep t. δ captures the largest change in any of the Q-table values (Q-values) in each timestep. ϵ is a threshold which we compare with δ to determine when to stop the algorithm, which we set to 0.01.

Algorithm 4 Value iteration, adapted from Hamadouche et al. [33]

Initialize Q_0 arbitrarily, e.g. $Q_0(s,a) = 0, \forall s \in S, a \in A$
$n \leftarrow 0$
repeat
  $\delta \leftarrow 0$
  $n \leftarrow n + 1$
  for all $s \in S, a \in A$ **do**
    $Q_n(s,a) = R(s,a) + \gamma \sum_{s' \in S}(s'|s,a) \ \max_{a' \in A} \ Q_{n-1}(s',a')$
    $\delta \leftarrow \max(\delta, |Q_{n-1}(s,a) - Q_n(s,a)|)$
  end for
until $\delta < \epsilon$
Return Q_n

3.4.1. States

The state space S consisted of a subset of the person's personal factors. These personal factors are factors that influence the physical activity of a person and can change on a daily basis. We considered the following factors: self-motivation, self-efficacy, rest, mood (valence), and available time. First, self-motivation is found to be a significant factor in exercise adherence [62] and is, next to that, one of the main influences of goal commitment [47]. Second, self-efficacy is also found to facilitate goal commitment and influence people's physical activity behaviour [79], and has, on top of that, been used in reinforcement learning for goal-setting before [92]. Third, being too tired is one of the most common barriers to physical activity [38, 60] and higher sleep quality is generally associated with higher levels of physical activity [6]. Next to this, feeling good, and with that implying a high valence mood, is found to increase physical activity [65]. Finally, having no time is a highly occurring reason to not participate in physical activity [38, 60, 70] and available time gives an opportunity for behaviour change [50]. We considered rest, available time, and mood as external factors, meaning that we could not influence those features much with our goal-setting dialogue. Instead, other parts of a person's life may change these features from day to day. Therefore, we combined these three features into one state feature, namely the *context state*. We also did not aim to improve this state feature during the goal-setting, as we could not really influence it. The state variables were collected by asking questions to the person during the conversation.

3.4.2. Actions

The action space A consisted of the different ways that the bounded step goal options could be modified. We selected the actions: increase, slightly increase, do not change, slightly decrease, and decrease. The *slightly increase* action increases all of the three step goal options by 200 steps. The *increase* action increments the options with 400 and for the *decrease* and *slightly decrease* actions it is a 400 and 200 steps decrease respectively. Finally, the *do not change* action does not change the options and proposes them as they are. Both increase and decrease actions do not change the step goal options outside of the bounds specified in the previous section. Therefore, if an upper or lower bound is reached, the respective increase and decrease actions do nothing anymore. In every conversation session, one of the actions was picked to create the personalized step goal options.

3.4.3. Reward function

The reward function R was determined by the absolute difference between the recommended step goal and the actual activity of the person towards that goal (Δ). This reward signal was chosen as it gives an indication of how close the person was to doing the recommended amount of steps for that day despite the set goal possibly being slightly different. Therefore, it can evaluate whether the chosen action to modify the initial step goal proposals was the correct choice. So, we have the following reward function: $R : r = 1 - (\Delta/recommended\_step\_goal)$. If the number of steps taken was higher than the recommended goal, the Δ would be halved to make taking more steps than the goal less penalizing compared to not reaching the goal. By dividing the absolute difference by the recommended step goal, we penalized a difference of 100 steps, for example, harder for lower goals than for higher ones. This is because being 100 steps off of a goal of 10,000 is relatively close in comparison to being 100 steps off of a goal of 2,000 steps.

3.4.4. Discount factor

The final part of the MDP, the discount factor γ, indicated the consideration of future rewards, where a higher γ means that there is more value attached to rewards in the distant future than in the near future. We set our discount factor to 0.85, as done by Albers et al. [3] for example, to focus on choosing actions that lead to higher rewards in the near future as we wanted to have effective goals early on. This was because ineffective goals at the start of the intervention could lead to people dropping out early on. But, we also still tried to be able to get a reasonable reward in the distant future, by accounting for that as well to allow for long-term behaviour change.

4 Evaluation

This chapter answers the following research question:

How effective is reinforcement learning for personalizing daily step goals for a collaborative dialogue with a virtual coach?

In the previous chapters, we showed the design of a reinforcement learning model to create daily personalized step goal proposals to be used in a dialogue with a virtual coach. To train and evaluate the model, an observational study was set up. The goal of this study was to gather data on people's states and behaviour, after proposing step goals that were randomly adjusted using one of the actions described in the previous chapter. This chapter describes the study setup, followed by the data analysis and results. Then, the results are discussed and the chapter ends with an explanation of the limitations of the observational study.

To answer the research question, we divided it into five analysis questions (AQ). We answered these questions by simulating the trained reinforcement learning model based on the data collected during the observational study. With the first question, we wanted to investigate the effect of the states described in the previous chapter on people's behaviour. People with higher self-motivation, for example, are more likely to attain physical activity goals [1] which could mean that they do not need a higher goal to push them to achieve their recommended goal for that day. Therefore, we wanted to see if looking at the states of people could predict how much people would over- or underachieve compared to their recommended step goal of that day when doing one of the actions described in the previous chapter. To see whether the state features help predict the behaviour of people, we formulated the following analysis question:

AQ1: How well do the states predict behaviour after proposing personalized step goals?

For the second analysis question, we looked at the future states of people. We expected that in certain states people would be more likely to attain the recommended step goal of that day. With that, it would be important to get people to such a state where they are more likely to do the recommended number of steps each day. For people in a low state (low self-motivation, low self-efficacy, and low context state), for example, setting a challenging goal could set them up for failure, decreasing their self-motivation and self-efficacy [1]. At the same time, for people in a high state (high self-motivation, high self-efficacy, and high context state), setting a challenging goal could be the only way to keep them motivated, as simple goals could lower their motivation [47]. So, since actions in different states could change people's state differently, we wanted to know whether we could predict people's future state based on their current state and we formulated the following analysis question:

AQ2: How well do the states predict the next states after proposing personalized step goals?

Then, for our third question, we started to look at the optimal policy for personalizing step goal proposals. The optimal policy describes which action is best to take in each state. Ideally, this optimal policy picks the action that modifies the proposed step goal to have people perceive it in such a way that they take their recommended number of steps trying to achieve the goal and get more confident and motivated from it. With that, it should ideally change their state to where they are more likely to reach their recommended step goal the next day, therefore, taking into account people's future states. Next to this, we were also interested in the long-term effect of the optimal policy on people's states, as it would only be effective if the states also remained in a way where people are more likely to reach their recommended step goal the next day over a longer period. So, to investigate the effect of the optimal policy on people's state, we created the following analysis question:

AQ3: What is the effect of (multiple) optimal step goal proposals on people's states?

Next to the effect on people's states, we also looked at the effect of the optimal policy on the behaviour of people. We were interested in whether the optimal policy had a better impact on the behaviour of people compared to non-optimal policies such as the worst policy or a policy where actions are taken randomly. With this, we could see if there was a best approach for each state, or whether taking a random action would also work. So, we wanted to see if proposing non-optimal step goals would have a different effect on how close people would get to their recommended goal, than proposing optimal goals. Therefore, we created the following analysis question:

AQ4: How do optimal and sub-optimal step goal proposals compare in their effect on behaviour?

Finally, we investigated the reward function of our model. Different reward signals could lead to different optimal policies. For example, if we were more interested in making people feel positive about the goal by looking at their perceived goal achievability, we wanted to know whether the optimal policy would change. Or, for example, if we wanted to push people more in general by giving higher rewards for proposing higher goals, we were curious whether a different policy was needed. So, next to the reward signal described in the previous chapter, we also wanted to consider the goal achievability and a standard reward for each action as possible reward signals. So, we investigated the effect of these different reward signals, and different reward functions, on the optimal policy. For this, we formulated the following analysis question:

AQ5: How do different optimal policies based on different reward functions compare?

4.1. Methods

The observational study to gather data for the training and evaluation of the reinforcement learning model was run in June and July 2023. Before running the study, we registered the design of the study on the Open Science Framework (OSF) [20] and ran a pilot study with 34 people to test if everything was ready for the full study. The pilot resulted in no major changes other than the addition of one question in the post-questionnaire to measure the perceived goal difficulty as the goal difficulty could affect the rewards. Furthermore, the study design was approved by the TU Delft Human Research Ethics Committee (HREC reference number: 3016).

4.1.1. Materials

In this study, we used multiple online services. We used Prolific[1] for recruiting, inviting, and communicating with participants, Qualtrics[2] for hosting the questionnaires and instructions for the conversational sessions, and Google Compute Engine[3] to host the virtual coach and the conversation sessions. Furthermore, we implemented the virtual coach using Rasa[4] version 3.2.8. The source code for the virtual coach can be found online [19].

4.1.2. Measures

[1]`https://www.prolific.co/`
[2]`https://www.qualtrics.com/`
[3]`https://cloud.google.com/`
[4]`https://rasa.com/`

Primary measures
For the primary measures, we considered the measures that were relevant to answering the analysis questions. These were the following measures:

- *State measures*, we measured six variables to determine the state of a person during the conversation sessions. These included self-efficacy, self-motivation, rest, and available time, which were measured on a scale from 0 to 10. Here, the self-efficacy measure was based on the definition from Park and Kim [60], the self-motivation and available time measures were based on the physical activity barrier descriptions from Robbins et al. [67], and the rest measure was based on the sleep quality rating from Dzierzewski et al. [26]. Furthermore, the valence dimension of mood was measured using adjectives from a study by Russell [68]. The labels and exact questions used for the primary measures and secondary measures mentioned in this section apart from the human interaction of the virtual coach can be found in Appendix B. All the state measures were subjective to simplify the collection procedure.

- *Steps taken*, we measured the number of steps people took each day by asking people to provide the number of steps they took on the previous day by looking it up on, for example, their smartwatch or smartphone.

- *Perceived goal achievability*, we measured how achievable the first step goal proposal of the previous day felt to people, which was asked on the next day. This was done using a question based on the perceived goal attainability question from Huang et al. [35], with a scale from 0 to 10.

Secondary measures
For the secondary measures, we considered the measures that we used for the exploratory analysis. These included the following measures:

- *Human interaction of the virtual coach*, to measure attributes such as the usability and acceptability of Steph, we used the short version of the Artificial Social Agent Questionnaire by Fitrianie et al. [28]. This gave 24 statements to be answered using a 7-point Likert scale from disagree to agree.

- *Perceived goal difficulty*, we measured how difficult the goals felt to achieve to people using a scale from -5 to 5. Again, we also asked why they felt so using an open question.

Other measures from the conversation sessions that were collected but not used in the analysis (for example, the number of rejected proposals, the reason for rejection, and weekly leisure activities) can be found in the OSF pre-registration [20].

Demographic measures
For the demographic measures, we considered measures that gave insights into our participants. These included the following:

- *Prolific measures*, we collected three measures from the Prolific profiles of our participants: age, gender, and engagement in physical exercise per week in minutes.

- *TTM-stage for becoming more physically active*, we asked people about their TTM (Transtheoretical Model [64]) stage of change regarding becoming more physically active using an adapted version of the question from the Cancer Prevention Research Center [56] where we replaced the question with “If regular physical activity was defined as taking 10,000 steps per day, would you say you are doing regular physical activity according to this definition?”.

- *Average steps per day*, we also asked people how many steps they took on average per day in the last week using the question “How many steps did you take per day on average in the last week?” and letting people respond with a number between 0 and 20,000. People take on average no more than 18,000 steps per day [80], so with an upper bound of 20,000, we allowed most people to enter their average steps per day while preventing typos where 4,000 becomes 40,000, for example.

- *Way of tracking steps*, we asked people how they were tracking the number of steps they took each day using the question "How do you track the number of steps you take each day?" with answer options: "Using a Smartwatch", "Using an iPhone's Health app", "Using the Samsung Health app", "I do not track the number of steps I take each day", and "Other, namely:" where people could give a free text response.

4.1.3. Participants

To determine our target sample size, we used the guidelines by Cohen [16] for multiple regression analysis and a medium effect, an alpha of 0.05, and three independent variables (i.e., the three state features that describe the state space). This resulted in a sample size of 76 samples. Since we had five possible actions in our model, we multiplied this sample size by five for a value of 380 samples. Each sample consisted of data from two conversation sessions with Steph, such as people's state and number of steps taken. Since every participant was invited to at most five conversation sessions, we could get four samples from one person. This meant that we needed at least 95 participants.

For our study, we used Prolific's screeners to only invite people who indicated that they:

- were fluent in English,
- were between 18 and 65 years old,
- engaged in physical activity for at most 150 minutes per week,
- lived in one of the following time zones: GMT, GMT+1, GMT+2, GMT+3, or GMT+4,
- had completed at least one previous study on Prolific,
- and had an approval rate of at least 90% on Prolific for their previously completed studies.

The timezone criteria made it so that the session duration from 4 AM until 12 AM GMT+2 would be in the morning for all participants. This was needed as setting a daily step goal at the end of the day did not make sense because people could not work towards that goal anymore by then. Next to the Prolific screening, people had to give consent, which meant that they indicated to:

- have a way to track their steps in the previous and coming five days,
- not be participating in a physical activity program at that moment,
- and have a low risk of getting injured because of walking according to the Physical Activity Readiness Questionnaire 2023 [18].

Finally, to be eligible to participate in the study, people had to be in the contemplating or preparing stage of the TTM for becoming more physically active and had to be taking no more than 9.000 steps on average per day in the last week. We ended up with 117 people to provide us with at least one sample, whose characteristics can be seen in Table 4.1. These 117 people gave us a total of 381 samples (data from two conversation sessions with Steph on consequetive days), reaching our target sample size.

4.1.4. Procedure

We ran a pilot from June 9 until June 19 where we tested if everything worked and collected 28 samples. Since we only added one question to the post-questionnaire to measure the perceived goal difficulty, we still considered the 28 samples useful and included them in the analysis as part of the 381 samples. The participants from the pilot were excluded from participation in the full study. The participants for our study were recruited from Prolific. They received monetary compensation based on the payment rules on Prolific (i.e., min. 6 GBP/hour). This would mean that a person completing the full study would get a total of £4.30. We had £707 as a budget for the study, of which we spent £53.60 on the pilot which meant we had £653.40 left for the full study. Near the end of the study, we found that we were not going to get to our target sample size because of a higher dropout rate than anticipated. Therefore, we added another £43.25 to our budget to be able to reach our target sample size, which we also adjusted in our OSF registration [20].

The study was split into three parts:

	Total (117)
Age	
Mean Standard deviation Range	28 8 18 - 56
Gender	
Females Males Other	60 (51%) 55 (47%) 2 (2%)
Engagement in exercise per week	
Less than 60 minutes Between 60 and 150 minutes	52 (44%) 65 (56%)
TTM-stage for becoming more physically active	
Contemplating Preparing	50 (43%) 67 (57%)
Average steps per day before the study	
Mean Standard deviation Range	4,402 2,383 30 - 9,000
Way of tracking steps	
iPhone health app Samsung health app Smartwatch Other: - “Huawei health app” - “Google fit” - “Sweatcoin” - “Step app on mobile phone” - “Pokemon Go” - “MapRunner” - “ITO pedometer” - “MStep app”	38 (32%) 36 (31%) 28 (24%) 15 (13%) 6 2 2 1 1 1 1 1

Table 4.1: Participants’ characteristics.

1. *Pre-screening questionnaire*, first, participants were asked to fill in a questionnaire on Qualtrics where they were asked to give consent to participate in the study. Next to that, they were asked some screener validation questions about their fluency in English and their weekly physical exercise and some questions to determine their eligibility about their stage of change for becoming more physically active and the number of steps they took on average per day in the past week. Also, people were asked about their weekly leisure activities by using the questions from Godin [30]. Finally, people had to pass two attention checks. This questionnaire took about six minutes.

2. *Conversation sessions with Steph*, second, people who were picked to continue after the pre-screening were invited to participate in up to five conversation sessions with Steph. The sessions ran from 4:00 until 12:00 (midday) GMT+2. Only one session could be taken per day and the sessions had to be done in consecutive days. In each session, people were asked about the number of steps they took in the previous days, their state (mood (valence), rest, available time, self-motivation, and self-efficacy), and the step goal they wanted to set. Each session randomly picked one of the actions, among the ones that were picked the least, to personalize the step goal proposals to gather data on the behaviour of people resulting from doing those actions. From the second session onwards, people were also asked about their perceived goal-achievability of the goal of the previous day and the accuracy of their indicated self-efficacy of the previous day. The first session took about ten minutes, the second one took about six minutes and the others about five minutes as they required less explanation and people got more familiar with the chat.

3. *Post-questionnaire*, the third and final part of the study consisted of a post-questionnaire in which the participants were asked about the goal-achievability of the goal of the previous day, the accuracy of their self-efficacy of the previous day, and the number of steps they took on the previous day. Next to that, people were asked to rate how personal the step goals felt, how difficult the step goals were to achieve, and which part of the conversation with Steph motivated them the most to go for a walk. On top of that, people were asked to answer questions about the human interaction of Steph. Finally, people had to pass two attention checks. This questionnaire also took about six minutes.

We started the pre-screening questionnaire round on eight different days between June 21 and July 3. For every pre-screening round, we invited as many people as we could pay for the whole study without exceeding our budget limit in case of no further dropout. We opted for a sample of at least 40% males and 40% females. After the pre-screening questionnaire, eligible people were invited to the first conversation session the next day. In these conversation sessions, randomization took place by randomly selecting which one of the actions described in the previous chapter (increase, slightly increase, do not change, slightly decrease, and decrease) to use. We tried to have each action be picked the same amount of times by choosing randomly amongst the actions that have been taken the least in total over all the conversation sessions of all participants. This is because we ideally wanted data on each state-action pair. Finally, the invitations for the post-questionnaire were sent one day after completing the fifth conversation session. Participants were informed that their payment was independent of whether they achieved the set step goal or not to account for self-interest and loss aversion biases [22]. As shown in Figure 4.1, we invited 859 people to the pre-screening questionnaire, also counting the people from the pilot study. Of those, only 75 completed the whole study.

4.1.5. Data preparation and analysis

The gathered data was cleaned in Python, where we removed the data from excluded people and anonymized the data for analysis. We also restructured the data to group data of two conversation sessions into one data sample. Finally, we transformed the collected mood data of people into numbers to extract the valence values and have the data in the same format as the rest and available time measures to be able to combine them into the context state feature. So, we transformed the moods into a valence score between 0 and 10. To do so, we mapped the moods onto the horizontal axis (valence axis) of the 2-D Emotion Wheel of the study by Kollias et al. [40]. We then mapped the range [-1.0, -0.9] on the axis to a score of 0, the range [-0.9, -0.7] was mapped to 1, etc., and the range [0.9, 1] on the axis was mapped to a score of 10. This way, a mood with high valence got a high valence score as it indicated a positive mood. The complete mapping of the moods can be found in Appendix C. After the mapping, we just added the values of rest, available time, and valence score to get the context state, which could take values between 0 and 30. For the data analysis, we used Python. The code and datasets can be found in the 4TU ResearchData Repository [21].

State feature selection To answer the analysis questions, we needed to know which of the state features mentioned in the previous chapter we wanted to use for our model and what possible values each of them would take. Ideally, we wanted each state feature to be binary since the state space would then be as small as possible which would mean that each state-action pair would have more data to use for training. However, using only binary features could mean that certain information about different states would be lost due to shrinking it down into a zero or a one. Hence, we determined which of the state features and what possible values to use by implementing an adaptation of the G Algorithm [14]. In the study by Chapman and Kaelbling, they describe how the G Algorithm identifies significant differences in the split data, according to the different state features, when calculating the Q-values. However, to use it for our study, we adapted this algorithm in two ways. First, to account for non-binary state features, state features with possible values from zero to three for example, we used an ANOVA test rather than a t-test to determine significant differences in the data. Second, instead of only checking differences when calculating Q-values, we also looked at differences when calculating reward predictions and state transitions to see which state features are significant for predicting those parts of the reinforcement learning model. This was done to investigate whether the features could explain different parts of the reinforcement learning model. We considered four possible values to be the maximum per state feature, as otherwise, the state space would become too big. For splitting the

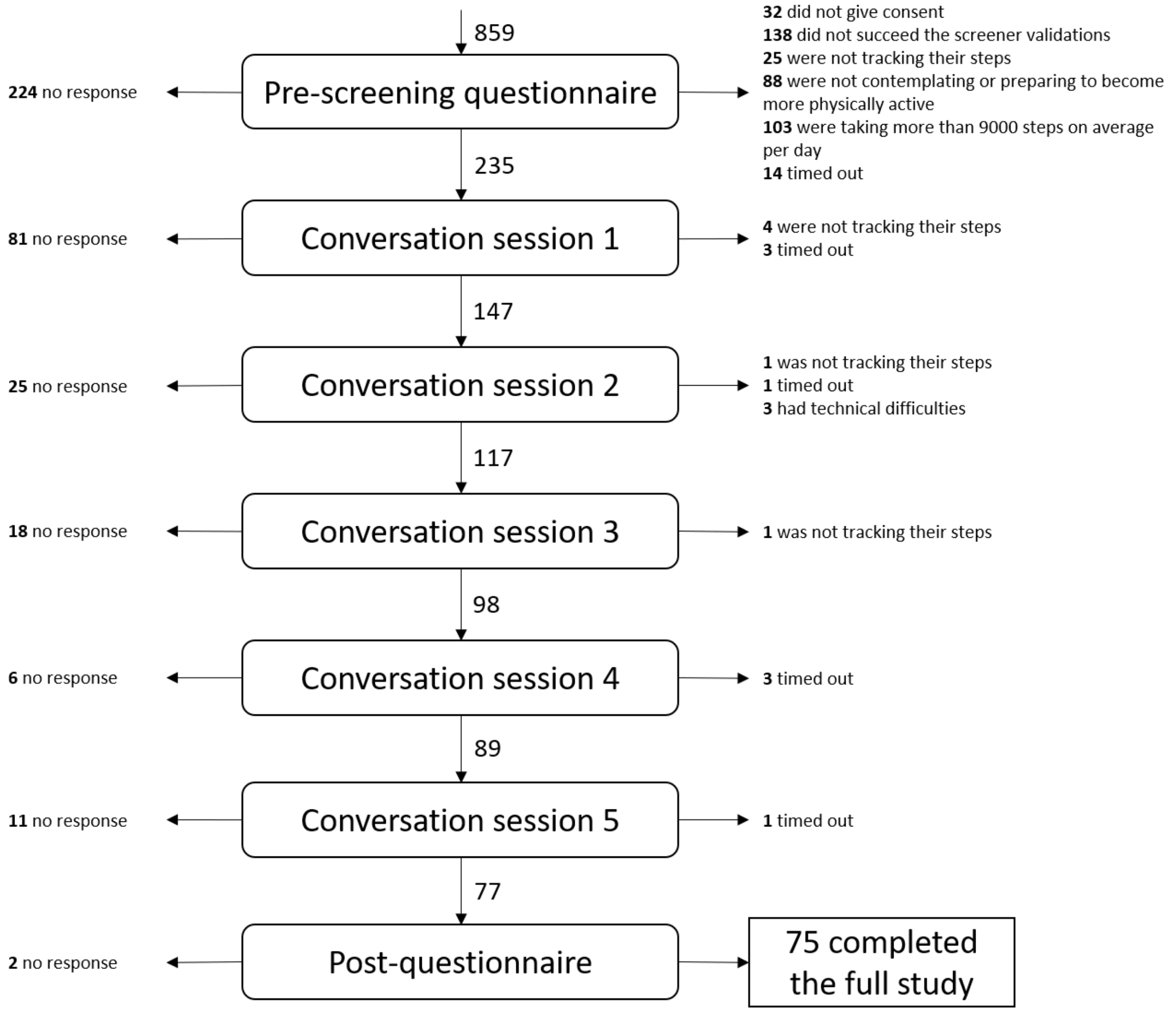


Figure 4.1: The participant flow of the observational study. The numbers next to the downward arrows denote how many people were invited to each of the study components. If people timed out, it meant that they started participating in that part but did not finish it within the time limit for the study set automatically by Prolific.

data, we looked at the percentiles of each data point. So, for a binary feature, for example, we split the data in the 50th percentile mapping the bottom half to 0 and the top half to 1. Also, we used a p-value of 0.05 to indicate significance. Next to the state features and their sizes, we also wanted to see if the three features that make up the context state were significant for the context state. So, if using the rest feature for the context state, for example, would change the value significantly compared to when not using it.

As can be seen in Table 4.2, we ran our adapted G Algorithm three times to check for significant state features and once to check the significance of each of the features of the context state. Running the algorithm for Q-value predictions was the only variation that gave a significant set of state features, where the context state was tertiary. In combination with that, both self-motivation and self-efficacy were significant when binary. However, when running our analysis with this setup, we found that some states had too little data. So, we decided to change our approach slightly from what was reported in our OSF form [20] and decreased our state space further to all binary features. With that, we had eight states ranging from [0,0,0] to [1,1,1]. Also, all three features (rest, available time and valence) were picked for the context state feature of our model, which meant that we could create our context state value by summing the values of rest, available time, and valence and after that look at the percentiles of the data for the mapping to a binary feature.

Run of adapted G Algorithm	Significance of state features with different sizes								
	Self-motivation			*Self-efficacy*			*Context state*		
	2	3	4	2	3	4	2	3	4
Reward predictions	No	No	No	No	Yes*	No	No	No	No
State transition predictions	Yes*	Yes*	Yes*	Yes*	Yes*	Yes*	Yes*	No	No
Q-value predictions	No	No	No	No	No	No	No	Yes**	Yes**
	Significance of context state features								
	Rest			*Available time*			*Valence*		
Context state predictions	Yes**			Yes**			Yes**		

* means the feature with that size was only significant on its own.
** means the feature with that size was significant in combination with the other features.

Table 4.2: Results of running our adapted G Algorithm for finding significant state features and their sizes. Denoted is the level of significance for picking a feature with a certain size to split the data on.

Sample balancing Finally, after creating the states using the aforementioned state features, we found that the distribution of samples across all state-action pairs was not ideal as one of the states still had very little data, as shown in Appendix D. Because of that, we injected some mean samples during the analysis to balance out the potential noise of the lack of samples for some state-action pairs. We had 380 target samples divided over 40 state-action pairs, so rounded up we wanted ten samples per state-action pair. For state-action pairs that had less than ten samples, we adjusted, for example, the mean reward for those pairs by "adding" samples with the overall mean reward. So, if a state-action pair had eight samples, for example, we used 80% of its mean reward and 20% of the overall mean reward to create the new mean reward for that state-action pair. We used this to also balance the transition probabilities and reward predictions to reduce the possible skewness of having only one sample for a state-action pair, for example. For the transition probabilities, we did not balance with the mean probability for each next state but we used an equal probability for each next state as a balancing distribution. We did not account for this before the analysis and, therefore, updated our OSF registration accordingly [20].

4.2. Results

To start off, we analyzed the results of the ASAQ questions in the post-questionnaire by summing the mean for each of the 24 items to get an ASA-Score of 19.32. In the study by Fitrianie et al. [28], the authors calculated the ASA-Score for 13 agents and a dog. The ASA-Score of our virtual coach Steph was higher than that of, for example, virtual healthcare agents Amy and Sim Sensei, who scored 9 and 17 respectively, and other disembodied agents such as Siri, who scored 13, as shown in Figure 4.2. This indicated a relatively good human interaction compared to other health and disembodied agents.

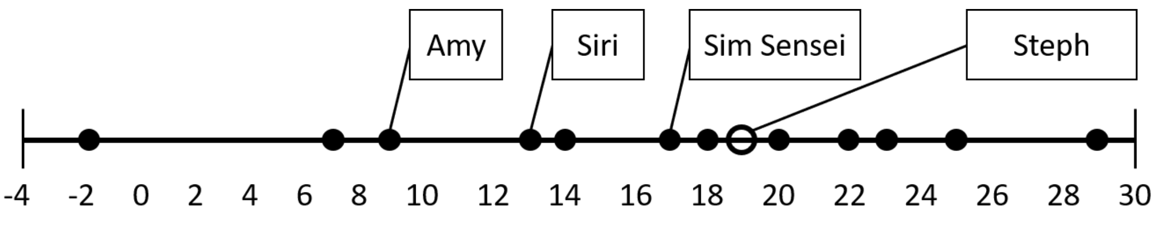


Figure 4.2: The ASA Scores from different agents from the study by Fitrianie et al. [28]. The higher the score, the better the human interaction of the agent.

Across all the 381 samples, the proposed step goals were rejected a total of 100 times. The distribution of the number of rejected proposals per sample is given in Figure 4.3. As shown, most samples had no rejections but there were occasions where people did adjust the goal proposals.

Furthermore, a total of 251 (66%) samples had people achieving their goals. On top of that, people indicated that the goals were relatively easy, as shown in Figure 4.4. Only a handful of people indicated that the goals were more on the difficult side.

The average number of steps taken per day before the start of the study was 4,549 steps (SD =

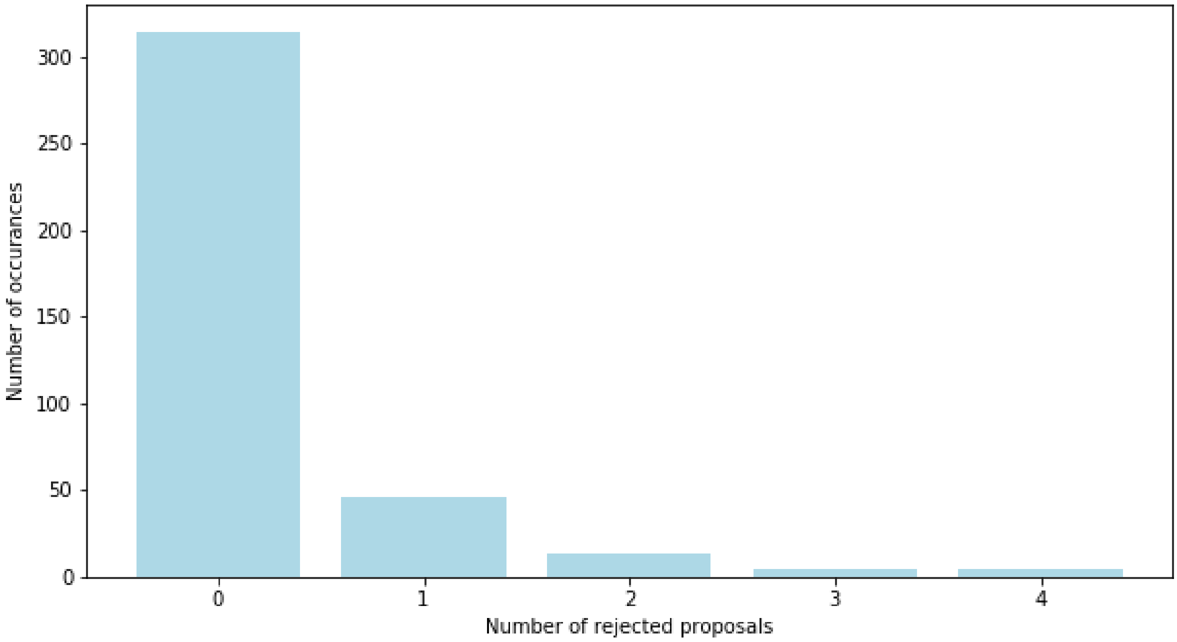


Figure 4.3: The number of rejected step goal proposals from all the samples.

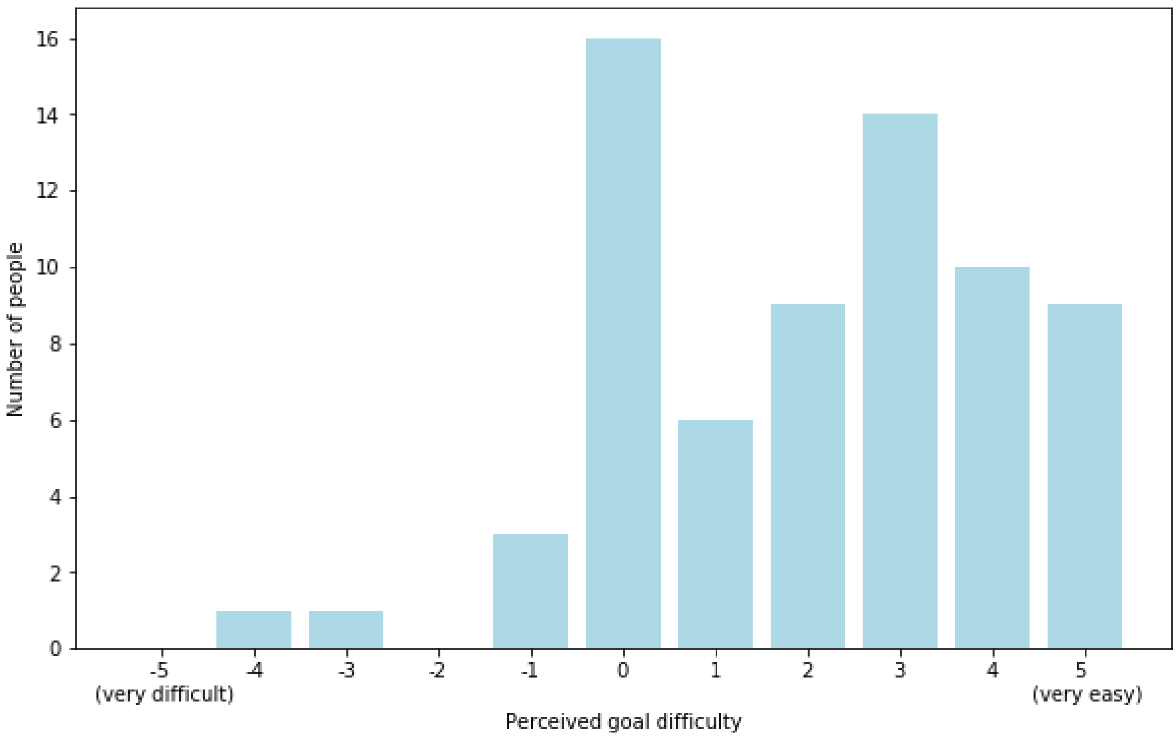


Figure 4.4: The number of people who gave each answer to the post-questionnaire question about their perceived goal difficulty. -5 was labelled as “It was very difficult to reach the daily goals”, 0 was labelled as “It was not difficult but also not easy to reach the daily goals”, and 5 was labelled as “It was very easy to reach the daily goals”.

4,350) by the people who completed the whole study. On the fifth and final day of the study, these people took on average 5,367 steps (SD = 3,353), showing an increase in the number of steps taken per day over the course of the study. Figure 4.5 shows that for the states where two or more state features were high, a relatively large group of people took more steps than the recommended goal.

In the state [1,1,1], for example, 75% of the people did on average 55% more than the recommended goal. On the other hand, for states [0,0,0] and [0,0,1] people were more often underachieving 45% of the recommended goal instead of overachieving.

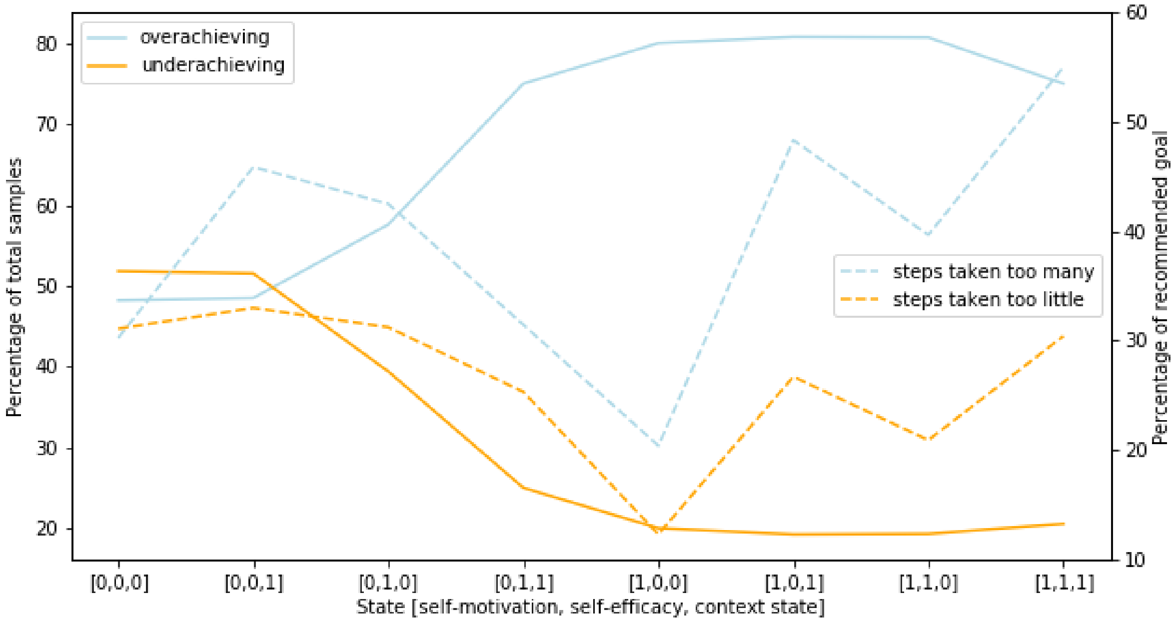


Figure 4.5: The percentage of samples where more or fewer steps were taken than the recommended goal per state. On the right axis is the percentage of the recommended goal that was overachieved or underachieved.

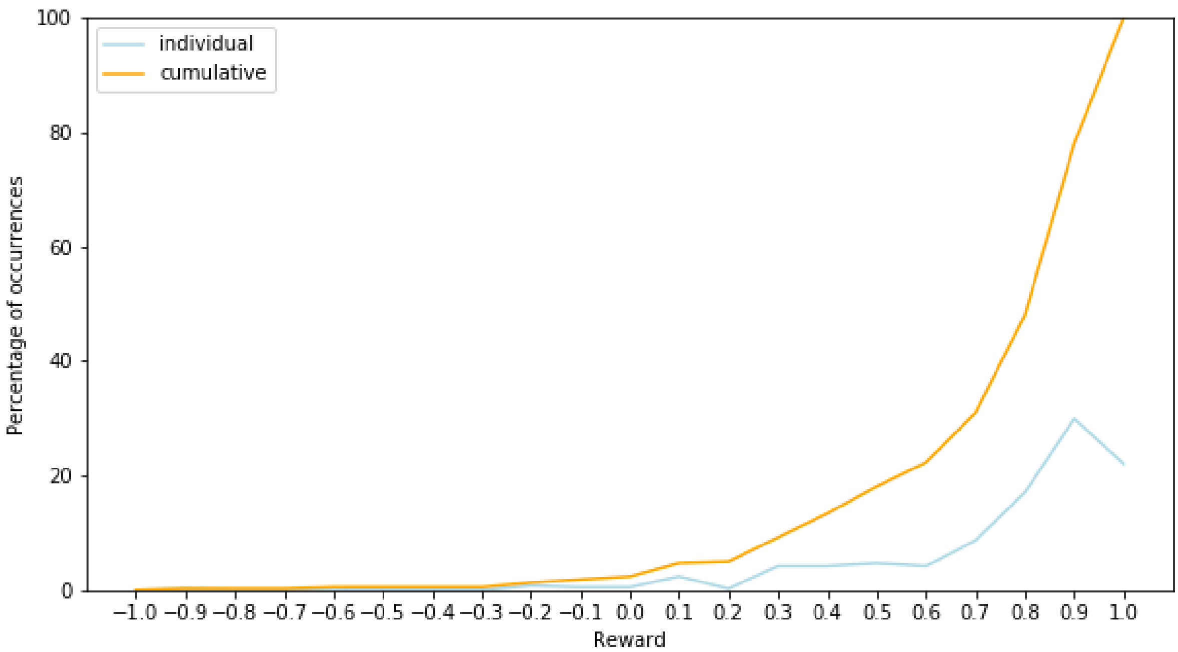


Figure 4.6: The percentage of occurrence of each rounded reward in the data set. The actual percentage and the cumulative percentage are shown.

Despite reaching our target sample size, one of the states had very few data points, as shown in Appendix D. The most occurring state over all the conversation sessions was the state [1,1,1] (112 times) and the least occurring was [1,0,0] (5 times). Next to this, the *slightly decrease* action was taken the most amount of times (80) while the *slightly increase* action was taken the least (73). The distribution of the rounded rewards is shown in Figure 4.6. It can be seen that a lot of people got relatively close to their recommended step goal, as more than 70% of the rewards were between 0.7 and 1. Appendix E shows the distributions of the rewards per state, indicating that each state had a large proportion of samples with a relatively high reward. Next to this, we calculated the optimal policy using the default reward function, which can be seen in Appendix F. As shown, different states had different optimal actions to take, but the *decrease* action was never the optimal action to take.

To investigate the trained reinforcement learning model further, we discuss each of the earlier explained analysis questions. For each question, we first explain the setup to answer the question and then our results.

AQ1: How well do the states predict behaviour after proposing personalized step goals?

Setup. It could be that in different states, people behave differently to the actions we choose. In our case, this behaviour is the number of steps people take on the day that they set a step goal. Since we ideally wanted to propose step goals that made people exactly reach their recommended step goal, we wanted to see if we could predict how much people would over- or underachieve compared to their recommended step goal of that day. This is represented by our reward function described in the previous chapter. So, to see if we could predict how much people would over- or underachieve compared to their recommended step goal, we had to see if we could predict the reward for a given data sample. We, therefore, looked at predicting the reward based on 1) the action and 2) both the state and the action of a data sample. We used leave-one-out cross-validation to leave out all the samples of a single person and predict the rewards of those samples. We did this for every person and calculated the mean $L1$-error for the reward prediction and its 95% credible interval (CI) [58] per state. If the mean of one of the approaches was outside of the credible interval of the other approach, we considered it a credible indication that the values were different. This was changed compared to what we originally wrote in our OSF registration [20], where we said that we would only consider a credible indication if both intervals were not overlapping. However, we figured that that approach was too conservative, which is why we changed it slightly.

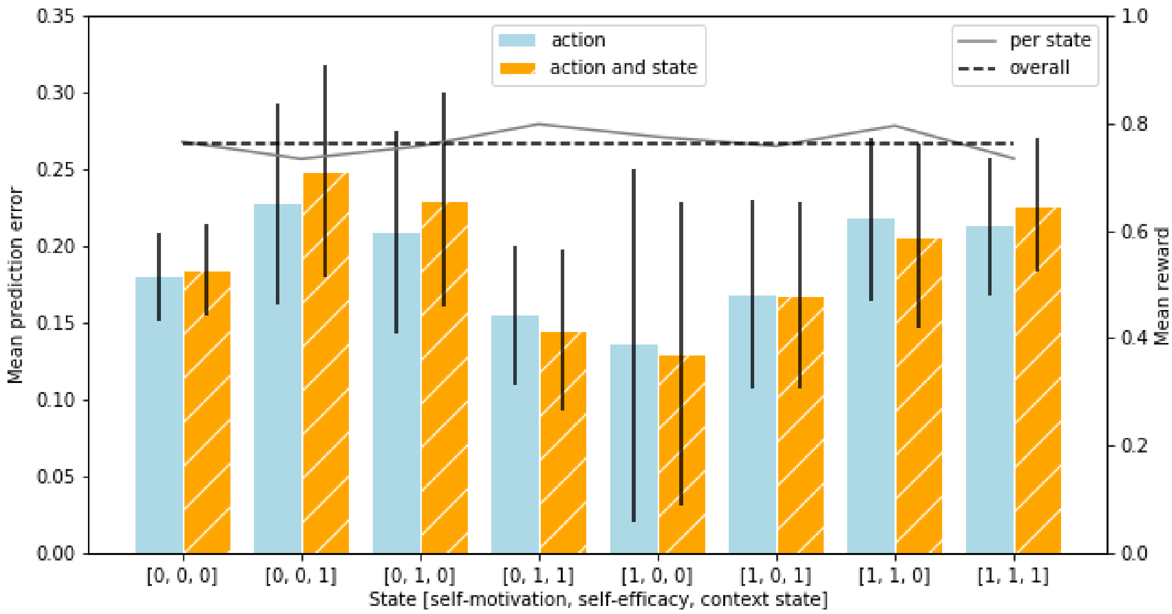


Figure 4.7: The left axis displays the mean prediction error per state when looking at only the action or both the state and the action. The right axis displays the mean reward for each state and overall.

Results. As can be seen in Figure 4.7, predicting the reward based on both the action and state did not necessarily give a lower error than basing it purely on the action. However, this result was not conclusive as the mean prediction error of one approach was inside the credible interval of the other approach. What can also be seen from Figure 4.7 is that the mean reward per state was the highest for states [0,1,1] and [1,1,0], while the lowest mean reward was in states [0,0,1] and [1,1,1]. However, the differences were very small. Finally, the figure shows that predicting the rewards for the people in the state [1,0,0] gave the lowest prediction error, while state [0,0,1] gave the highest.

AQ2: How well do the states predict the next states after proposing personalized step goals?

Setup. By proposing a personalized step goal, we wanted people's state to change to a different state in which they were more likely to follow the recommended step goal each day and, with that, made more progress towards the long-term goal of 10,000 steps per day. So, to be able to cause this state change for people, we needed to know if we were able to predict people's next state. We again used leave-one-out cross-validation to compare three ways of predicting the next state of the samples of the left-out person. We compared 1) using an equal probability for all states, 2) predicting that people have the same next state as their current state, and 3) using a probability for each next state based on the calculated transition function from the training data. For each approach, we calculated the mean-likelihood of each next state and their 95% credible intervals. Here, we also consider the mean of one of the approaches being outside of the credible interval of the other approach to give a credible indication that the values are different.

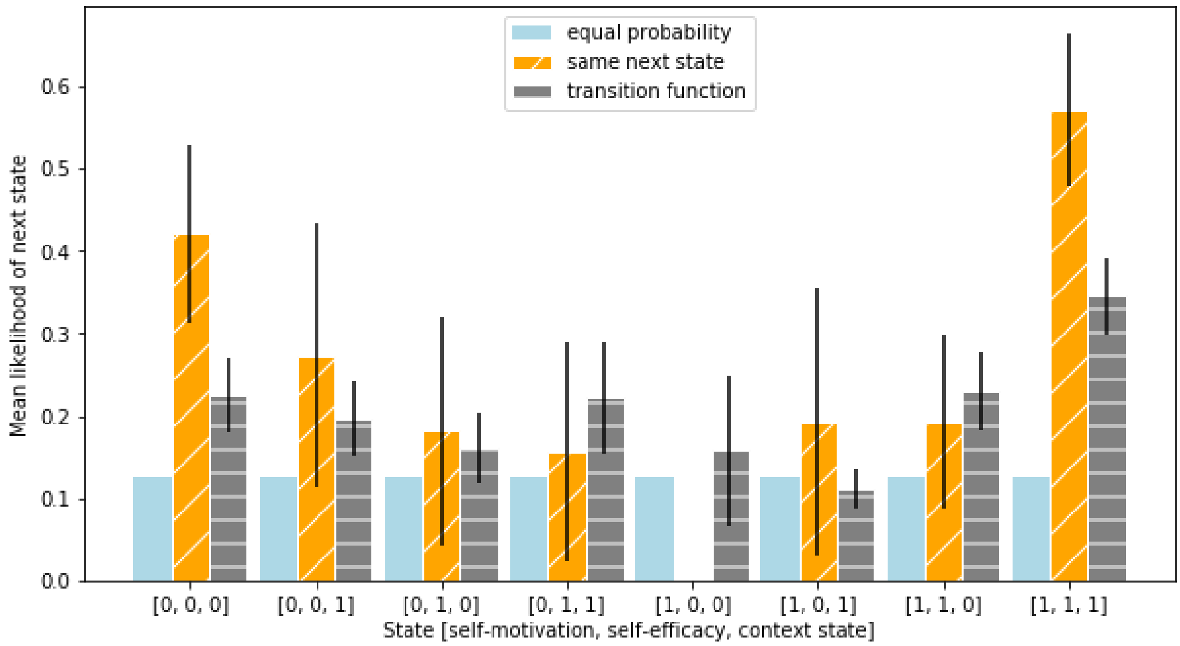


Figure 4.8: The mean likelihood of a transition to a next state when using an equal probability for each next state, using the same next state as the current state, and using the calculated transition function.

Results. The results from predicting the next state based on the current state were inconclusive for three states because the mean likelihood for predicting with equal probability for each next state was inside the credible intervals of the other approaches, as can be seen in Figure 4.8. For the other states, it can be concluded that either predicting that people stay in the same state or transition according to the calculated transition function after proposing a personalized goal was more accurate than predicting the next state being a random one. For states [0,0,0], [0,0,1], [1,0,1], and [1,1,1] there is a credible indication that predicting that people stay in those states is more accurate than predicting that they transition according to the transition function. For states [0,1,1] and [1,0,0] it was the other way around. For the latter state, state [1,0,0], the few samples we had for it did not have people stay in that state which is why the bar for that approach is empty.

AQ3: What is the effect of (multiple) optimal step goal proposals on users' states?

Setup. After looking at whether we could predict the change in people's states, we also wanted to see how the states would change when proposing step goals according to the optimal policy. This is interesting, as we wanted to see whether people's states changed towards the states where they were more likely to exactly take the recommended number of steps per day. To do so, we used the training data to calculate the optimal policy, which can be found in Appendix F, and simulated people's change of state over time. We simulated 8,000 people who were evenly distributed over each possible state (so 1,000 people per state) and had them follow the optimal policy for 20 timesteps. For each timestep, we checked the distribution of people across the states.

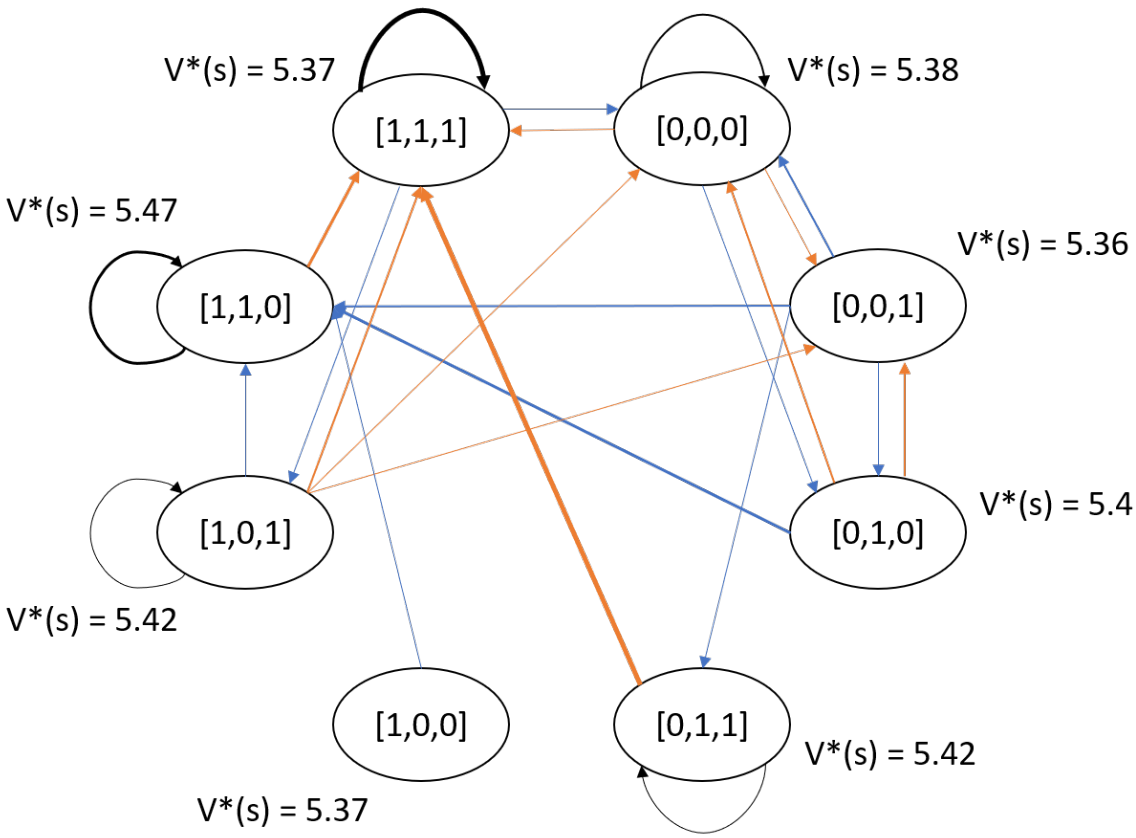


Figure 4.9: Transitions from each state to their next states. Red arrows indicate next states with a lower value, black arrows indicate transitions to the same state, and blue arrows are transitions to a next state with a higher value. The thicker the arrow, the higher the probability of the transition.

Results. Figure 4.9 displays the transitions where the probability of the transition was higher than $\frac{1}{|S|}$ for each of the states following the optimal policy for one timestep. The cumulative discounted reward for taking the optimal action in the state s (V*(s)) is shown for each of the states as well. It can be seen that the optimal policy had many transitions to state [1,1,1], including a high probability of staying in that state, while the value of that state was lower than for most of the other states. However, it can also be noted that the difference in V*(s) between the states was very small, as the values only ranged from 5.36 to 5.47. Next to that, for the states where only one of the features was high, there were no major transitions to stay in those same states but only transition towards other states. The distribution of people after multiple timesteps of following the optimal policy is displayed in Figure 4.10. It can be seen that following the optimal policy transitioned people generally to the state where all state features were high, being [1,1,1]. This meant that people generally got to a state where their self-motivation, self-efficacy and context state were above average. Next to this, the state [1,0,0] had very few transitions towards it, so after a couple of timesteps there were very few people left as they all transitioned to other states.

AQ4: How do optimal and sub-optimal step goal proposals compare in their effect on behaviour?

Setup. When we knew the effect of the optimal policy on both short and long-term state changes, we wanted to see whether different policies had different effects on the behaviour of people. Different policies could give different step goal proposals which could make people get closer to their recom-

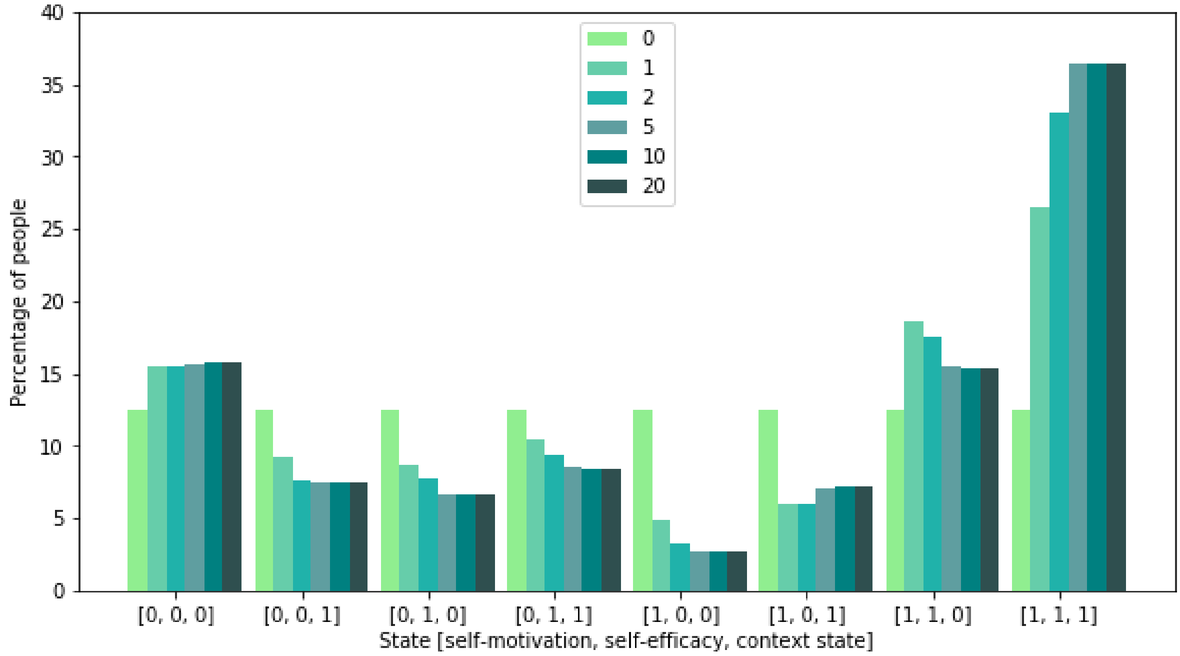


Figure 4.10: Percentage of people in each state after simulating multiple time steps of following the optimal policy.

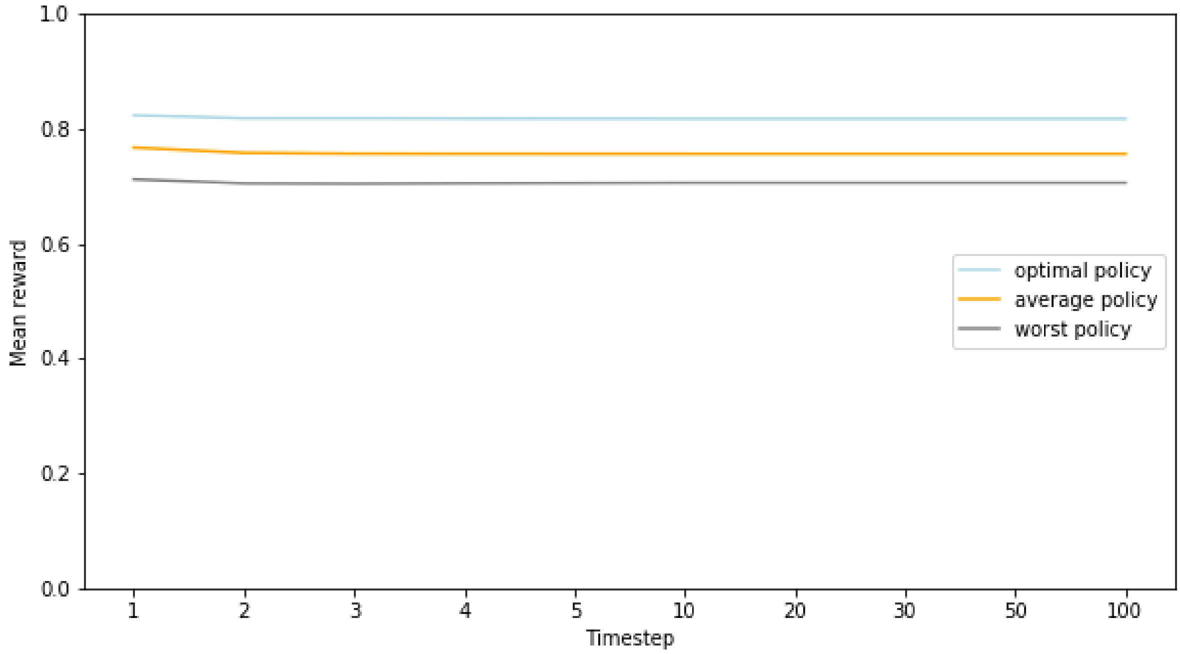


Figure 4.11: Mean reward per timestep while simulating the optimal policy, the average policy, and the worst policy.

mended step goal. To see this effect, we calculated the mean reward on different timesteps following three different approaches: 1) the optimal policy, 2) the average policy, and 3) the worst policy. The optimal policy and the worst policy can be found in Appendix F. For the average policy, we considered actions such that each action was taken an equal amount of times per timestep. We again simulated people, this time we distributed our 117 participants across the states according to the state of people in the first session of our study to simulate a realistic starting distribution across the states.

Results. As shown in Figure 4.11, following any of the policies decreased the mean reward per timestep slightly in the first few timesteps but kept it roughly the same over a longer period. The difference in reward between the optimal policy and the average policy was roughly 0.06. For the worst policy, this gap was even bigger, with a difference of roughly 0.12 compared to the optimal policy. So, following the optimal policy always gave the highest mean reward, and following the worst policy always gave the lowest mean reward while following the average policy kept the rewards on an average level. However, the difference between the mean rewards following the different policies was very small.

AQ5: How do different reward functions compare in their optimal policy?

Setup. We used a default reward function to answer the previous analysis questions, however, we also collected data on other reward signals of which we wanted to see whether they give a different optimal policy. So, we calculated the optimal policy using four different reward functions:

1. The default function, which was $R1 : r = 1 - (\Delta/recommended\_step\_goal)$ where Δ is the absolute difference between the recommended step goal and the actual number of steps taken. If the goal was attained, this Δ was divided by two as taking more steps than the goal was more desired than taking fewer steps.
2. Including the perceived goal achievability, which was $R2 : r = 1-(\Delta/recommended\_step\_goal)+ 0.1 * perceived\_goal\_achievability$. Since the perceived goal achievability would go from 0 to 10, we gave it a weight of 0.1 to have it add 1 to the reward function at maximum, which was also the amount that the $\Delta/recommended\_step\_goal$ could subtract from the reward if no steps were taken at all.
3. Including a standard reward based on the action taken, which gave $R3 : r = standard\_reward- (\Delta/recommended\_step\_goal)$ where the standard reward was 1 for the *increase* action, 0.95 for the *slightly increase* action, 0.9 for the *do nothing* action, 0.85 for the *slightly decrease* action, and 0.8 for the *decrease* action. This was done to give preference to the *increase* action to try to make the goals a bit higher when possible.
4. Including both the perceived goal achievability and a standard reward per action, which was $R4 : r = standard\_reward - (\Delta/recommended\_step\_goal) + 0.1 * perceived\_goal\_achievability$.

To compare for similarities between the optimal policies of the different reward functions and the default reward function, we used the Weighted Kappa [17]. We compared the list of optimal actions for each state using linear weights which made it so that the similarity between an *increase* and *slightly increase* action, for example, was higher than between an *increase* and *decrease* action.

State	Default function	Goal achievability	Standard reward per action	Goal achievability + Standard reward per action
s[0,0,0]	+	+	+	++
s[0,0,1]	++	++	++	++
s[0,1,0]	-	- -	++	+
s[0,1,1]	-	-	0	-
s[1,0,0]	- -	- -	+	+
s[1,0,1]	-	-	++	++
s[1,1,0]	+	+	++	+
s[1,1,1]	++	-	++	++

- - is the *decrease* action
- is the *slightly decrease* action
0 is the *do nothing* action
+ is the *slightly increase* action
++ is the *increase* action

Table 4.3: The optimal action per state using different reward functions, namely the default reward function, adding perceived goal achievability to the reward function, adding a standard reward per action to the reward function, and adding both perceived goal achievability and a standard reward per action to the reward function.

Results. The optimal action to take in each state while using the different reward functions is depicted in Table 4.3. As can be seen, adding the perceived goal achievability to the reward function gave a policy that was very similar to the default reward function with only the action for state [1,1,1] being very different. Their similarity had a kappa score of 0.67 which, according to Landis and Koch [42], meant a substantial agreement. Then, looking at the policy from the reward function with a standard reward per action, only one *do nothing* action was suggested and the rest were only *increase* and *slightly increase* actions. The similarity between this reward function and the default one has a kappa score of only 0.19, meaning that this policy was very different from the policy of the default reward function, having a slight agreement according to Landis and Koch. Third, when looking at the policy from the reward function where both the perceived goal achievability and the standard reward per action were added, it can be seen that it was somewhat similar to the policy from the default reward function but it had in general more *increase* and *slightly increase* actions. The similarity measure between this most complicated function and the default reward function was 0.33, meaning that it was a fair agreement on the scale from Landis and Koch. Next to this, it can also be seen that the fourth reward function was more similar to the function with only the standard reward per action than the function with only the perceived goal achievability, with kappa scores of 0.41 and 0.16 respectively. Finally, it can be seen that state [0,0,1] was the only state for which the policies from all the reward functions suggested the same optimal action being *increase*.

4.3. Discussion of the results

Looking at the fact that two-thirds of the goals were achieved and that almost no one indicated that the goals were difficult, we could say that the goals were relatively low in general. Since the *decrease* and *slightly decrease* actions were not taken that much more often than the other actions across all the samples, the relatively low goals were probably because the recommended goals themselves were already on the lower end. This can also be concluded from the fact that in all the states, at least 20% of the people took more steps than the recommended goal and in some states, people did on average 50% more than recommended. Despite this, the goal-setting was found to be effective as the average steps of people who completed the whole study increased by 818 steps (18%). Compared to a study from Adams et al. [2], this increase was the same as the group with a static goal instead of the group with an adaptive goal, which got a 48% increase. This difference could be explained by the fact that the study of Adams et al. ran for 180 days and we only had five days. So, it could be that, over time, this increase in steps would grow further to also hit the 48%. Additionally, the three states that had the highest percentage of steps taken over the recommended goal were states where the context state was high. This could be because the context state consisted of features that resemble barriers to physical activity. Therefore, if the context state was high, people were less affected by those barriers which could have allowed for way more steps than recommended.

Next to the general discussion, we discuss our findings for each of the analysis questions below.

AQ1: How well do the states predict behaviour after proposing personalized step goals?

Predicting rewards based on the action taken was roughly as good as predicting them based on both the action and the state because the mean prediction errors for both approaches were roughly similar. However, this result is not credible as the mean of one of the approaches fell in the credibility interval of the other approach. This could mean that more data is needed to get better insights into the effect of states on predicting the rewards. Additionally, making predictions with very few samples, such as in state [1,0,0], creates uncertainty which could have increased the size of the credible intervals. Next to this, the state with the least number of samples, state [1,0,0], also had the lowest prediction error, which probably was because the few samples that were in that state all had a relatively similar reward, making them easier to predict. Finally, the differences in mean rewards were very small between states. This could mean that across all the states, people got similarly close to their recommended goal in general. However, the closeness of rewards also partially happened because of the high percentage of overachieving of the goals, depicted in Figure 4.5. States [1,0,1] and [1,1,1], for example, had a very high percentage of people overachieving and a very high percentage of the amount of overachieving. This caused the rewards to be lower, as the reward function also slightly punishes people to do more than the recommended goal. Figure 4.12 shows that when removing this penalty for overachieving, the mean rewards for the states with features being high increased in comparison to the states with low features. Also, the importance of considering states for predicting the reward for states with higher features becomes clear, as states [1,1,0] and [1,1,1] had a credible indication that predicting the reward based on both the action and state gives a lower error. So, despite having a relatively lower mean reward in our model, the states with the higher state features had more people achieve their recommended goal. Therefore, there is in fact a difference in the behaviour of people when in one of the states with low features versus when in states with high features.

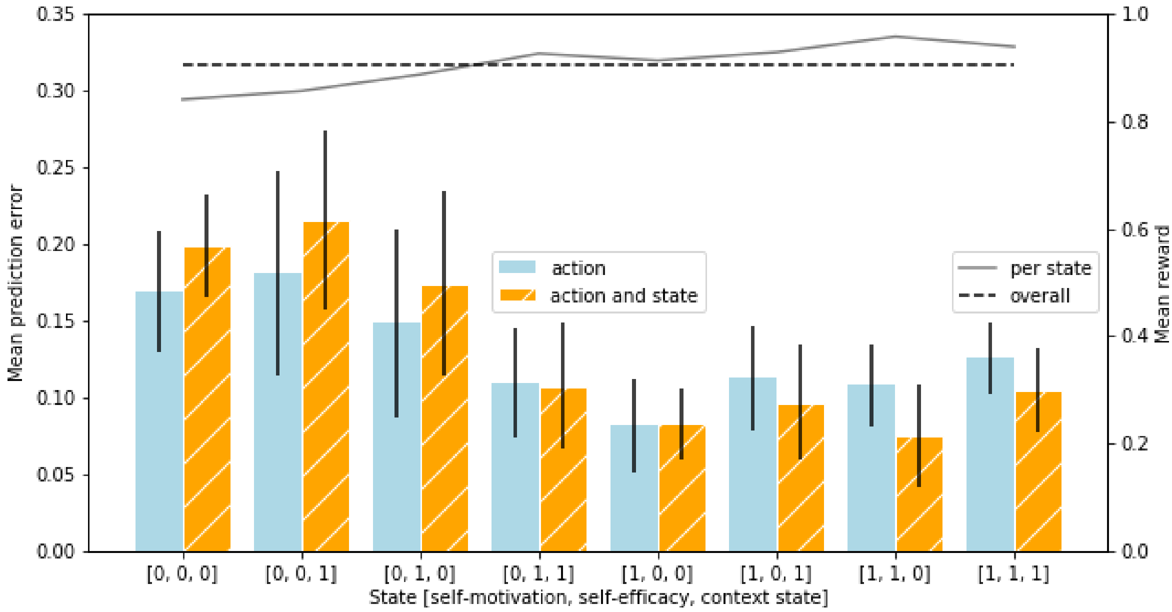


Figure 4.12: Reward prediction when the reward function no longer penalizes overachieving. The left axis displays the mean prediction error per state when looking at only the action or both the state and the action. The right axis displays the mean reward for each state and overall.

AQ2: How well do the states predict the next states after proposing personalized step goals?

The fact that, for most states, we were able to conclude that predicting to either stay in the same state or transition according to the transition function was better than predicting a random one, meant that we were generally able to use states to predict next states. However, since the mean likelihood of the next state when using the calculated transition function was never higher than 0.35, it means that there is quite some variability in where people transition to for each state. Also, the higher mean likelihood of predicting to stay in certain states, such as [0,0,0] and [1,1,1], indicates that for those states it might be more difficult to change people's states. Next to this, the difference in the size of the credible intervals of the prediction to stay in the same state and the prediction of transitioning according to the transition function could be explained by the difference in probabilities. Predicting that people stay in the same state either gave a zero, people did not stay, or a one, people stayed in that state. The mean likelihood resulted in an estimate of the probability but the two different values to get to this made the intervals stretch out quite much. The predictions from the transition function could take values anywhere between zero and one, so there was less variability in those values, therefore making the credible intervals smaller.

AQ3: What is the effect of (multiple) optimal step goal proposals on users' states?

Every timestep, a lot of people transitioned to state [1,1,1], of which the value was lower than for most other states but that difference was very small. As explained earlier, this was because people took more steps than the recommended goal in the states with higher features which lowered the rewards, despite people achieving the goals. So, people transitioning to the state [1,1,1] was good in the sense that it was one of the states where people were more likely to achieve their recommended step goal. But at the same time, it was bad in the sense that people were more likely to massively overdo their walking which could increase the risks of injuries. Next to people transitioning to state [1,1,1], people were also likely to stay in that state, which could mean that people generally remained motivated and confident to achieve their daily recommended step goal. Apart from state [1,1,1], people also tended to transition to the state [0,0,0] indicating that for some people the goal-setting would lower their motivation and confidence. This could partially be because the goals were generally found to be easy and easy goals are less motivating than more difficult ones [47]. So, making the goals more difficult could have transitioned fewer people to the state [0,0,0]. It could also just be because some people had a bad day, did not provide their state in an honest way, or were less motivated and confident in general. State [1,0,0] had very few transitions to it and the people who started in that state almost all transitioned away from it, which meant that people generally did not have just a high self-motivation. Instead, if people had high motivation, they would also be confident or have a high context state, as can be seen from the sample distribution in Appendix D. In a study by Neace et al. [54], the authors said that exercise self-efficacy positively influences exercise motivation. This relationship could explain why both states where self-motivation was high and self-efficacy was low had the least number of samples, as it appears that high self-efficacy was needed to have high self-motivation but not the other way around.

AQ4: How do optimal and sub-optimal step goal proposals compare in their effect on behaviour?

The mean reward over time following different policies is roughly similar which was because of the small differences in mean rewards per state as shown in Figure 4.7. Therefore, even though people transitioned to different states following the worst policy compared to the optimal one, as shown in Figure 4.13, it did not change the reward much. Next to that, because the distribution of people across the states did not change much after two timesteps, as shown in Figure 4.14, the mean reward did not change much either over time. The difference of 0.12 between the mean reward of following the optimal and worst policy meant that people following the worst policy were on average 12% further off of their recommended goal than people following the optimal policy. Hence, there was little difference between the overall mean reward following different policies. However, given that people in the state [0,0,0] were less likely to achieve their recommended step goal, as depicted in Figure 4.5, the difference in policies does have a larger impact on the behaviour than shown by the reward comparison, as now more people would fall short on their goal. So, despite not being directly visible in the mean rewards, there is a difference in behaviour comparing optimal and sub-optimal policies.

AQ5: How do different reward functions compare in their optimal policy?

The policy from the reward function with goal achievability was relatively similar to the policy from the default reward function. This indicated that the policy from the default reward function created goals that were not only getting people the closest to their recommended goal but were also found to be achievable in general. Only in the state [1,1,1], the policies from the two reward functions were very different, where the policy from the reward function with goal achievability would suggest a lower goal, meaning that increasing the goals in that state felt less achievable to people. In contrast, the policy from the reward function with the standard reward per action was not very similar to the policy from the default reward function. This made sense, as the reward function was now pushed towards *increase* and *slightly increase* actions and therefore those were chosen more often. But these actions did make it so that people's actual number of steps taken did not get too far off the recommended step goal, otherwise the reward would have been lower again and the optimal policy would not have suggested those actions. So, this might also point towards the fact that the proposed goals were a bit low in general and could have been increased and a little nudge in the reward function could have helped with that. Similarly, the policy from the reward function with both goal achievability and the standard reward per action was not that similar to the policy from the default reward function. Finally, the combined reward function gave a policy that was more similar to the one from the reward function with the standard reward per action instead of the one from the reward function with the goal achievability. This meant that, with the reward function being pushed to suggest *increase* actions, the perceived goal achievability did not get so low that it was better to not actually use the *increase* action, but rather pick an action that gets people a higher feeling of achievability. So, in general, increased goals were also found to be quite achievable.

All in all, our reinforcement learning model with binary states got people to transition towards and stay in states with higher self-motivation and self-efficacy following the optimal policy. Despite the results for AQ1 being inconclusive, from the fact that the optimal policy did not suggest the same action for every state, it can be concluded that the states were important in choosing which action to take

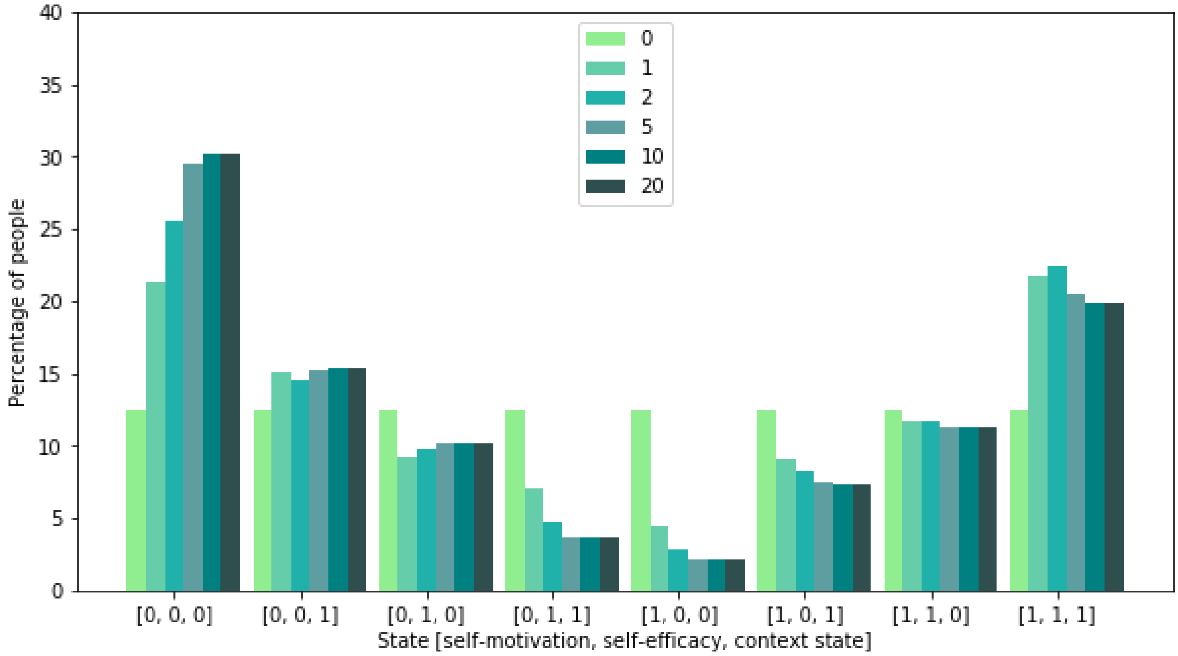


Figure 4.13: Percentage of people in each state after simulating multiple time steps of following the worst policy.

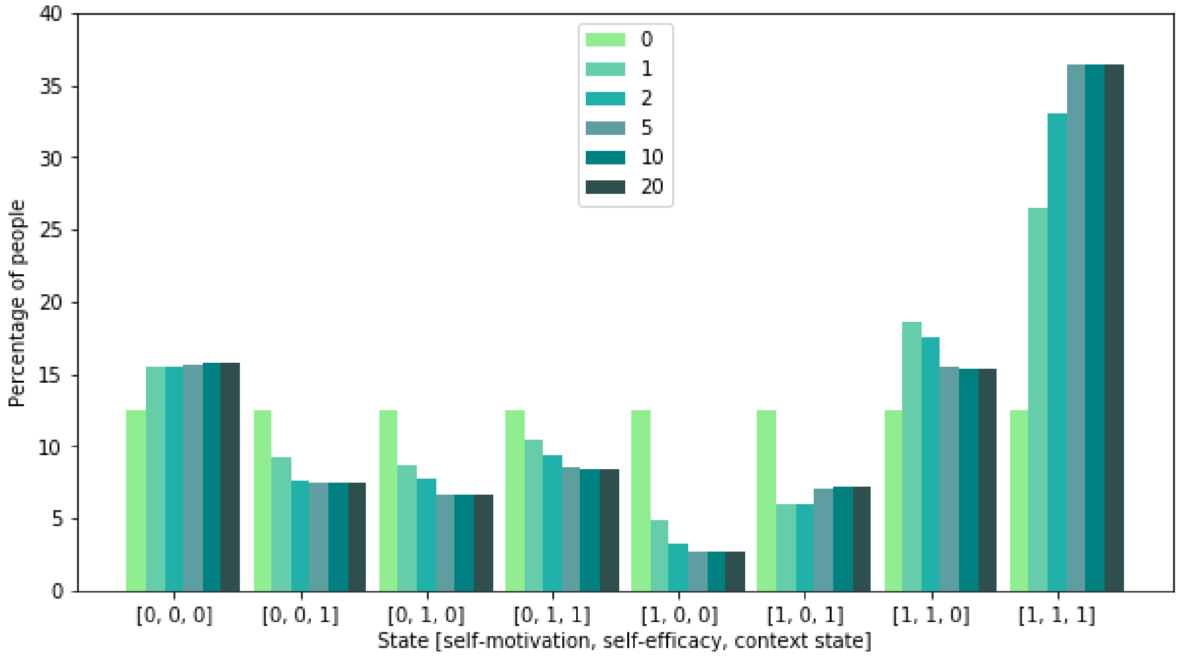


Figure 4.14: Percentage of people in each state after simulating multiple time steps of following the optimal policy.

for optimal rewards. On top of that, the optimal policy suggested actions that created goals that were generally found to be achievable, even though the reward function did not take that into account. Also, picking the actions from the optimal policy was more effective than picking random actions, as the mean reward was lower for the latter. Adjusting the reward function to not penalize overachieving would make more clear differences between the rewards for states with low features (e.g. [0,0,0]) and states with high features (e.g. [1,1,1]). However, the states with high features had a large proportion of people overachieving which could lead to injuries. Especially since the target group of our model was people who were inactive.

4.4. Limitations

One limitation we identified in our observational study is the target group of participants. Our participants were only people who resided in certain time zones, which excluded people living in, for example, the USA or China. On top of this, we targeted people between 18 and 65 years old but our oldest participant was only 56 years old and the mean age was only 28. This meant that our data may not generalize well for the full world population, so it would be interesting to gather data from people in other time zones and from other ages to compare the results and see whether our conclusions generalize over a larger population.

A second limitation is that, despite reaching our target sample size, we were still lacking data for some of the state-action pairs, as shown in Appendix D. We compensated for this by imputing average data, which may not be representative of actual people in those states. Therefore, the results of the analysis questions could be slightly skewed. So, a further data collection study is needed to gather data on the missing state-action pairs to complete the results.

Another limitation of our observational study is our approach to congratulating and encouraging people who were doing more steps than their recommended goal. In the data analysis, we showed that a relatively large group of people were overachieving and we had not designed our virtual coach to account for that. Instead, we complimented them for achieving the goal which could have made it so that they would be more likely to keep overachieving. This may have caused the relatively large proportion of overachieving and might not have occurred if those people had been informed about it during the conversation sessions.

Finally, a limitation is that the data gathered on the primary and secondary measures was subjective. We did not, for example, collect rest data using people's heart rates and such because it would make data collection too complicated and subjective measures would give better insights into how people actually felt towards walking. So, instead, we asked people to indicate how well they slept, for example, meaning that dishonest answers from people could have skewed the data and our analysis. We also used mostly non-validated questions to gather this data which were, however, based on existing questions and definitions. Despite this, we could not always guarantee the validity of the questions.

5

Discussion and Conclusion

In this chapter, we discuss and conclude our work. First, the research questions are presented and answered. Then, the contributions of this thesis are discussed. After that, a discussion on the limitations is given, and finally, we give suggestions for future research.

5.1. Conclusion

The main research question was as follows:

How can reinforcement learning be used to personalize daily step goals for a collaborative dialogue with a virtual coach?

This research question was divided into three sub-questions which we answer separately. Our first sub-question was as follows:

- *What relevant factors and concerns arise when using reinforcement learning to personalize daily step goals for a collaborative dialogue with a virtual coach?*

We conducted an expert evaluation and explored previous work on goal-setting, goal personalization, and dialogues with virtual coaches. In total, we found eight relevant factors and concerns that arise with a virtual coach for collaboratively setting daily step goals that are personalized using reinforcement learning. We found that a person's commitment (FC1) and SMART goals (FC2) are important for effective goal-setting. We also found that a person's involvement in the goal-setting process (FC3) increases the commitment of someone to the goal because it feels more important to them. Next to this, we found that three factors and concerns need to be considered when personalizing a goal. The first of these is the consideration of people's personal factors that could change on a daily basis such as previous activity and self-efficacy (FC4), as the number of steps people take on a day could change depending on these factors. The second one is goal boundaries (FC5), as goals need to be of the right difficulty. Third, we found future goals of the intervention (FC6), as setting a goal on one day may impact the personal factors and physical activity of the coming days. Finally, we found that people's trust and willingness to depend on the advice of the virtual coach (FC7) is important for an effective interaction between a person and the virtual coach and that the bond between them (FC8) influences the motivation and commitment of a person.

Our second subquestion was as follows:

- *How can a reinforcement learning model and virtual coach be designed for daily collaborative personalized step goal-setting?*

Keeping the identified factors and concerns in mind, we designed a virtual coach that would collaboratively set daily personalized step goals with people. This conversation involved getting the personal factors of a person, proposing personalized step goal options and letting the people react and change the proposed options to their liking. We also adapted an algorithm from Adams et al. [2] to create the recommended step goals for people based on their previous activity and bound them to make them not too high or too low. Finally, we trained a reinforcement learning model to create three personalized step goal options based on the recommended step goal which made people more likely to reach their recommended step goal. For this, we used a combination of people's self-motivation, self-efficacy, rest, available time and mood (valence) to create the states. We selected the actions: increase, slightly increase, do not change, slightly decrease, and decrease of the recommended step goal for the model. The reward function was made by dividing the absolute difference between the recommended step goal and the actual activity of the person towards that goal with the recommended goal, where that absolute difference was halved when the recommended step goal was reached. This value was subtracted from one to get the reward which could be interpreted as the relative closeness of people to their recommended goal.

Our third and final subquestion was the following:

- *How effective is reinforcement learning for personalizing daily step goals for a collaborative dialogue with a virtual coach?*

To answer this question, we ran an online observational study where we collected data on state features, actions, and reward signals. With that data, we created people's states with self-motivation and self-efficacy and the context state which consisted of rest, available time and mood. All three state features were binary, so either a 0 or a 1. So we ended up with eight possible states ranging from [0,0,0] to [1,1,1]. After training the model, we found that for different states, our model used different actions to create goal proposals, indicating that there was not one best action to take overall and the creation of the goal proposals should be done based on the state of a person. Also, we found that the model creates step goal options that generally increase people's self-motivation and self-efficacy. However, we also found that people with high self-motivation and self-efficacy tended to overachieve a lot, meaning that they did more steps than the recommended goal which could lead to injuries. So, when purely looking at people achieving goals, the model is effective for making people more likely to achieve their recommended goals. However, for our target group, it might be less effective because of the risk of injuries associated with overachieving. Next to this, we found that accounting for the perceived goal achievability of people did not change the goal proposals that much compared to our base model, so our model was effective in creating goals that were perceived as achievable. Finally, proposing step goal options that were not personalized following the optimal policy of our model, would lower people's self-motivation and self-efficacy more often. This meant that people would be more likely to not achieve their recommended goal if the wrong action was taken for people in a certain state.

5.2. Contributions

Our first contribution is an overview of insights into collaborative goal-setting, personalization of step goals, and goal-setting with a virtual coach, which we gathered from our literature review and expert consultations. This information can be used in future research regarding virtual coaches and daily goal-setting in the context of physical activity. Additionally, we contribute insights into the setup and possibilities of reinforcement learning for personalizing goals and changing people's behaviour. These insights could, for example, form the inspiration for future research on reinforcement learning in health behaviour change.

A second contribution is the implementation of a virtual coach for collaboratively setting daily personalized step goals. The setup of the messages and the personalization of the step goals could be used as a basis for other virtual coaches for goal-setting or other software with goal-setting functionalities. We also contribute the implementation of a reinforcement learning model for personalizing step goals. This implementation could be used to be extended in further research into personalized goal-setting.

Thirdly, we contribute a fully functional virtual coach for daily step goal-setting. It is able to have a conversation and set a step goal which allows it to be used as part of an intervention to get people walking, for example. The virtual coach is also able to collect and store data which might be useful for further research on goal-setting and people's walking behaviour.

5.3. Limitations

One of our limitations is that we did not use our final model to actually create personalized step goals in order to evaluate it. Instead, we used random actions during the observational study to collect data and train the model and only hypothesised about the effectiveness of the reinforcement learning model for personalizing daily step goals. However, we do not know for certain if the results (e.g. the increase in people's self-motivation and self-efficacy) would be the same if the actual model was used to pick actions to create step goals for people in different states.

Next, a limitation of our study is that we only considered step goals to increase people's physical activity. We did not investigate, however, to what extent taking steps supports the increase in the overall level of physical activity. Also, we focussed on short-term goal-setting over a couple of days but do not know if our results of increasing the number of achieved goals and increasing the amount of overachieving generalize well over a longer period of time.

Another limitation is that we only consulted three experts and focussed on the literature on physical activity goal-setting and virtual coaches, which could have caused a biased viewpoint on certain topics, such as the level of participation in goal-setting. This could mean that our final model and virtual coach may not be directly applicable to other domains of goal-setting, such as mental health goals or study goals, as these might require different levels of participation of people in the goal-setting. Also, our view on the lower and upper limits of the step goals may not be suitable for different groups of people, such as more active people or people with chronic illnesses, who might require different limits.

Finally, we only considered self-motivation, self-efficacy, rest, available time and mood (valence) as state features. These factors can change on a daily basis which makes them useful as state features in our model for daily personalized step goal-setting. However, there might be other factors that also influence the physical activity of people that could be included in the model, such as people's social environment [60, 70]. Since we only looked at factors that could change on a daily basis, we did not include those in our model. Additionally, our research into our five state features was limited as we did not look at factors influencing our state features, for example, the effect of the attitude of sleep on self-reported sleep quality [71], even though these factors might predict people's behaviour more effectively. We also did not look into the effect of the state features on each other, e.g. the effect of exercise self-efficacy on exercise motivation [54], and their possible combined effect on people's behaviour which might have limited our results.

5.4. Future work

The first direction future work can take is to investigate the notion of overachieving we found in our study. Currently, our model gets people to take more steps than their recommended goal, which increases the risk of injuries. However, little research is done on what exact amount of walking is too much for an individual or what an effective distribution of exercise throughout the day is [8]. On top of this, future work could look into ways to monitor and provide feedback to people to reduce overachieving using monitoring systems that look at, for example, heart rate, duration of walks, and frequency of walks [24]. Also, the notion of overtraining syndrome, where training too much would cause negative effects, is still not fully explored [9].

Secondly, the implementation of our virtual coach could be used in future work to be integrated into a larger system to see whether this way of collaborative goal-setting is useful in a larger setting. It could then also be investigated what the long-term effect of personalized step goals is on a person's behaviour.

Finally, future work could experiment with more different reward signals and state features to see if the model could be improved. Especially because the difference in the number of steps people take per day is quite large between people [25, 34]. Personality, for example, is found to be correlated to physical activity [89], so future work could consider adding certain personality traits as features into the model. Additionally, there could be future research done towards the reward function in our model to see if there is a more sophisticated function for people with different kinds of activity levels. Also, since some state-action pairs had no data in our study, more data could be collected on the current model to get more conclusive results.

5.5. Final remarks

In this thesis, we presented a reinforcement learning model to personalize daily step goal proposals for a collaborative dialogue with a virtual coach. From simulations, we found that adjusting people's step goal proposals based on their daily personal factors (e.g. mood) in an optimal way could make people more likely to achieve their daily recommended number of steps. We also found that people with higher values of personal factors (e.g. a higher mood) were more likely to do more than their daily recommended number of steps. Finally, we made suggestions for future work to investigate this notion of overachieving to better understand our results.

Bibliography

[1] B. Abernethy et al. *Biophysical foundations of human movement 3rd ed*. Human Kinetics, 2012. ISBN: 9781450431651.

[2] M. A. Adams et al. "An Adaptive Physical Activity Intervention for Overweight Adults: A Randomized Controlled Trial". In: *PLOS ONE* 8.12 (2013). DOI: 10.1371/journal.pone.0082901.

[3] N. Albers, M. A. Neerincx, and W. P. Brinkman. "Addressing people's current and future states in a reinforcement learning algorithm for persuading to quit smoking and to be physically active". In: *PLoS ONE* 17.12 (2022). DOI: 10.1371/journal.pone.0277295.

[4] N. Albers et al. "Attitudes Toward a Virtual Smoking Cessation Coach: Relationship and Willingness to Continue". In: *International Conference on Persuasive Technology*. 2023, pp. 265–274. DOI: 10.1007/978-3-031-30933-5_17.

[5] N. Albers et al. "Setting Physical Activity Goals with a Virtual Coach: Vicarious Experiences, Personalization and Acceptance". In: *Journal of Medical Systems* 47.15 (2023). DOI: 10.1007/s10916-022-01899-9.

[6] S. Atoui et al. "Daily associations between sleep and physical activity: A systematic review and meta-analysis". In: *Sleep Medicine Reviews* 57 (2021), p. 101426. DOI: 10.1016/j.smrv.2021.101426.

[7] T. Bodenheimer and M. A. Handley. "Goal-setting for behavior change in primary care: An exploration and status report". In: *Patient Education and Counseling* 76 (2009), pp. 174–180. DOI: 10.1016/j.pec.2009.06.001.

[8] M. Cabrita et al. "Automated Personalized Goal-setting in an Activity Coaching Application". In: *Proceedings of the 3rd International Conference on Sensor Networks* (2014), pp. 389–396. DOI: 10.5220/0004878703890396.

[9] F. A. Cadegiani and C. E. Kater. "Novel insights of overtraining syndrome discovered from the EROS study". In: *BMJ Open Sport & Exercise Medicine* 5.1 (2019), e000542. DOI: 10.1136/bmjsem-2019-000542.

[10] J. Cao. *Why Goal Setting Is So Hard*. 2022. URL: https://www.habitify.me/blog/why-goal-setting-is-so-hard (visited on 01/09/2023).

[11] J. M. Carroll, ed. *Making use: scenario-based design of human-computer interactions*. MIT Press, Cambridge, 2003, p. 382. ISBN: 0262032791.

[12] L. Ceci. *Number of health and fitness app users in the United States from 2018 to 2022*. 2021. URL: https://www.statista.com/statistics/1154994/number-us-fitness-health-app-users/ (visited on 12/15/2022).

[13] B. Chang. "Reflection in learning". In: *Online Learning* 23.1 (2019), pp. 95–110. DOI: 10.24059/olj.v23i1.1447.

[14] D. Chapman and L. P. Kaelbling. "Input Generalization in Delayed Reinforcement Learning: An Algorithm and Performance Comparisons." In: *Ijcai*. Vol. 91. 1991, pp. 726–731.

[15] B. C. K. Choi, A. W. P. Pak, and J. C. L. Choi. "Daily step goal of 10,000 steps: a literature review". In: *Clinical and Investigative Medicine* (2007), E146–E151. DOI: 10.25011/cim.v30i3.1083.

[16] J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, New York, 1977. ISBN: 0-12-179060-6.

[17] J. Cohen. "Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit." In: *Psychological bulletin* 70.4 (1968), p. 213. DOI: 10.1037/h0026256.

[18] PAR-Q+ Collaboration. *2023 PAR-Q+*. 2023. URL: http://eparmedx.com/ (visited on 06/12/2023).

[19] M. Dierikx. *Using Reinforcement Learning to Personalize Daily Step Goals for a Collaborative Dialogue with a Virtual Coach: code for the virtual coach*. Version 1.0. Sept. 2023. DOI: 10.5281/zenodo.8382413. URL: https://github.com/PerfectFit-project/virtual_coach_daily_step_goal_setting.

[20] M. Dierikx, N. Albers, and W. P. Brinkman. *Daily Collaborative Personalized Step Goal-Setting with a Virtual Coach*. June 2023. DOI: 10.17605/OSF.IO/6JQPK.

[21] M. Dierikx, N. Albers, and W. P. Brinkman. *Data and code underlying the master thesis: Using Reinforcement Learning to Personalize Daily Step Goals for a Collaborative Dialogue with a Virtual Coach*. Sept. 2023. DOI: 10.4121/6f8e6750-7494-4226-b6f9-299a9edbb077.

[22] T. Draws et al. "A checklist to combat cognitive biases in crowdsourcing". In: *Proceedings of the AAAI conference on human computation and crowdsourcing*. Vol. 9. 2021, pp. 48–59. DOI: 10.1609/hcomp.v9i1.18939.

[23] P. F. Drucker, ed. *The practice of management*. Harper & Row, New York, 1954, p. 404. ISBN: 9780060913168.

[24] Peter Düking et al. "Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies". In: *Frontiers in physiology* 7 (2016), p. 71. DOI: 10.3389/fphys.2016.00071.

[25] T. Dwyer et al. "The inverse relationship between number of steps per day and obesity in a population-based sample–the AusDiab study". In: *International journal of obesity* 31.5 (2007), pp. 797–804. DOI: 10.1038/sj.ijo.0803472.

[26] J. M. Dzierzewski et al. "Exercise and sleep in community-dwelling older adults: evidence for a reciprocal relationship". In: *Journal of sleep research* 23.1 (2014), pp. 61–68. DOI: 10.1111/jsr.12078.

[27] J. Feine et al. "Gender bias in chatbot design". In: *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers 3*. Springer. 2020, pp. 79–93. DOI: 10.1007/978-3-030-39540-7_6.

[28] S. Fitrianie et al. "The Artificial-Social-Agent Questionnaire: Establishing the long and short questionnaire versions". In: *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*. 2022, pp. 1–8. DOI: 10.1145/3514197.3549612.

[29] S. Ghanvatkar, A. Kankanhalli, and V. Rajan. "User Models for Personalized Physical Activity Interventions: Scoping Review". In: *JMIR Mhealth Uhealth* 7.1 (2019). DOI: 10.2196/11098.

[30] G. Godin and R. J. Shephard. "Leisure Time Exercise Questionnaire". In: *Canadian Journal of Applied Sports Sciences* (1997). DOI: 10.1037/t31334-000.

[31] K. Gunasti and T. Ozcan. "The role of scale-induced round numbers and goal specificity on goal accomplishment perceptions". In: *Marketing Letters* 30 (2019), pp. 207–217. DOI: 10.1007/s11002-019-09492-w.

[32] Y. Guo et al. "Designing for trust: A set of design principles to increase trust in chatbot". In: *CCF Transactions on Pervasive Computing and Interaction* 4.4 (2022), pp. 474–481. DOI: 10.1007/s42486-022-00106-5.

[33] M. Hamadouche et al. "Comparison of value iteration, policy iteration and q-learning for solving decision-making problems". In: *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*. 2021, pp. 101–110. DOI: 10.1109/ICUAS51884.2021.9476691.

[34] M. Hirvensalo et al. "Daily steps among Finnish adults: Variation by age, sex, and socioeconomic position". In: *Scandinavian journal of public health* 39.7 (2011), pp. 669–677. DOI: 10.1177/1403494811420324.

[35] S. Huang, L. Jin, and Y. Zhang. "Step by step: Sub-goals as a source of motivation". In: *Organizational Behavior and Human Decision Processes* 141 (2017), pp. 1–15. DOI: 10.1016/j.obhdp.2017.05.001.

[36] G. Jain et al. “Revisiting attribute framing: The impact of number roundedness on framing”. In: *Organizational Behavior and Human Decision Processes* 161 (2020), pp. 109–119. DOI: https://doi.org/10.1016/j.obhdp.2020.04.006.

[37] E. Kang et al. “Person-Centered Goal Setting: A Systematic Review of Intervention Components and Level of Active Engagement in Rehabilitation Goal-Setting Interventions”. In: *Archives of Physical Medicine and Rehabilitation* 103.1 (2022), 121–130.e3. DOI: 10.1016/j.apmr.2021.06.025.

[38] S. Y. S. Kimm et al. “Self-perceived barriers to activity participation among sedentary adolescent girls”. In: *Medicine & Science in Sports & Exercise* 38.3 (2006), pp. 534–540. DOI: 10.1249/01.mss.0000189316.71784.dc.

[39] H. J. Klein et al. “The Assessment of Goal Commitment: A Measurement Model Meta-Analysis”. In: *Organizational Behavior and Human Decision Processes* 85.1 (2001), pp. 32–55. DOI: 10.1006/obhd.2000.2931.

[40] D. Kollias et al. “Deep affect prediction in-the-wild: Aff-wild database and challenge, deep architectures, and beyond”. In: *International Journal of Computer Vision* 127.6-7 (2019), pp. 907–929. DOI: 10.1007/s11263-019-01158-4.

[41] A. Konrad et al. “Finding the Adaptive Sweet Spot: Balancing Compliance and Achievement in Automated Stress Reduction”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Seoul, Republic of Korea, 2015, pp. 3829–3838. DOI: 10.1145/2702123.2702512.

[42] J. R. Landis and G. G. Koch. “The measurement of observer agreement for categorical data”. In: *biometrics* (1977), pp. 159–174. DOI: 10.2307/2529310.

[43] J. de Lannoy. *The effect of chatbot personality on emotional connection and customer satisfaction*. 2017. URL: http://essay.utwente.nl/74166/.

[44] I. Lee et al. “Association of step volume and intensity with all-cause mortality in older women”. In: *JAMA internal medicine* 179.8 (2019), pp. 1105–1112. DOI: 10.1001/jamainternmed.2019.0899.

[45] M. Lee et al. “Caring for Vincent: A Chatbot for Self-compassion”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* 702 (2019), pp. 1–13. DOI: 10.1145/3290605.3300932.

[46] P. Liao et al. “Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.1 (2020), pp. 1–22. DOI: 10.1145/3381007.

[47] E. A. Locke and G. P. Latham. “Building a Practically Useful Theory of Goal Setting and Task Motivation”. In: *American Psychologist* 57.9 (2002), pp. 705–717. DOI: 10.1037/0003-066X.57.9.705.

[48] Y. Lu, K. Yu, and X. Gan. “Effects of a SMART Goal Setting and 12-Week Core Strength Training Intervention on Physical Fitness and Exercise Attitudes in Adolescents: A Randomized Controlled Trial”. In: *International Journal of Environmental Research and Public Health* 19.13 (2022). DOI: 10.3390/ijerph19137715.

[49] A. Mariamo et al. “Emotional reactions and likelihood of response to questions designed for a mental health chatbot among adolescents: Experimental study”. In: *JMIR human factors* 8.1 (2021), e24343. DOI: 10.2196/24343.

[50] S. Michie, M. M. van Stralen, and R. West. “The behaviour change wheel: A new method for characterising and designing behaviour change interventions”. In: *Implementation Sci* 6.42 (2011). DOI: 10.1186/1748-5908-6-42.

[51] A. Middelweerd et al. “Apps to promote physical activity among adults: a review and content analysis”. In: *International Journal of Behavioral Nutrition and Physical Activity* 11.97 (2014). DOI: 10.1186/s12966-014-0097-9.

[52] H. L. Morris, K. E. Carlyle, and J. E. Lafata. “Adding the patient’s voice to our understanding of collaborative goal setting: How do patients with diabetes define collaborative goal setting?” In: *Chronic Illness* 12.4 (2016), pp. 261–271. DOI: 10.1177/1742395316648748.

[53] S. A. Munson and S. Consolvo. “Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity”. In: *6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops* (2012), pp. 25–32. DOI: 10.4108/icst.pervasivehealth.2012.248691.

[54] S. M. Neace et al. “Trait mindfulness and intrinsic exercise motivation uniquely contribute to exercise self-efficacy”. In: *Journal of American College Health* 70.1 (2022), pp. 13–17. DOI: 10.1080/07448481.2020.1748041.

[55] S. M. Nelis et al. “Goal-setting to Promote a Healthier Lifestyle in Later Life: Qualitative Evaluation of the AgeWell Trial”. In: *Clinical Gerontologist* 41.4 (2018), pp. 335–345. DOI: 10.1080/07317115.2017.1416509.

[56] C. Nigg. *Exercise: Stages of Change (Short Form)*. 2023. URL: https://web.uri.edu/cprc/measures/exercise/stages-of-change-short-form/ (visited on 08/24/2023).

[57] K. O’Leary et al. “Persuasive Reminders for Health Self-Management”. In: *AMIA Annu Symp Proc* (2016), pp. 994–1003.

[58] T. E. Oliphant. “A Bayesian perspective on estimating mean, variance, and standard-deviation from data”. In: *Faculty Publications* 278 (2006). URL: https://scholarsarchive.byu.edu/facpub/278.

[59] A. E. Paluch et al. “Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts”. In: *Lancet Public Health* 7.3 (2022), pp. 219–228. DOI: 10.1016/S2468-2667(21)00302-9.

[60] H. Park and N. Kim. “Predicting Factors of Physical Activity in Adolescents: A Systematic Review”. In: *Asian Nursing Research* 2.2 (2008), pp. 113–128. DOI: 10.1016/S1976-1317(08)60035-3.

[61] E. S. Pearson. “Goal setting as a health behavior change strategy in overweight and obese adults: A systematic literature review examining intervention components”. In: *Patient Education and Counseling* 87.1 (2012), pp. 32–42. DOI: 10.1016/j.pec.2011.07.018.

[62] D. J. Plonczynski. “Measurement of motivation for exercise”. In: *Health Education Research* 15.6 (2000), pp. 695–705. DOI: 10.1093/her/15.6.695.

[63] D. Pope and U. Simonsohn. “Round numbers as goals: Evidence from baseball, SAT takers, and the lab”. In: *Psychological science* 22.1 (2011), pp. 71–79. DOI: 10.1177/0956797610391098.

[64] J. O. Prochaska and C. C. DiClemente. “The transtheoretical approach”. In: *Handbook of psychotherapy integration* 2 (2005), pp. 147–171.

[65] M. Reichert et al. “Within-subject associations between mood dimensions and non-exercise activity: an ambulatory assessment approach using repeated real-time and objective data”. In: *Frontiers in Psychology* 7 (2016), p. 918. DOI: 10.3389/fpsyg.2016.00918.

[66] D. Resnick et al. “Designing a Cancer Prevention Collaborative Goal-Setting Mobile App for Non-Hispanic Black Primary Care Patients: An Iterative, Qualitative Patient-Led Process”. In: *JMIR Form Res* 6.3 (2022). DOI: 10.2196/28157.

[67] L. B. Robbins, N. J. Pender, and A. S. Kazanis. “Barriers to physical activityperceived by adolescent girls”. In: *Journal of Midwifery & Women’s Health* 48.3 (2003), pp. 206–212. DOI: 10.1016/S1526-9523(03)00054-0.

[68] J. A. Russell. “A circumplex model of affect.” In: *Journal of personality and social psychology* 39.6 (1980), p. 1161. DOI: 10.1037/h0077714.

[69] M. D. Schmidt et al. “Cardiometabolic Risk in Younger and Older Adults Across an Index of Ambulatory Activity”. In: *American Journal of Preventive Medicine* 37.4 (2009), pp. 278–284. DOI: 10.1016/S2468-2667(21)00302-9.

[70] V. Seefeldt, R. M. Malina, and M. A. Clark. “Factors Affecting Levels of Physical Activity in Adults”. In: *Sports Med* 32 (2002), pp. 143–168. DOI: 10.2165/00007256-200232030-00001.

[71] Enrico Sella et al. “Personality traits and sleep quality: The role of sleep-related beliefs”. In: *Personality and Individual Differences* 156 (2020), p. 109770. DOI: 10.1016/j.paid.2019.109770.

[72] Donghee Shin. “The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI”. In: *International Journal of Human-Computer Studies* 146 (2021), p. 102551. DOI: 10.1016/j.ijhcs.2020.102551.

[73] H. Y. Shum, X. D. He, and D. Li. “From Eliza to XiaoIce: challenges and opportunities with social chatbots”. In: *Frontiers of Information Technology & Electronic Engineering* 19 (2018), pp. 10–26. DOI: 10.1631/FITEE.1700826.

[74] S. Srewaradachpisal et al. “Effects of cushioning pads on heel and forefoot plantar pressure reduction in walking and running.” In: *Songklanakarin Journal of Science & Technology* 43.6 (2021). DOI: 10.14456/sjst-psu.2021.221.

[75] V. Strecher et al. “Goal Setting as a Strategy for Health Behavior Change”. In: *Health education quarterly* 22 (1995), pp. 190–200. DOI: 10.1177/109019819502200207.

[76] P. J. Teixeira et al. “A classification of motivation and behavior change techniques used in self-determination theory-based interventions in health contexts”. In: *Motivation Science* 6.4 (2020), pp. 438–455. DOI: 10.1037/mot0000172.

[77] R. H. Thaler and C. R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.

[78] J. Treasure. “Motivational interviewing”. In: *Advances in Psychiatric Treatment* 10.5 (2004), pp. 331–337. DOI: 10.1192/apt.10.5.331.

[79] S. G. Trost et al. “Correlates of adults’ participation in physical activity: Review and update”. In: *Medicine and Science in Sports and Exercise* (2003), pp. 1996–2001. DOI: 10.1249/01.MSS.0000038974.76900.92.

[80] C. Tudor-Locke et al. “How many steps/day are enough? For adults”. In: *International Journal of Behavioral Nutrition and Physical Activity* 8.1 (2011), pp. 1–17. DOI: 10.1186/1479-5868-8-79.

[81] S. T. Völkel et al. “User perceptions of extraversion in chatbots after repeated use”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–18. DOI: 10.1145/3491102.3502058.

[82] S. R. Wack, K. A. Crosland, and R. G. Miltenberger. “Using Goal Setting and Feedback to Increase Weekly Running Distance”. In: *Journal of Applied Behaviour Analysis* 47 (2014), pp. 181–185. DOI: 10.1002/jaba.108.

[83] A. 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* 14.1 (2012). DOI: 10.2196/jmir.1629.

[84] A. Wattanapisit and S. Thanamee. “Evidence behind 10,000 steps walking”. In: *Journal of Health Research* 31.3 (2017), pp. 241–248. DOI: 10.14456/jhr.2017.30.

[85] T. G. Weimann, H. Schlieter, and A. B. Brendel. “Virtual coaches: Background, theories, and future research directions”. In: *Business & Information Systems Engineering* 64.4 (2022), pp. 515–528. DOI: 10.1007/s12599-022-00757-9.

[86] D. K. White et al. “The association of obesity with walking independent of knee pain: the multicenter osteoarthritis study”. In: *Journal of obesity* (2012). DOI: 10.1155/2012/261974.

[87] WHO. *Physical activity*. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity (visited on 12/15/2022).

[88] D. Wilkinson et al. “Why or Why Not? The Effect of Justification Styles on Chatbot Recommendations”. In: *ACM Trans. Inf. Syst.* 39.4 (2021), pp. 1–21. DOI: 10.1145/3441715.

[89] K. E. Wilson and R. K. Dishman. “Personality and physical activity: A systematic review and meta-analysis”. In: *Personality and individual differences* 72 (2015), pp. 230–242. DOI: 10.1016/j.paid.2014.08.023.

[90] Wendy Wlasak, Sander Paul Zwanenburg, and Chris Paton. “Supporting Autonomous Motivation for Physical Activity With Chatbots During the COVID-19 Pandemic: Factorial Experiment”. In: *MIR formative research* 7 (2023). DOI: 10.2196/38500.

[91] Y. Xu et al. “Enhancing customer satisfaction with chatbots: the influence of anthropomorphic communication styles and anthropomorphised roles”. In: *Nankai Business Review International* ahead-of-print (2022). DOI: 10.1108/NBRI-06-2021-0041.

[92] M. Zhou et al. “Personalizing Mobile Fitness Apps using Reinforcement Learning”. In: *CEUR workshop proceedings* 2068 (2018).

A

Expert consultation scenarios

In this appendix, the scenarios used in the expert consultations with the psychology and physical activity experts are given. The aim of the consultations was to find out and understand why experts would choose certain designs regarding the goal-setting dialogue over others.

A.1. Psychology experts discussion

In this section, the scenarios used in the expert consultation with the psychology experts are given. The psychology experts consisted of a psychologist and assistant professor who is an expert on changing health behaviour, and a therapist and PhD student in health psychology

A.1.1. Scenario 1

In scenario 1, shown in Figure A.1, the user is asked about his mood, after which the virtual coach recommends a step goal. The user can then react in one of three ways:

- A: The user can choose one of the proposed step goals.
- B: The user is asked to give a goal himself.
- C: The user can choose to take the recommended goal or indicate that he wants a lower or higher goal.

Figure A.1: Scenario 1 for the discussion with psychology experts.

A.1.2. Scenario 2

In scenario 2, shown in Figure A.2, the user is asked about his mood, after which the virtual coach gives a couple of step goals to the user to choose from. These options can be given in one of two ways:

- A: The virtual coach explains what is recommended and why that is the recommendation. After that, it explains where the other options come from and asks the user to choose one of them.
- B: The virtual coach just gives the options and asks the user to choose one.

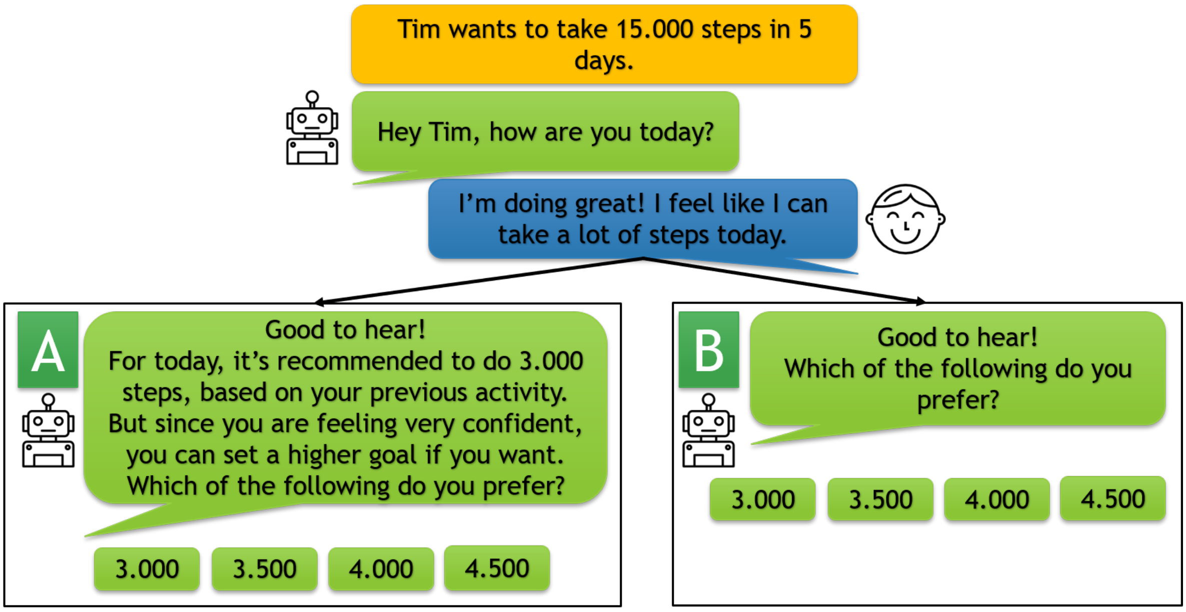


Figure A.2: Scenario 2 for the discussion with psychology experts.

A.1.3. Scenario 3

In scenario 3, shown in Figure A.3, the user indicates that he does not want to take any steps. The virtual coach can then react in one of two ways:

- A: The virtual coach does not try to negotiate and immediately accepts a goal of zero steps for that day.
- B: The virtual coach does not accept a lower goal and tries to push the user to take the recommended step goal as initially proposed.

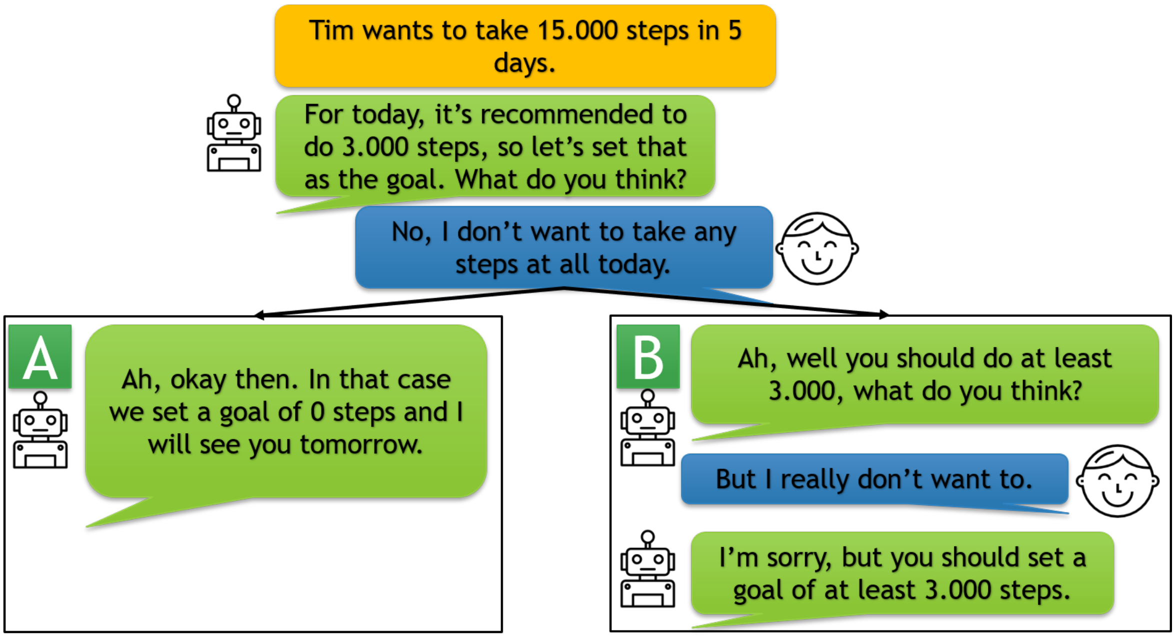


Figure A.3: Scenario 3 for the discussion with psychology experts.

A.1.4. Scenario 4

In scenario 4, shown in Figure A.4, virtual coach recommends a goal, after which the user indicates that he wants to set a lower goal. The virtual coach can then react in one of two ways:

- A: The virtual coach proposes a lower goal as long as the user wants to do less.
- B: The virtual coach tries to motivate the user by reminding them to keep up with the recommendation. After that, if the user still wants to set a lower goal, the virtual coach accepts it.

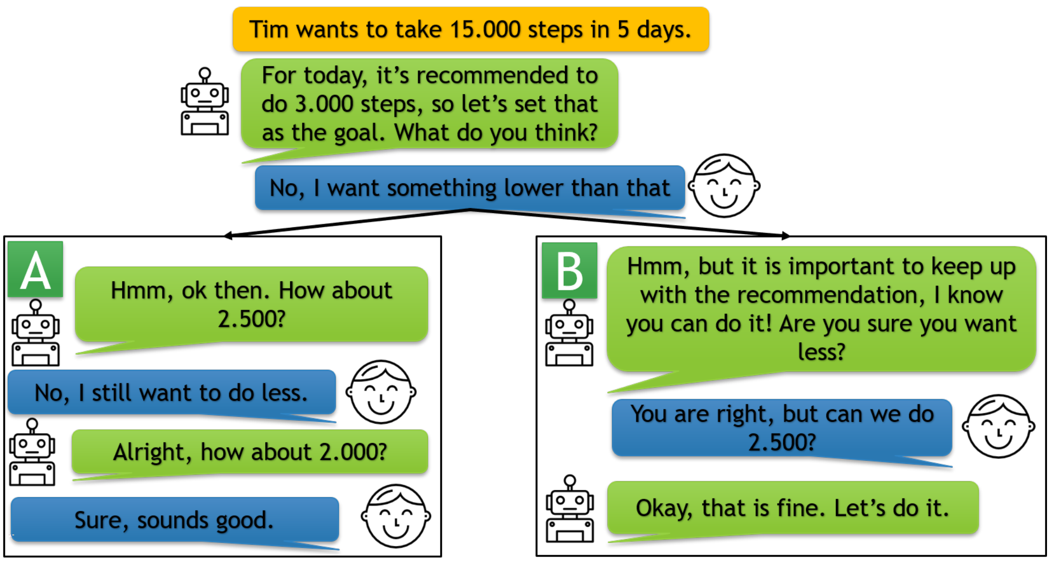


Figure A.4: Scenario 4 for the discussion with psychology experts.

A.2. Physical activity expert discussion

In this section, the scenarios used in the expert consultation with the physical activity expert are given. The physical activity expert is a PhD student in the field of biomedical signals and systems.

A.2.1. Scenario 1

In scenario 1, shown in Figure A.5, the user is asked about his mood, which the user reacts to by saying he is in a good mood. After that, the virtual coach can recommend a goal in one of two ways:

- A: The virtual coach indicates the recommended step goal but proposes a higher goal since the user is in a good mood.
- B: The virtual coach indicates the recommended step goal but asks the user to propose a goal which can be higher since the user is in a good mood.

Figure A.5: Scenario 1 for the discussion with sports expert.

A.2.2. Scenario 2

In scenario 2, shown in Figure A.6, the user is asked about his mood, which the user reacts to by saying he is in a bad mood. After that, the virtual coach can recommend a goal in one of two ways:

- A: The virtual coach indicates the recommended step goal but proposes a lower goal since the user is in a bad mood.
- B: The virtual coach indicates the recommended step goal but asks the user to propose a goal which can be lower since the user is in a bad mood.

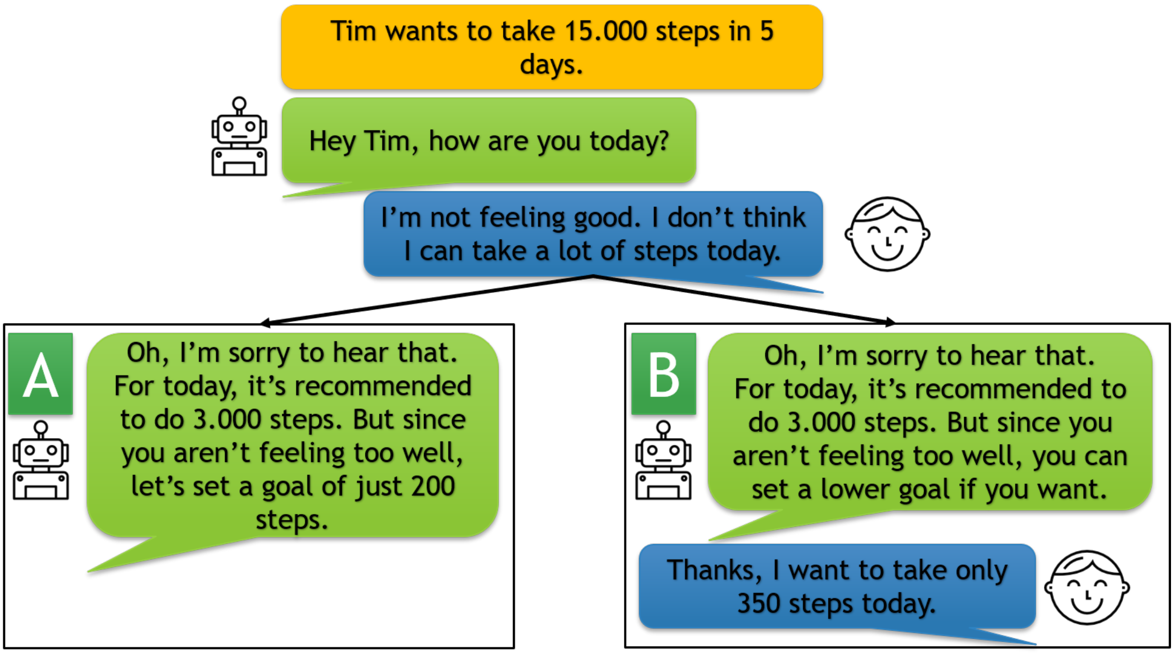


Figure A.6: Scenario 2 for the discussion with sports expert.

A.2.3. Scenario 3

In scenario 3, shown in Figure A.7, the user indicates that he does not want to take any steps. The virtual coach can then react in one of two ways:

- A: The virtual coach does not try to negotiate and immediately accepts a goal of zero steps for that day.
- B: The virtual coach lowers the proposed goal and tries to push the user to take that lowered step goal instead of setting a goal of zero steps.

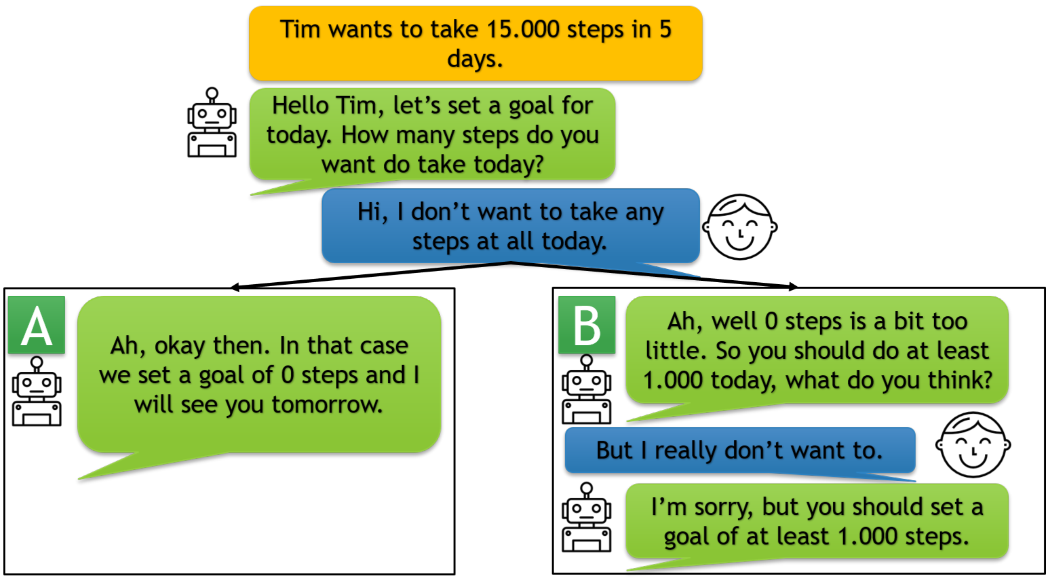


Figure A.7: Scenario 3 for the discussion with sports expert.

A.2.4. Scenario 4

In scenario 4, shown in Figure A.8, the user and the virtual coach set a long-term goal together. After that, the virtual coach can propose the step goal for the first day in one of three ways:

- A: The virtual coach proposes a relatively high goal but indicates that the user does not have to achieve it if it is not possible.
- B: The virtual coach proposes a relatively low goal but indicates that the user can do more steps that day if possible.
- C: The virtual coach proposes a goal based on the previous activity of the user.

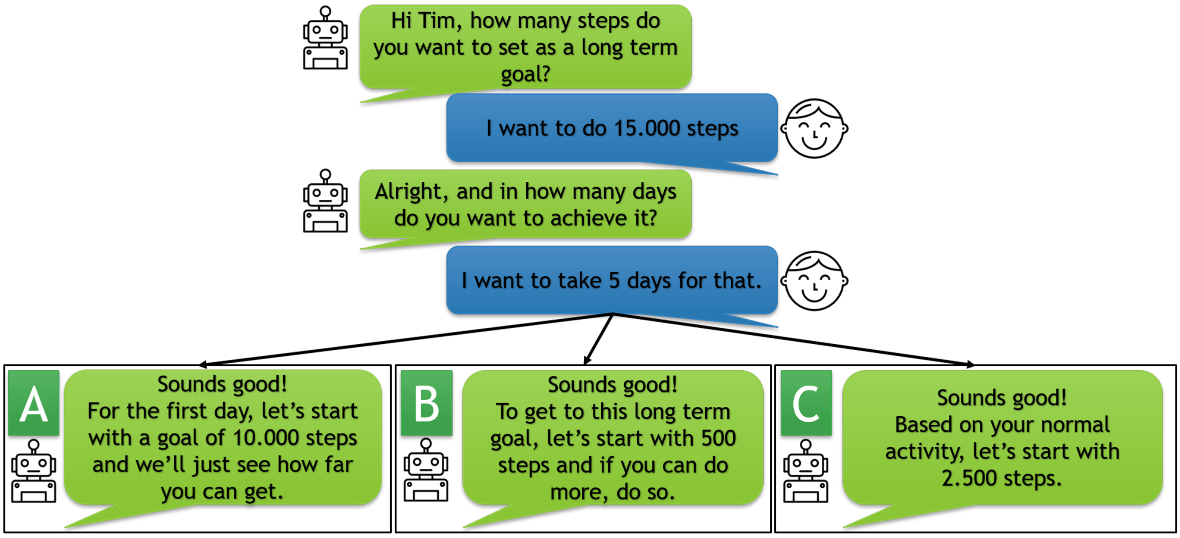


Figure A.8: Scenario 4 for the discussion with sports expert.

B

Measure details

The exact questions, scales and labels used for the primary and secondary measures in our observational study are provided in the following table. None of the questions were existing questions but were created according to the definitions of the measures as explained in Section 4.1.2.

Measure	**Question**	**Scale**
Self-efficacy	If you were given a step goal of your average number of steps from the last few days, how confident or not confident are you that you can achieve that step goal today on a scale from 0 to 10?	0 (I'm not confident at all that I can achieve that step goal today) to 10 (I'm very confident that I can achieve that step goal today)
Self-motivation	How motivated or demotivated are you to go for a walk today on a scale from 0 to 10?	0 (I am very demotivated to go for a walk today) to 10 (I am very motivated to go for a walk today)
Rest	How did you sleep last night on a scale from 0 to 10?	0 (I slept terribly last night) to 10 (I slept amazing last night)
Available time	How much time do you have available today to go for a walk on a scale from 0 to 10?	0 (I have no time at all to go for a walk today) to 10 (I have a lot of time to go for a walk today)
Valence dimension of mood	How would you describe your current mood?	Options from a study by Russell [68]
Steps	How many steps did you take yesterday?	No scale, but a number between 0 and 20.000
Perceived goal achievability	Knowing what you know now, how achievable or unachievable do you think a step goal of X steps would have been yesterday on a scale from 0 to 10?	0 (I think that goal would have been very unachievable) to 10 (I think that goal would have been very achievable)
Perceived goal difficulty	How easy or difficult did you find it to reach the daily step goals?	-5 (It was very difficult to reach the daily goals) to 5 (It was very easy to reach the daily goals) with 0 (It was not difficult but also not easy to reach the daily goals)

Table B.1: The questions, scales and labels used for each of the measures.

C
Mood to valence score mappings

To transform the mood data of people into a number to be used in the analysis, we mapped the moods according to their valence (horizontal axis position) in the 2-D Emotion Wheel of Kollias et al. [40]. We mapped the range [-1.0, -0.9] on the valence axis to a score of 0, the range [-0.9, -0.7] was mapped to 1, etc., and the range [0.9, 1] on the valence axis was mapped to a score of 10. This way, a mood with high valence got a high valence score as it indicated a positive mood. This led to the following mappings:

Mood	**Valence score**
Glad	10
Happy	10
Pleased	9
Delighted	9
Serene	9
Content	9
Satisfied	9
Relaxed	9
Calm	9
Excited	8
Astonished	7
Aroused	7
Sleepy	5
Neutral	5
Tired	5
Tense	5
Alarmed	5
Afraid	4
Droopy	3
Bored	3
Angry	3
Annoyed	3
Frustrated	2
Distressed	1
Depressed	1
Sad	1
Gloomy	1
Miserable	0

Table C.1: Mood to valence score mappings.

D
Samples details

After collecting the data and identifying the significant state features and their sizes, we ended up with the following number of samples per state-action pair:

	decrease	**slightly decrease**	**do nothing**	**slightly increase**	**increase**	total
s[0,0,0]	15	14	15	19	20	83
s[0,0,1]	8	6	7	7	5	33
s[0,1,0]	5	10	8	3	7	33
s[0,1,1]	6	9	9	4	4	32
s[1,0,0]	1	3	0	1	0	5
s[1,0,1]	3	6	4	7	6	26
s[1,1,0]	7	9	17	10	14	57
s[1,1,1]	29	23	17	22	21	112
total	74	80	77	73	77	381

Table D.1: The number of samples belonging to each state-action pair.

E

Distributions of rewards per state

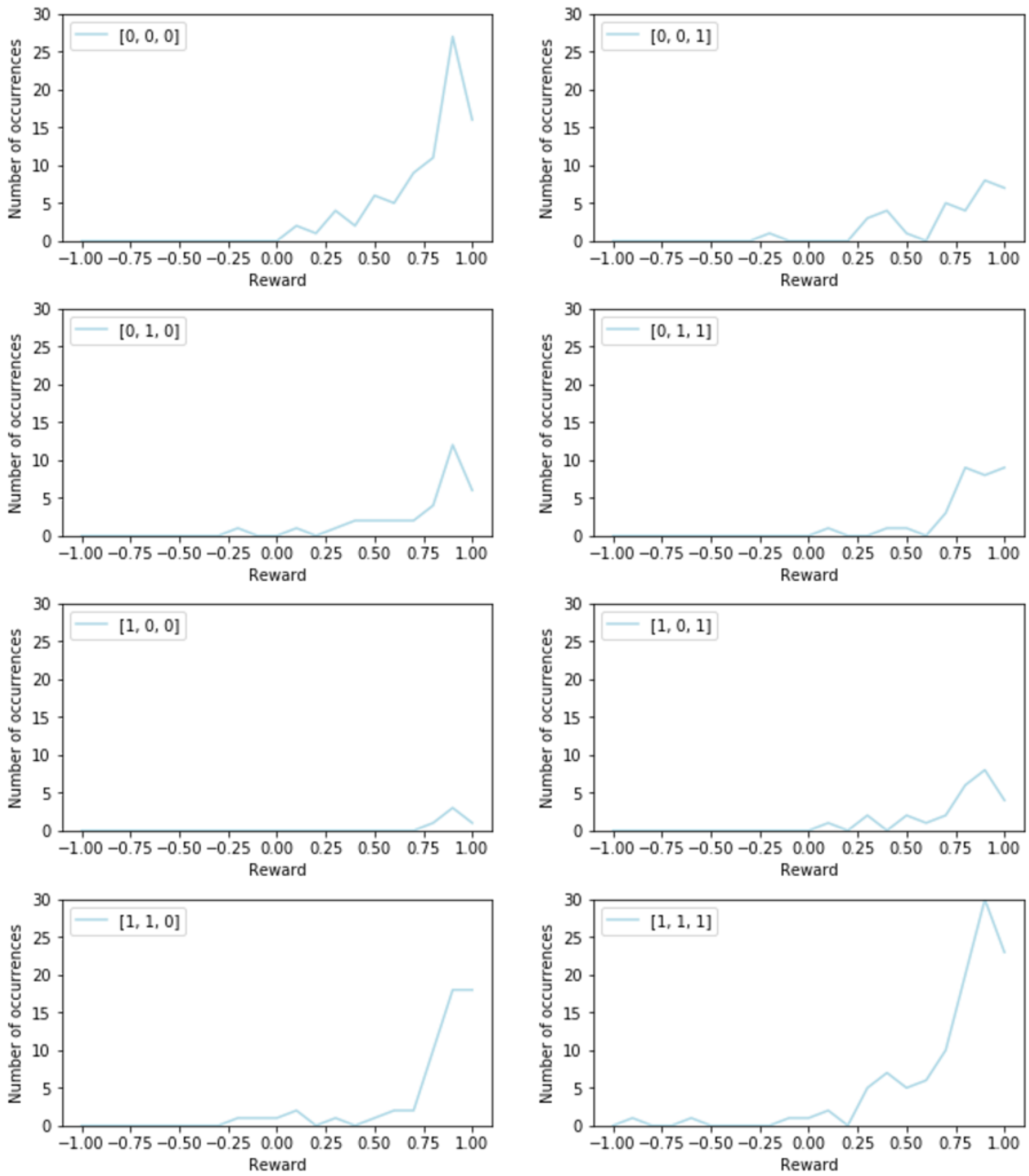


Figure E.1: The number of times each reward occurred for the samples per state.

F

Policies of the model

Calculating the q-values using the default reward function gave the following q-table:

State	decrease	slightly decrease	do nothing	slightly increase	increase
s[0, 0, 0]	5.33	**5.32**	5.33	**5.38**	5.33
s[0, 0, 1]	5.28	5.32	5.32	**5.27**	**5.36**
s[0, 1, 0]	5.34	**5.4**	**5.26**	5.36	5.32
s[0, 1, 1]	5.37	**5.42**	5.42	**5.33**	5.36
s[1, 0, 0]	**5.37**	5.37	**5.34**	5.37	5.34
s[1, 0, 1]	**5.23**	**5.42**	5.35	5.35	5.34
s[1, 1, 0]	5.35	**5.28**	5.36	**5.47**	5.43
s[1, 1, 1]	**5.24**	5.37	5.27	5.27	**5.37**

Table F.1: The Q-values for each state-action pair. Blue is the **optimal policy** and red is the **worst policy**.