

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/356809866>

Tailoring coaching conversations with virtual health coaches

Thesis · December 2021

DOI: 10.3990/1.9789036552608

CITATION

1

READS

52

1 author:



Tessa Beinema

University of Twente

14 PUBLICATIONS 41 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

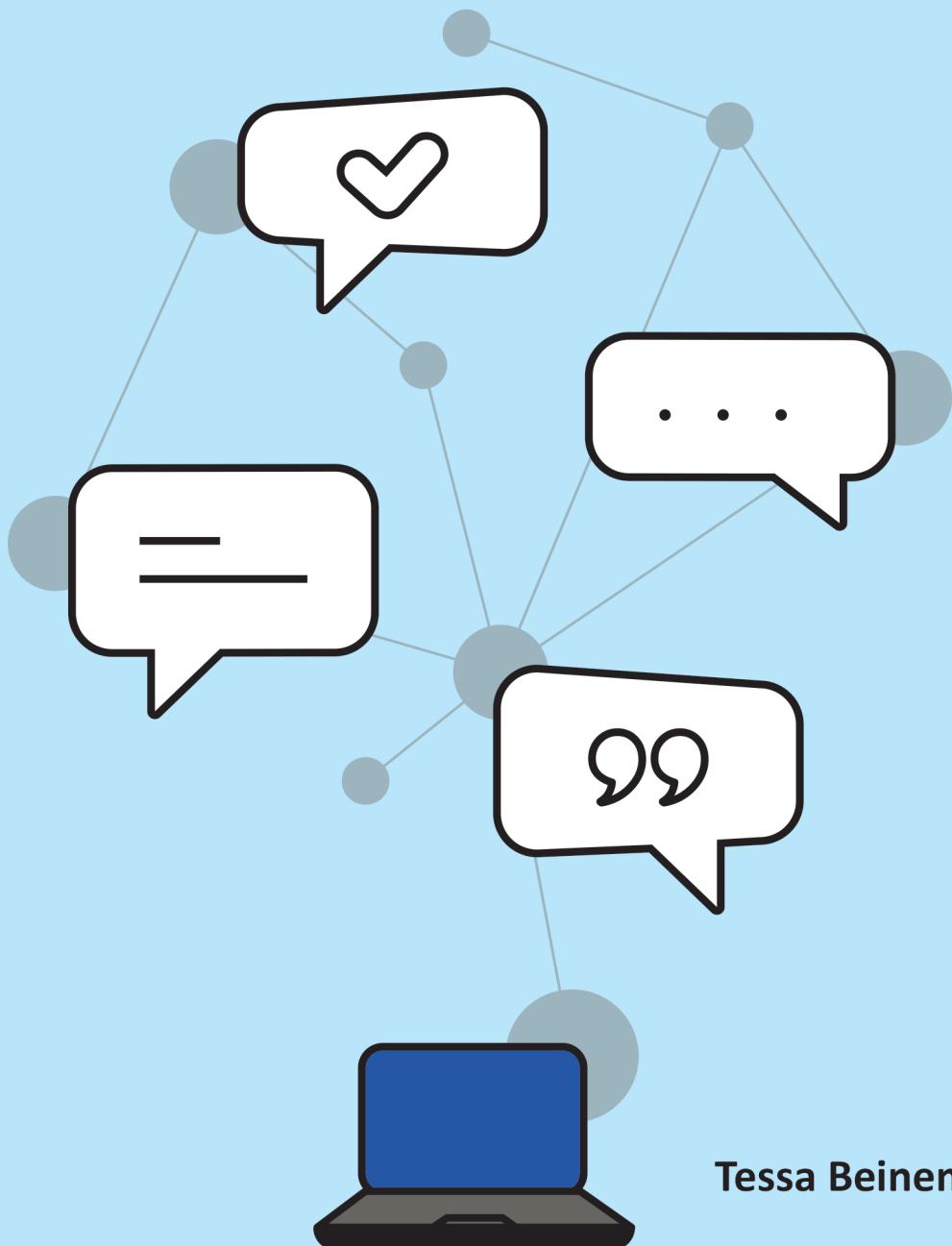


Council of Coaches [View project](#)



Agents United [View project](#)

Tailoring coaching conversations with virtual health coaches



Tessa Beinema

TAILORING COACHING CONVERSATIONS WITH VIRTUAL HEALTH COACHES

Tessa Beinema

The work in this thesis was financially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 769553 (Council of Coaches).

The publication of this thesis was financially supported by:



Cover design: Tessa Beinema

Printed by: Ipkamp Printing

Lay-out: Tessa Beinema

ISBN: 978-90-365-5260-8

DOI: 10.3990/1.9789036552608

© Tessa Beinema, Enschede, the Netherlands, 2021.

All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.

TAILORING COACHING CONVERSATIONS WITH VIRTUAL HEALTH COACHES

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on Thursday the 16th of December 2021 at 12.45 hours

by

Theresa Charlotte Beinema

born on the 23rd of June, 1991
in Heumen, the Netherlands

This dissertation has been approved by:

Supervisor:

prof. dr. ir. H.J. Hermens

Co-supervisor:

dr. ir. H. op den Akker

Graduation committee

Chair / secretary:

prof. dr. J.N. Kok, University of Twente

Supervisor:

prof. dr. ir. H.J. Hermens, University of Twente

Co-supervisor:

dr. ir. H. op den Akker, Innovation Sprint, Roessingh Research and Development

Committee Members:

prof. dr. J.E.W.C. van Gemert-Pijnen, University of Twente

prof. dr. D.K.J. Heylen, University of Twente

prof. dr. C. Pelachaud, Sorbonne University

prof. dr. D. Richards, Macquarie University

Contents

Chapter 1	General introduction	9
Chapter 2	The WOOL Dialogue Platform: Enabling interdisciplinary user-friendly development of dialogue for conversational agents	21
Chapter 3	Tailoring coaching strategies to users' motivation in a multi-agent health coaching application	47
Chapter 4	What to discuss? – A blueprint topic model for health coaching dialogues with conversational agents	79
Chapter 5	Automatic topic selection for embodied conversational agents in health coaching: A micro-randomized trial	117
Chapter 6	General discussion	145
&	Appendices References Summary Samenvatting Dankwoord About the author Progress range	165

Chapter 1

General introduction

Health and coaching

Our **health** is important. It is a resource for everyday life, which lets us interact with our loved ones, go to school or work, and even go on adventures. In 1948, the World Health Organization defined health as “*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*” (World Health Organization, 1946). While this definition has not changed since, it is argued that it should be updated to fit a society in which life expectancy has increased and the prevalent types of illnesses have changed. The proposed new concept of health has a focus on people’s resilience and ability to cope with new situations: “*Health as the ability to adapt and to self-manage, in the face of social, physical and emotional challenges.*” (Huber et al., 2011; Huber et al., 2016). The ‘right to health’ is recognised by the United Nations and many other international organisations as one of the fundamental rights of every human being (Office of the United Nations High Commissioner for Human Rights & World Health Organization, 2008).

While recent advances in medicine have led us to live longer, these additional years are not always spent in good health. The number of persons aged 60 or above is expected to more than double by 2050, and even triple by 2100 (United Nations Department of Public Information, 2017). In Europe specifically, the percentage of the population that is older than 60 is expected to increase from 25% of the population to 35%. Where in 1948 infectious diseases were a leading cause of death, nowadays, an increasing percentage of the population suffers from noncommunicable diseases (NCDs) or chronic diseases. These include cardiovascular diseases, cancers, chronic respiratory diseases and diabetes (World Health Organisation, 2021), but dementia and depression are also common. Chronic diseases are not only responsible for 71% of all deaths (globally) each year (41 million people), but people also suffer from them for a long time. This not only influences their quality of life, but also puts pressure on our health care system.

An effective approach for the prevention and control of chronic diseases is to reduce risk factors in multiple domains. Adopting a healthier lifestyle by, for example, following a healthy diet, exercise habits and sleep habits, can delay chronic diseases and improve quality of life. Research has shown that health coaching by health care professionals improves the management of chronic diseases (Kivelä et al., 2014), although the most suitable approach can vary between target groups. Coaching “*can be considered as a series of conversations between two individuals – the coach and coachee – for the benefit of the coachee in a way that relates to the coachee’s learning process*” (Starr, 2008).

However, it has been shown that behaviour change, especially in the long-term, is difficult (Bouton, 2014), even if supported through frequent sessions with a

coach. Furthermore, resources in our healthcare system are finite and can be expensive, and human coaches do not offer continuous support (for example, when there is a moment of weakness at the end of a long day at work). Ideally, people would be continuously supported in changing their behaviour towards a healthy lifestyle so that chronic diseases can be prevented or their effects limited.

1

eHealth applications for continuous support

One solution for the limited availability of human coaches can be found in digital health applications for personalised coaching. These applications are referred to by a number of names, which include digital health (eHealth) applications, digital behaviour change interventions (DBCIs), health coaching applications, health behaviour change applications, or e-coaching systems. They are generally developed to assist a user in adopting a healthier lifestyle by adjusting modifiable behaviours, such as being physically active or making dietary choices (World Health Organisation, 2021). In recent years, research and development in the area of such eHealth applications has increased (Brinkman, 2016).

Where initially there was a focus on providing health information via websites or telemedicine services, in later years this focus shifted to building interventions that could also actively send notifications, use information from wearables and other sensors, and provide the user with insight into their personal situation (e.g., through personalised feedback). These systems have the potential to inform, assist and empower, and thus support, people in adopting a healthier lifestyle. The fact that they could be always accessible, scalable and low-cost make them an interesting candidate for providing continuous support and coaching. In this manner, they can be employed to treat health conditions, but also prevent them.

While eHealth applications have the potential to provide continuous support, a major challenge for these applications is users' adherence (Nijland, 2011; Wangberg et al., 2008). Users tend to interact with an application at first, but then stop using it or stop following its suggestions after the novelty effect wears off. This is problematic, since health applications need to capture their users attention (i.e., be *engaging*) for a longer period of time to impact users' behaviour change (Perski et al., 2017; Yardley et al., 2016). Engagement is therefore a prerequisite for the system to have further impact (Bickmore et al., 2010).

In general for interaction with applications, short-term engagement tends to be characterised as flow (Hamari et al., 2016; Nakamura & Csikszentmihalyi, 2002) or enjoyment (O'Brien & Toms, 2013), while long-term engagement can be seen as the duration and depth of usage of a system over time (Couper et al., 2010). There are several objective measures for long-term engagement, such

as the number of voluntary interactions users choose to have, the number of logins, and amount and type of content used (Perski et al., 2017; Trinh et al., 2018). The relevance of engagement for eHealth applications is also addressed by Cole-Lewis et al. (2019), who distinguish two types of engagement, 'Big E' (health behaviour engagement) and 'Little e' (application engagement), with a subdivision for the latter in user interaction with a) features that encourage frequency of use and b) behaviour change intervention components. They also emphasise that 'Big E' is dependent on 'Little e'. That is, once again, if a user is not engaged with the application, they will not be engaged with the suggested health behaviour. Yardley et al. (2016) also distinguish two similar types of engagement, but call these micro (moment-to-moment) and macro (with the intervention goals) engagement.

The causes for the lack of adherence to and engagement with digital health applications are actively being researched. It appears that contributing factors are a lack of direct involvement of a health care professional (*no social incentive*) and content that does not always fit the user's personal situation (*relevance of content*) (Andersson et al., 2009; Buimer et al., 2017). Therefore, two possible solutions investigated to tackle these issues are the inclusion of *conversational agents* as a natural human-computer interaction paradigm to stimulate social incentive, and *tailoring* the application to the user (op den Akker et al., 2014) to increase its relevance.

Conversational agents as coaches

Conversational agents (CAs) have been widely researched and are already actively employed in many domains. They are "*computer systems that imitate natural conversation with human users through images and written or spoken language*" (Laranjo et al., 2018). These agents are present in various commercial settings, where they are digital assistants, such as Siri, Alexa, or Cortana. They are also prevalent in the customer service field (Kuligowska, 2015; Xu et al., 2017), and in the digital entertainment industry, for example, as non-player characters in video games (Grey & Bryson, 2011). In research, they are incorporated in applications to make those applications more accessible, engaging and effective; for example, as pedagogical agents for children (Kim & Baylor, 2016; Veletsianos & Russell, 2014), providing various types of skills training (Tanaka et al., 2017), and for a wide range of health related applications such as coaches (Kramer et al., 2020; van Wissen et al., 2016) or digital healthcare professionals (Zhou et al., 2014). In some cases, conversational agents are just a component of a broader eHealth application (ter Stal et al., 2020; van Velsen et al., 2020), while in other cases they are the intervention's main component (op den Akker et al., 2018; Sebastian & Richards, 2017). In all cases, agents are designed to

interact with the user and in this manner hopefully increase engagement with the application itself and ultimately with the application's aim. Figure 1.1 shows some examples of conversational agents for health with varying embodiments.

The inclusion of CAs in eHealth applications serves multiple purposes. One is that CAs can make the use of a health application easier, more satisfying, and less frustrating (André & Pelachaud, 2010; Bickmore et al., 2016), and potentially more effective (Ma et al., 2019). Second, their potential to engage the user can help improve user's adherence and participation (Bickmore et al., 2010; Scholten et al., 2017) and can potentially increase an application's effect in changing the user's behaviour (Ma et al., 2019). CAs can give the system social ability, which is an important factor for establishing and maintaining a collaborative relationship in such applications (Bickmore et al., 2010; Bickmore, Trinh et al., 2018; Kamphorst, 2017). Third, a CA's credibility, similarity, authority, power, and social attractiveness can all contribute to the persuasion of a user (Pickard, 2012, pp. 73-74), which can be useful when providing behaviour change support. Furthermore, they are always available and their dialogues can be tailored dynamically to the user, for example by elaborating when needed (Bickmore & Giorgino, 2006). They never grow tired of answering (the same) questions and can provide continuous support. These aspects make that CAs could support users in their daily life - after a diagnosis, but also in preventative care.

While some agent applications are designed for relatively short single session interactions (e.g., ordering a bookcase at Ikea or changing your flight), others aim to have the user interact with an agent in several sessions that are spread out over a longer period of time. This long-term interaction is especially relevant for CAs that are incorporated in eHealth applications (Bickmore, Trinh et al., 2018; Callejas & Griol, 2021). It is therefore important that the dialogue development process of such virtual coaches takes this into account.

Coaching dialogues

A conversational agent's capability to communicate with a user, through natural language, is important in health coaching applications, as it provides an intuitive human-computer interaction (HCI) metaphor. Such *interactivity* is generally acknowledged to be a natural attribute of human face-to-face communications, but is also an attribute that is fundamental in HCI concepts such as the computers as social actors (CASA) paradigm (Nass et al., 1994), which states that people apply social rules and expectations to computers. Once the user is engaged however, they need to stay engaged and concepts like *goal agreement*, *task agreement* and *development of a personal bond* become important (Castonguay et al., 2006; Hatcher & Gillaspy, 2006; Horvath & Greenberg, 1989). Writing

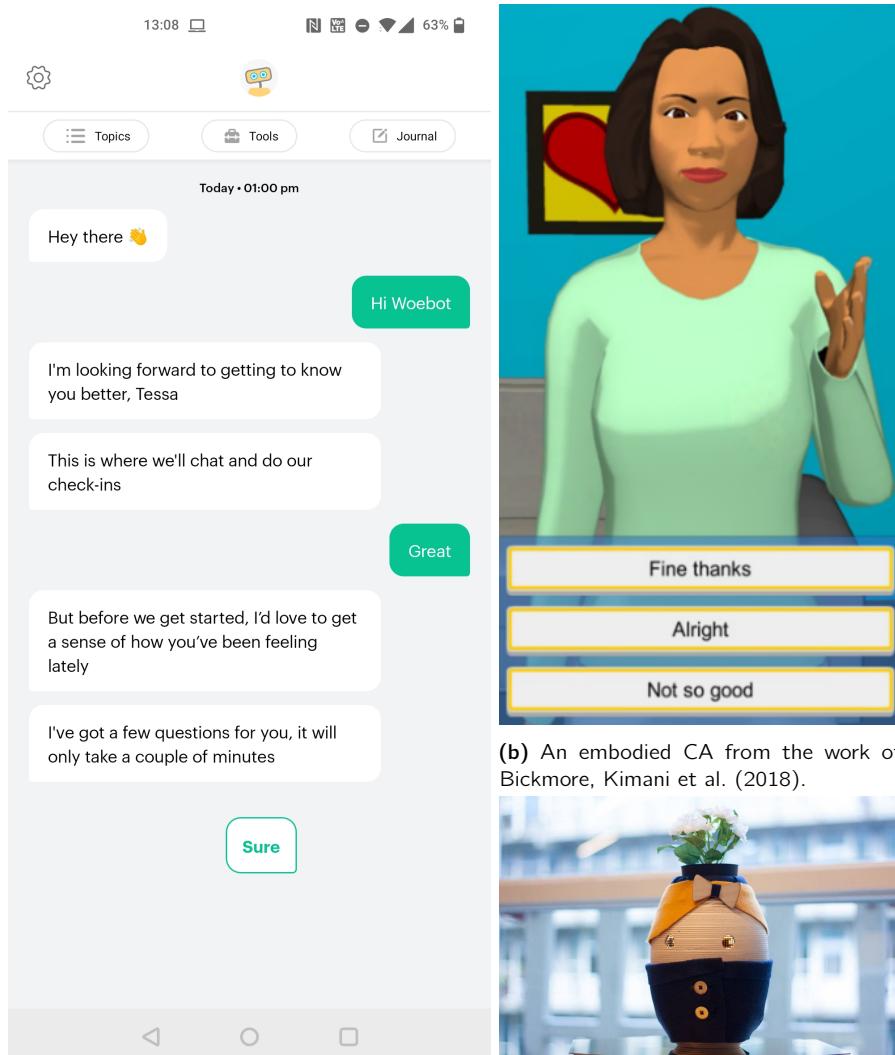


Figure 1.1: Examples of different types of conversational agents for health.

dialogues for the agent that follow a carefully designed strategy can fulfil these needs, but not all users have the same preferences or respond well to the same approach.

Dialogue content in eHealth applications, when not automatically generated, is carefully designed by domain experts based on existing interventions (e.g., by including motivational interviewing principles (Lisetti et al., 2011; Olafsson et al., 2019)). This process begins with translation of intervention content (e.g., guidelines, exercises and conversational techniques) to interactive dialogues that can be executed between a user and the conversational agent. In addition, conversational agents also need to motivate the user to complete the objective of their interaction (Bickmore et al., 2010). One way to do so is by giving the agent a personality and to not only include *task related* (coaching) dialogues, but to also include *social* elements (Bickmore et al., 2005; Schulman & Bickmore, 2009). Furthermore, a virtual coach should ask questions, give feedback and offer tailored advice (Kamphorst, 2017). In addition, health coaching dialogues with agents should be designed to take place over several sessions (Bickmore, Kimani et al., 2018). This means that over time, dialogues should be coherent and that there should be variation between dialogues in terms of content and structure to keep participants engaged (Bickmore et al., 2013).

Tailoring

Tailoring an application and its content to its users has proven to be an effective approach in both conversational agents and eHealth applications in general. For example, for embodied conversational agents the non-verbal behaviour was adjusted (Krämer et al., 2010) and switching between topics on a sentence level was investigated (Glas & Pelachaud, 2018; Smith et al., 2011), and in several eHealth applications adjustments were made to the content of SMS messages or suggested goals (Krebs et al., 2010; Ryan et al., 2019; Wangberg et al., 2008). Tailoring of content can be seen as the adjustment of a communication's timing, intention, content and representation to a user (op den Akker et al., 2014). In eHealth applications, communication can be tailored to many aspects of a user's personal profile. Examples from the context of sending messages include tailoring to Big Five personality traits (de Vries et al., 2016), self-efficacy and measured level of physical activity (Achterkamp et al., 2013), susceptibility to persuasive strategies (Kaptein et al., 2012), stage of change (de Vries et al., 2016; Uribe et al., 2011), Combi model state (Klein et al., 2013), or gender (de Vries et al., 2017).

Tailoring is investigated to make a digital health application's content, such as the interaction with a conversational agent, more personally relevant for the specific user (e.g., Krebs et al. (2010), Ryan et al. (2019) and Wangberg et al.

(2008)). Most research on tailoring coaching content either has its focus on communication in a single direction (e.g., sending messages and notifications) or on the adjustment of the final presentation steps (e.g., inserting tailored goals or adjusting wording of sentences). When it comes to coaching dialogues with CAs participants are often all presented with the same general coaching programme or have to select what they want to discuss themselves. In some cases, the coaching programme is tailored manually by an involved healthcare professional in the system's back end (e.g., Abdullah et al. (2018), Benítez-Guijarro et al. (2019) and Fadhil et al. (2019)). However, all of these approaches have their disadvantages. That is, the first approach presents all participants with the same general approach, the second places the initiative with the coachee (while the coach should be credible and proactive (Kamphorst, 2017)), and the third approach – which seems most optimal from a tailoring and initiative perspective – requires involvement of healthcare professionals.

Ideally, a coaching system would take available information about the user, their behaviour, their context, and the coaching domain and it would adapt its communication to that knowledge. In the case of coaching conversations with virtual agents, this would mean that the dialogues with the agent are tailored to such available information and knowledge. The application would dynamically respond to the behaviour and context of the user (e.g., through multi-modal sensor input), would be able to adjust to the type of user (e.g., through profiling), their responses in the dialogues, and their preferences and values. However, where this process is quite intuitive for a human coach, all of these processes have to be modelled and implemented when it comes to virtual coaches.

Research context: Council of Coaches

The research context for this thesis was the European Horizon 2020 Council of Coaches project (op den Akker et al., 2018). Most coaching applications with conversational agents feature a single agent as a coach who has a specific expertise, such as physical activity (King et al., 2017; Watson et al., 2012). However, as previously stated, health often requires support on multiple domains (Huber et al., 2016; World Health Organization, 1946). To that end, recent research has been investigating coaching with a single agent on a combination of domains (Gardiner et al., 2017; Klaassen et al., 2018). The Council of Coaches project explored the concept of virtual coaching with multiple embodied conversational agents. It aimed to develop a state-of-the-art multi-agent health coaching application that combined holistic behaviour analysis, smart adaptive coaching, dialogue management, and realistic embodied conversational agents to aid its users in obtaining a healthier lifestyle.

Interaction with multiple agents at the same time provides opportunities for

vicarious persuasion and engagement (André & Rist, 2001). That is, coaches can explain concepts to each other, which can indirectly persuade the user (Kantharaju et al., 2018) and this might also keep the user involved in the conversation, even if they do not want to actively provide answers at a certain moment. It also allows for a decentralised presentation of domain information, for example by casting the agents as coaches that each have their own expertise (e.g., a physical therapist or dietitian), thus providing the possibility to include multiple viewpoints without an agent contradicting itself (Kantharaju et al., 2019).



(a) The 2D interface for the functional demonstrator.



(b) The 3D interface for the technical demonstrator.

Figure 1.2: The two final applications in the Council of Coaches project.

CHAPTER 1

The work presented in this thesis was situated in the work package that focused on the shared knowledge base and coaching strategies. The related software component was situated at the centre of the system and was included in both a functional demonstrator (to test the concept and usability of the application) and a technical demonstrator (to test the application and development of state-of-the-art 3D agents) (see Figures 1.2a and 1.2b).

This thesis

The effectiveness of a health coaching application with conversational agents depends on the use of those applications. Tailoring coaching content has been found to be an effective approach for increasing engagement with eHealth applications and also seems to be a promising approach for presenting users with relevant and engaging coaching conversations. Previous research has shown that tailoring principles such as suggesting personalised goals and personalisation of content (e.g., by addressing users with their first name) are effective approaches from eHealth tailoring that can be transferred to the dialogue context. However, much is still unknown when it comes to tailoring the dynamic two-way interaction that is involved with coaching dialogues. Therefore, the aim of this thesis is to investigate how we can further tailor coaching dialogues with conversational agents to increase the engagement with these agents. The specific research question that is addressed in this thesis is the following:

How can we tailor users' coaching conversations with conversational agents to improve engagement?

To address this research question, this thesis focuses on tailoring and approaches this with three essential aspects of the tailoring process in mind: 1) content and system design; 2) tailoring content and conversations; and 3) evaluation of tailored content. An overview of the chapters is shown in Figure 1.3.

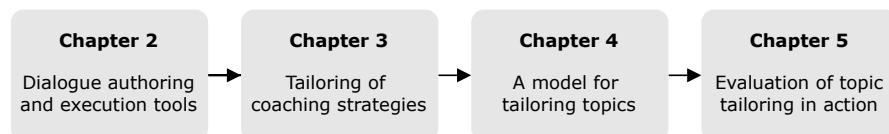


Figure 1.3: A schematic representation of the chapters in this thesis.

The development of tailored coaching conversations does not begin with tailoring, but with the content of the coaching dialogues. Providing correct advice is

essential when it comes to health coaching, and it is therefore important that (non-technical) domain experts can participate in dialogue authoring and testing. **Chapter 2** discusses the development of specialised tools that facilitate dialogue authoring and development as part of such multidisciplinary collaboration. In addition, the dialogue language and execution libraries had to include features that facilitate the tailoring process, both for conventional tailoring approaches (e.g., including a first name or personal goal in conversations) as well as allowing for more dynamic approaches on a higher level (e.g., seamlessly chaining together several dialogues to build up a full conversation).

Following the development of tools to facilitate authoring and tailoring of coaching dialogues for conversational agents, the focus shifts to tailoring this dialogue content. **Chapter 3** describes a study that explores if the underlying strategy for coaching conversations can be tailored to users' motivation to live healthy. Such strategies are important to ensure that the conversations with a coach are coherent over several sessions. In addition, it investigates if the likeability of a conversational agent influences a users' appreciation of the strategy it presents and if users' demographics are of influence on strategy appreciation.

Where **Chapter 3** focused on the strategy that is underlying to a coaching approach, **Chapter 4** focuses on tailoring the conversation itself during interaction with the user. It begins with the introduction of the concept that coaching conversations can be tailored in five steps, one of which is tailoring topics of conversation. In order to implement such automatic topic selection, a structured set of topics is needed. The chapter therefore investigates which topics should be included in a blueprint topic model for health coaching conversations with CAs and how these should be structured. The model is constructed using literature from the relational agents, persuasive technology, and behavioural psychology fields and evaluated in a card-sort study with experts.

With the introduction of the automatic topic selection concept in **Chapter 4**, **Chapter 5** presents the evaluation of a proof of concept implementation of such automatic topic selection. It describes the setup and results of a micro-randomized trial that was conducted as part of a long-term daily life study. In this trial, we compare the condition in which an ECA automatically suggests a topic to discuss with the condition in which the user has to select a topic themselves. We evaluate the length of interactions in both conditions (as a measure of engagement), and the acceptance of suggestions for automatically selected topics by participants.

Finally, in **Chapter 6** the main findings of the previous chapters are summarised and discussed. This thesis concludes with a discussion of implications, challenges and opportunities for the development of personally relevant conversational coaches.

Chapter 2

The WOOL Dialogue Platform: Enabling interdisciplinary user-friendly development of dialogue for conversational agents

Based on:

Beinema, T., op den Akker, H., Hofs, D., van Schooten, B. The WOOL Dialogue Platform: Enabling interdisciplinary user-friendly development of dialogue for conversational agents (*submitted for publication*).

Abstract

Health coaching applications can include (embodied) conversational agents as coaches. The development of these agents requires an interdisciplinary cooperation between eHealth application developers, interaction designers and domain experts. Therefore, proper dialogue authoring tools and tools to integrate these dialogues in a conversational agent system are essential in the process of creating successful agent-based applications. However, we found no existing open source, easy-to-use authoring tools that support multidisciplinary agent development. To that end, we developed the WOOL Dialogue Platform.

The WOOL Dialogue Platform provides the eHealth and conversational agent communities with an open source platform, consisting of a set of easy to use tools that facilitate virtual agent development. The platform consists of a dialogue definition language, an editor, application development libraries and a web service. To illustrate the platform's possibilities and use in practice, we describe two use cases from EU Horizon 2020 research projects.

The WOOL Dialogue Platform is an 'easy to use, and powerful if needed' platform for the development of conversational agent applications that is seeing a slow but steady increase in uptake in the eHealth community. Developed to support dialogue authoring for embodied conversational agents in the health coaching domain, this platform's strong points are its ease of use and ability to let domain experts and agent technology experts work together by providing all parties with tools that support their work effectively.

2.1 Introduction

The world's population is ageing, and the pressure on healthcare is increasing. Digital health (eHealth) applications are investigated to not only support people with lifestyle changes following a health condition, but also as supporting tools for preventative lifestyle changes. These applications, however, face adherence challenges (Nijland, 2011; Wangberg et al., 2008), and people need to actively engage with them for the interventions to have an effect. To increase the engagement in eHealth applications, the inclusion of (embodied) conversational agents as a natural human-computer interaction paradigm is a promising field of investigation. These agents can take on the role of virtual health experts or coaches, and allow users to interact with the application in a natural manner, namely through conversations (dialogues). The agents can be one element in a broader application (L. van Velsen et al., 2020), or they can be the application's main component (op den Akker et al., 2018; Sebastian & Richards, 2017).

Conversational agents have been widely researched and are employed in many domains. They are present in various commercial settings, such as customer service (Kuligowska, 2015; Xu et al., 2017) or the digital entertainment industry, such as non-player characters in video games (Grey & Bryson, 2011). In research, they are incorporated in applications to make those applications more accessible, engaging and effective; for example, as pedagogical agents for children (Kim & Baylor, 2016; Veletsianos & Russell, 2014), providing various types of skills training (Tanaka et al., 2017), and for a wide range of health related applications (op den Akker et al., 2018; van Wissen et al., 2016; Zhou et al., 2014).

Building conversational agents is an interdisciplinary process that requires domain experts, experts on agent technology, and agent designers to work together (Wang et al., 2019; Watson et al., 2012) – especially when building agents for contexts where presenting incorrect domain information might have serious adverse effects, such as the healthcare domain or other mission critical domains. The tools that are used to author the dialogues and the software libraries used for dialogue management in these conversational agents are therefore an important factor in making the development process as efficient and user-friendly as possible – both for technical and non-technical members of a development team (Wang et al., 2019).

While tools and libraries are important in the process of developing conversational agents with high quality content and many human-computer interaction labs have their own private development tools, we found that there was no open source dialogue authoring platform that supported this process, was user-friendly and could be used for developing web and mobile applications on an operating system of choice. In this paper, we therefore present the WOOL Dialogue Platform (Roessingh Research and Development, 2020) for dialogue authoring

and execution (see Figure 2.1), which has been developed over the past three years in the context of the EU Horizon 2020 project Council of Coaches (grant agreement #769553).



Figure 2.1: The WOOL Dialogue Platform logo.

The aim of the Council of Coaches project was to develop a health coaching application in which multiple embodied conversational agents support a user with advice on leading a healthy lifestyle (op den Akker et al., 2018). The WOOL Dialogue Platform was developed to fulfil a need for tools that support the division of tasks in the dialogue development process between domain experts and system developers, that allow novice authors (who were experts on the coaching domains) to write dialogues, and that support functional testing while also allowing for the developed content to be used for a state-of-the-art technical prototype.

2.2 The dialogue authoring process

One possible representation for dialogues are dialogue scripts. These scripts can serve multiple uses in agent development: they can be used directly, as design examples, or as training data for more complex dialogue systems. Often applied in applications with a user interface that follows a speech-bubble and reply-button paradigm, scripted dialogues have a simple learning curve and a low entry threshold for first-time authors; especially when supported by intuitive authoring tools. This ensures that domain experts can focus on the quality of the dialogues instead of struggling to write them. Furthermore, scripted dialogues have a set structure and order, which makes dialogues relatively predictable – an important feature if you develop an eHealth application that provides health coaching and you need to ensure that the resulting advice by the coach is always sound.

Authoring scripted dialogues can be an intensive process. A few phases can be distinguished in this process and each of them has its challenges. First, once the general idea for the application is clear (e.g., the coaching domain has been selected and the role of the agent defined), there is a phase in which system

designer and domain experts determine the topics for which dialogues should be written. For example, for a physical activity coach this could involve determining the type of feedback that should be given or which health education sessions can be defined. Some of these topics can be repeated over coaching sessions (e.g., feedback on sensor data), while others are less likely to be repeated (e.g., a first introduction). Therefore, one major challenge in dialogue authoring is to make dialogue scripts *reusable*, as opposed to duplicating (parts of) dialogues. A possible solution could be to divide content into multiple dialogue scripts that refer to each other, instead of creating one large script.

In the second phase, a domain expert writes the dialogues. This requires them to keep an overview of the network of possible statements by the agent and replies by the user. In a dialogue script, each defined statement could have multiple reply options for a user, which in turn might all receive different responses by the agent, etc. This *branching* is another major challenge in dialogue authoring. Keeping an overview of the network (a directed dialogue graph) can be challenging for authors and could be assisted by dialogue visualisation (Gaffney et al., 2008). Furthermore, authoring dialogues tends to be an iterative process and in this phase, it could be beneficial for the author to go back and forth between editing and interacting with the dialogue they are writing (Gaffney et al., 2008).

In the authoring phase it is also important to include elements to tailor the dialogue to the user. Examples are using the user's first name in a sentence or responding to a user's personal data. This can be done during the authoring process by the expert or afterwards by a developer. *Tailoring* dialogues is a challenge that can be addressed by allowing authors to use variables (e.g., `$userFirstName`) and conditionals (e.g., if-then-else statements) in a dialogue script, that can then be filled in or executed with information stored for a specific user. Storing information following user replies or about which parts of a dialogue have been completed is therefore essential.

Once the domain expert has finished a dialogue, in the third phase, the dialogue is reviewed and tested by the system developers. They will add variables for dialogue management – and possibly other purposes, depending on the authoring experience of the domain expert – and they will embed the dialogue in the intended system. They will also test the dialogue in the actual system to ensure that the flow fits the agent and context. After the dialogue has been successfully integrated in the application, the domain expert is then asked to review it again.

Finally, completed dialogues can be translated to other languages. At the minimum, this involves management of multiple versions of a dialogue's text (*translation management*), while keeping the structure the same. Translating sentences to another language, while avoiding, for example, anglicisms (the use of a word or construction from English in another language) is an art. However, localisation in dialogue can be more complicated than that, and sentence-by-

sentence translation might not always suffice. For example, dialogue branches in which an agent recommends local events, or discusses a typical local recipe, may not be suitable to “translate”, but must be replaced with completely different localised dialogue branches. Furthermore, there are numerous other translation challenges, for example, formality can differ between languages (e.g., ‘your’ in English can already be translated to either ‘jouw’ or ‘uw’ in Dutch) and the gender of the speaker or addressee can influence the inflection of words (e.g., in Portuguese).

To summarise, major challenges in the authoring process that a dialogue platform should address are to develop *reusable* dialogues, keeping an overview of *branching* dialogues, *tailoring* of dialogues, and *translation management*.

2.3 Related work: Dialogue authoring tools

There is a large body of research on the topic of dialogue, conversational agents, and chatbots. Various tools and software platforms have been developed in the past to enable their development. Different tools are suited for different types of applications, authors, and developers. In the following subsections, we briefly discuss platforms that specifically include tools for authoring the dialogues, since we see dialogue authoring by (possibly non-technical) domain experts as an important step in the conversational agent development process. In our overview, we have included commercial, scientific and open source tools (as distinguished by Green et al. (2018b)). We will discuss the tools by dividing them into three groups, based on the type of agent that they are intended for and discuss our conclusions for development of the WOOL Dialogue Platform.

2.3.1 Tools for building chatbots

One group of tools that can be distinguished are so-called *chatbot building services*. Examples include commercial services such as Amazon’s Lex (Amazon, 2020), Microsoft’s Azure Bot Service (Microsoft, 2020), Google’s Dialogue Flow (Google, 2020), IBM’s Watson (IBM, 2020), (Facebook’s) Wit.ai (Wit.ai, Inc., 2020), Landbot I.O. (Landbot.io, 2020), Botsociety (Botsociety.io, 2020) and Bot Builder (SAP Conversational AI, 2020) or open source chatbot services (e.g., Rasa (Rasa Technologies GmbH, 2020) and DialogOS (Saarland University, 2020)); to name a few. These services are typically aimed at users who want to build agents that have short interactions with their users, often used in contexts like customer support, question answering or operating smart home appliances. Generally speaking, these services let developers define *intents* that the user of

their application can have (e.g., book a flight), *entities* that are relevant to those intents (e.g., Amsterdam - as the intended city to travel to) and add *actions* that can be triggered by those intents (e.g., a search for available plane tickets or a dialogue that asks more questions about the date and time).

While these chatbot tools can be used to build conversational agent systems that respond to the user with short dialogues, these dialogues are typically limited to short dialogues, that are not really aimed to mimic a real human-human dialogue. Furthermore, they are usually not intended for designing interactions with agents that can carry over multiple sessions – an aspect that is essential in the health coaching domain.

2.3.2 Tools for building ECAs

A second group that can be distinguished are scientific platforms or tools aimed at building embodied conversational agents (ECAs) in research settings. These tools are prevalent in the educational domain, where they are used to build intelligent learning systems such as tutoring agents or (3D) simulations.

Tools such as ASPIRE (Mitrovic et al., 2009), CTAT (Aleven et al., 2006) and ASAT (Cai et al., 2015), and GIFT authoring tools (Ososky & Sottilare, 2016) have been developed to generate the rules that form the domain model for knowledge-based intelligent tutoring systems following a domain expert's input (Mitrovic et al., 2009), but these are not intended for ECA development. CO-AUTHOR (Wang et al., 2019) is a tool which has a conversational interface aimed at building knowledge-based agents. The FAtiMA Authoring Tool (Guimarães et al., 2019) is a tool that allows agent developers to define dialogue actions and cognitive rules for knowledge-based FAtiMA agents. It also provides a (text-based) simulator to test the resulting interactions and inspect changes in underlying variables. Tools such as NPCEditor (Leuski & Traum, 2011), and Virtual Human Toolkit (USC Institute for Creative Technologies, 2021) are aimed at building the natural language processing components for agents that can respond to user utterances (question-answering agents), while tools like TuTalk (Jordan et al., 2007) are aimed at agents for tutoring conversations using natural language.

Then there are tools intended for authoring serious games, which include functionalities for, for example, building storylines, aspects of the 'world' that a user interacts with (e.g., Mehm et al. (2012)) and even serious games with agents, such as virtual humans for tactical questioning (Gandhe et al., 2005) or believable 3D agents in interactive training applications (e.g., Johnson and Valente (2008) and M. van Velsen (2008)), that are intended to be easy to use in the development of complex applications. The SceneMaker toolkit (Gebhard et al., 2003) was designed for rapid prototyping of interactive agents by non-experts, but it

was designed for a specific target application. Furthermore, Gaffney et al. (2008) review three tools (VISIOn, Experience Builder, Captivate 3) for authoring soft skill simulations.

While authoring tools for knowledge-based systems make it possible for domain experts to author content, the underlying representations for the resulting systems and often the dependencies on specific computationally-heavy agent applications make these tools difficult to use for the (rapid) development of prototypes.

2.3.3 Tools for building interactive stories and (narrative) games

A third group of tools, mostly from the open source domain, is used to build interactive stories and (narrative) games, which often feature conversational agents as characters. These tools typically include dialogue flow graph-based editors (e.g., Inklewriter (Inkle Studios, 2019), Texture Writer (Texture, 2019), and StorySpace 3 (Eastgate Systems, Inc., 2020)) in which the statements by an agent are represented as a network of connected blocks, or natural language programming languages (Inform 7 (Graham Nelson, 2019)). TADS 3 (Roberts, 2020) and Villanelle (scientific) (Martens & Iqbal, 2019) on the other hand are more similar to programming languages.

Articy is a commercial tool aimed at professional game developing teams (Articy Software GmbH & Co. KG, 2020). Fungus (Fungus Games, 2020) is developed for scripting narrative games in Unity 3D. Other examples include Quest (textadventures.co.uk, 2020a), Ren'Py (The Ren'Py Community, 2019), Genarrator (The Genarrator Community, 2019), Dialogue Designer (The Yarn Team, 2020), Yarn (Radmatt, 2020) and the Yarn editing tool Yarn Spinner (SecretLab, 2020), and Squiffy (textadventures.co.uk, 2020b). Many tools within this category have been inspired by Twine (Interactive Fiction Technology Foundation, 2020).

Scenejo (Spierling & Szilas, 2009; Spierling et al., 2006), Cyranus (lurgel, 2008), Story Canvas (Skorupski & Mateas, 2010) and Scribe (Medler & Magerko, 2006) are all scientific tools for building interactive stories with agents, which, as was the case for tutoring agents have knowledge-based representations. Furthermore, a state-of-the-art scientific tool in this group is Mimisbrunnur by Stefansson and Thue (2018), which uses natural language processing and mixed-initiative exploration (the system providing suggestions to the author) to help the author create an outline for their generative, interactive story. Finally, Novella is a recently published scientific tool that allows for graph-based editing (Green et al., 2019) of interactive narratives.

In the end, the best candidates for what we needed were found in this category of dialogue tools. These candidates were Yarn (Radmatt, 2020) and the Yarn Spinner tool (SecretLab, 2020), and Twine (Interactive Fiction Technology

Foundation, 2020), which thus form the basis and inspiration for the WOOL Dialogue Platform. Novella (Green et al., 2019) would also have been a candidate but was released after our initial search.

2.4 Requirements

In our search for suitable dialogue authoring and agent development tools, there were several platforms available in the commercial, scientific and open source communities that ticked many of the boxes on our wish-list. However, there was not one tool or platform that completely suited our needs. Based on our experience with the dialogue authoring process and its major challenges, and our exploration of available tools, we created a set of requirements for our dialogue platform that are listed below.

2

2.4.1 Non-functional

The authoring tools had to (a) have an intuitive user interface that made dialogue authoring very user-friendly and easy to learn; even for novice authors, (b) include a testing feature that would allow authors to test their dialogue while editing.

2.4.2 Functional

The dialogue language had to support (a) structuring scripted dialogues into several files, which would allow them to be easier to reuse, (b) conversations with multiple conversational agents, (c) a simple way to define basic dialogues, but also (d) powerful features when needed, such as the use of variables, conditionals, input fields, and special actions. The editor had to have a (e) graph-based dialogue flow editing interface, since that would provide a clearer overview of dialogues steps than form-based or text-based interfaces (Green et al., 2018a).

2.4.3 Technical

The platform's tools should support (a) inclusion of the developed content in all kinds of applications (ranging from mobile apps to tablet applications to desktop applications), (b) dialogue authoring on multiple operating systems, (c) dialogue authoring with an easy editor setup/installation, and similarly, (d) dialogue execution without major requirements for an operating system or installed software. Finally, the platform had to be (e) open source, so that it could benefit from community suggestions and contributions.

2.5 Methods

2.5.1 Implementation

Development of the WOOL Dialogue Platform started with a version of the Yarn (Radmatt, 2020) narrative game language. We made adjustments that suited our context with multiple conversational agents, and added functionality to support various data input fields, which in turn resulted in the development of custom execution libraries and addition of testing functionality to Yarn's editing tool (Yarn Spinner (SecretLab, 2020)). In the end, this process resulted in a language and a set of tools that was significantly different from the original language and tools: The WOOL Dialogue Platform.

Figure 2.2 shows a typical setup of a WOOL application that has a client-server architecture. The Author uses the WOOL Editor to write a set of scripts that are stored in the WOOL Web Service. The WOOL Core libraries can be used inside the web service to parse (WOOL Parser) and execute (WOOL Executor) these scripts, triggered by a request from a client application. During execution, variables in the scripts are retrieved from and then stored in the WOOL Variable Store. The open source platform provides all these functionalities, and lets developers focus on the creation of the specifics of their next user interfaces (Agent UI).

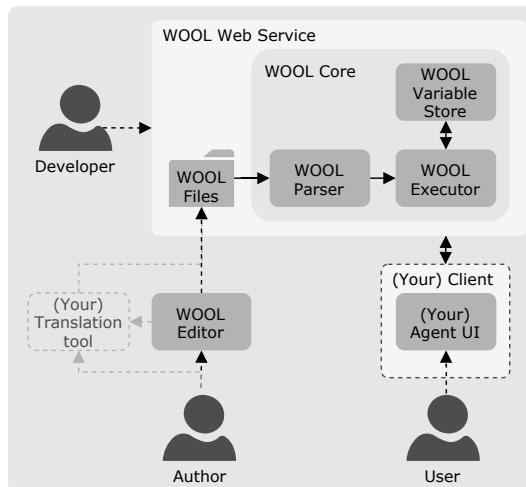


Figure 2.2: Typical setup of an application using the WOOL Dialogue Platform.

In the following three subsections, we will elaborate on the platform's three main components: the language, the editor, and the core libraries and web service.

The WOOL dialogue definition language

While detailed language documentation is available on the platform's website (Roessingh Research and Development, 2020), and may undergo changes as additional requirements emerge and functionalities are added, in this paper we will describe the basics of the WOOL language, and highlight features that support authors and developers by addressing dialogue authoring challenges and fulfilling our requirements.

The basics

The WOOL dialogue definition language defines, in essence, a series of dialogue steps or *nodes*. Each of these nodes has a header and a body. Consider, for example, the following example code for a single node:

```
title: Start
speaker: Robin
position: -416, 112
color: cyan
---
Hello, my name is Robin!

[[Nice to meet you, Robin!|NodeRobin1]]
[[Goodbye.|End]]
==
```

The header in this example contains a *title* and a *speaker*. The node's title ('Start' in this case) can be used when referring to the next node from a reply, and the speaker indicates the name of the agent that makes the statement ('Robin' in this case). The *position* and *color* are technically optional, but are used for depiction of the node in the editor. It is also possible to include any number of optional *tags* in the header.

The body of the example node contains two other key elements for defining a WOOL dialogue, namely a *statement* and two *replies*. The statement ("Hello, my name is Robin!") is a basic example of what an agent can say. The replies contain the user's response options (e.g., "Nice to meet you Robin!") and a reference (*node pointer*, e.g., 'NodeRobin1') to the next node when that reply is chosen.

By linking nodes such as the one in the example, an entire dialogue can be defined, which makes WOOL easy to learn and use for novice authors (especially when they use the editor). However, the language can also be very powerful if needed. Authors have the possibility to add *variables*, *set-statements*, *input fields*, *conditionals*, and *actions*, which give them more control of the dialogue flow, means to personalise content, and influence other user interface elements.

The setup of the language in this manner allows for a non-technical expert to

write an entire dialogue using the language's most basic features (perhaps including *comments* for the developers), and then a developer could add additional elements that make the dialogue more dynamic or tailored. More experienced authors, however, could immediately write more complicated dialogues.

Language features addressing authoring challenges

To support *reusability* of dialogue scripts, the reference to a next dialogue step from a reply does not have to refer to a reference to a node in the same dialogue script, but it can also be a reference to a specific node in a different dialogue script. Through this feature and the obligatory use of a 'Start' node, authors can keep the size of dialogues they write limited and various dialogue scripts can be chained together as needed. Since this allows authors to limit a dialogue to one topic, they can also have a better overview of that specific dialogue, which helps with *branching*.

Tailoring within dialogues is supported in the WOOL language through the use of variables, conditionals, set-statements and input fields. Authors can include variables in sentences, which are filled in with stored information about the user on dialogue execution (e.g., Welcome back, \$userFirstName!). These variables are also included in conditionals. For example, an agent could advise a user to set a goal of 10000 steps when they are younger than 65, and a goal of 7500 steps when they are older than 65. The text in the editor could then be the following:

```
<<if $userAge >= 65>>
A daily step goal of 7500 steps would be enough!
<<else>>
How about a daily step goal of 10000 steps?
<<endif>>
```

To refer to the user's personal goal in a future interaction, a set-statement could be included to store that goal (e.g., «set \$dailyStepGoal = 7500»). These set-statements can be included in the statement-part of a node, but they can also be attached to specific replies (e.g., to store an answer to an 'are you satisfied with your current physical activity?' question).

Storing information can also be done through input fields. These can be included in a reply and allow users to input information. For example, My name is «input type="text" value="\$userFirstName"»..

Finally, though not addressing an authoring challenge, actions can be included in dialogues. These can be links to other webpages, triggers to open a widget (e.g., the coach opens a recipe book or shows an activity graph), or any other functionality that the developers implement for the agent application.

ENABLING USER-FRIENDLY DEVELOPMENT OF DIALOGUE

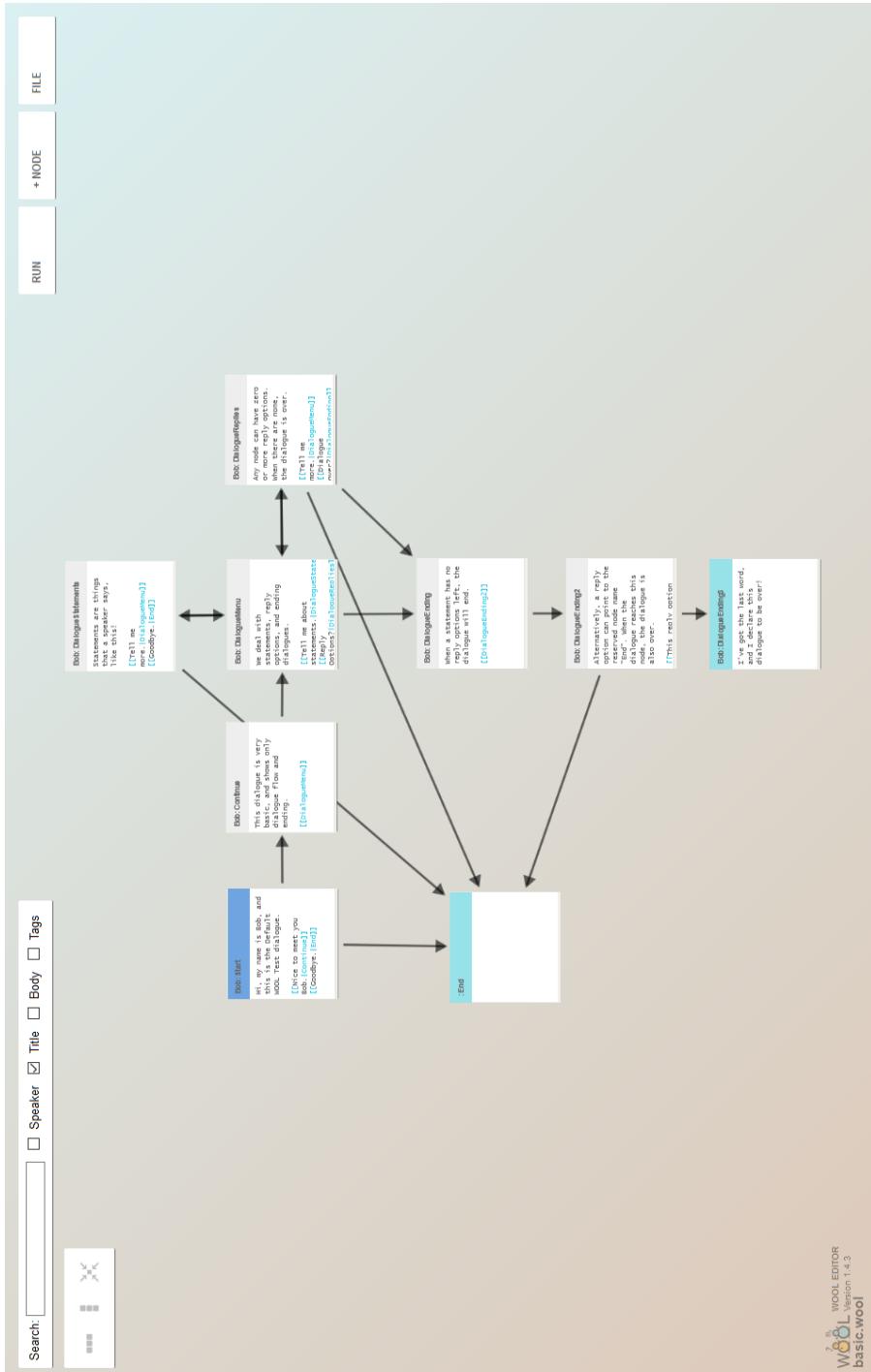


Figure 2.3: The overview screen in the WOOL Editor.

The WOOL Editor

The editor is a web-based editor that allows users to write and test dialogues. It can be used in a browser or as a desktop application (using the open source Electron framework (Github Inc., 2020)). The desktop version also includes file management features. The editor has two main screens, namely a dialogue overview screen (see Figure 2.3) and a ‘run dialogue’ screen (see Figure 2.4).



Figure 2.4: The node editing overlay in the WOOL Editor.

The dialogue overview screen allows users to create new nodes and edit their content. This part of the editor has its origins in the Yarn Spinner tool (SecretLab, 2020). When a node refers to another node (through a node pointer in a reply), the two are connected by an arrow, thus creating a directional graph that provides users with an overview of the structure and the dependencies in their dialogue (which helps with the *branching* challenge). When a node in the editor is double clicked it opens the node editing overlay, which allows users to edit the node, for example, by adding or changing the agent name, node name, tags and node content in its body (statements, replies, conditionals, etc.).

In the desktop version of the editor, we added file management, to support authors when they create multiple dialogues on multiple topics (to make dialogues *reusable*). We also added syntax checking, which is performed on the content of nodes. Notifications for syntax errors can be seen both in the node view and in front of the line with the error in the editing overlay.

The second main screen, the ‘run dialogue’ screen, features a randomly generated

2D agent and background (which can be customised) (see Figure 2.5). This screen allows users to interact with the dialogue to test its functionality (e.g., for node pointers and conditionals) and experience; that is, is the flow of the conversation as intended? If a dialogue contains multiple agents as speakers, these are shown sequentially – that is, each name is assigned a different agent image, which is shown when that agent participates in the dialogue. Authors can open an overview of the variables in their dialogue, to verify that these are updated correctly throughout the dialogue. They can also go to a current node in the editor screen to make a change and then continue testing where they left the dialogue. Additionally, the editor can export a dialogue as a shareable URL that can function as a prototype.

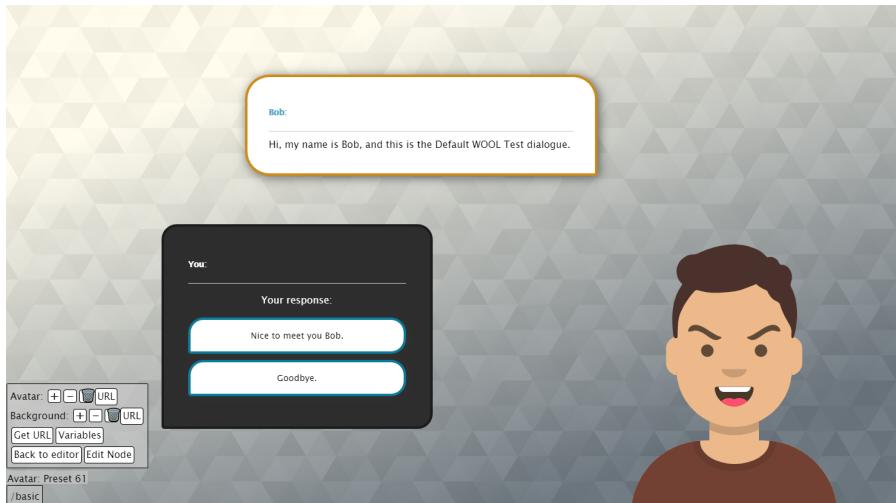


Figure 2.5: The run and test screen in the WOOL Editor.

The editor allows the user to save their created dialogues as .wool files. When dialogues need to be translated to other languages for multi-lingual applications, the editor also allows to export all the unique sections of text ("terms") that occur in a dialogue to a JSON file. These exported terms can subsequently be imported in popular tools for translation, such as POEditor (Code Whale Inc., 2020), which can be used to generate the translated terms files. The resulting translations can be imported in the editor. The execution libraries for the WOOL Dialogue Platform have built-in support for reading these translation files and combining them with the original source .wool script.

The WOOL libraries and Web Service

The third key component of the WOOL Dialogue Platform is a set of Java libraries (the WOOL Core) that conversational agent developers can use to build their own

conversational agents. These can be used in both web services and Android apps. The Java libraries provide classes to parse WOOL scripts, store variables and execute the dialogue definitions, for example, by providing methods for starting and progressing a dialogue based on selected replies.

The platform also includes the WOOL Web Service, which is a Java Spring Web Service built around the WOOL Core. It allows developers to build their own web service using a basic service with endpoints for starting and progressing dialogues and basic user management. It also supports the execution of dialogues in multiple languages (if translation files are added). The inclusion of basic authentication/authorisation means that a first version of an application has basic security features when deployed to a webserver. User information and data are stored in files, but developers can exchange this for their database of choice.

2.5.2 Operation

The desktop version of the WOOL Editor requires an installation of NodeJS (≥ 14). The WOOL Core libraries are intended to be used with developer's own source code (Java, JDK ≥ 8). Running the WOOL Web Service requires Java (JDK ≥ 8) and Tomcat (8.5). Setup instructions can be found in the ReadMe files included in the WOOL repository. Language documentation and tutorials can be found in the documentation section of the WOOL website¹.

2.6 Use cases

The main motivation for developing the WOOL Dialogue Platform was to facilitate the creation of conversational agents for both (novice) authors and developers. In this section we present and discuss two use cases for the platform. The first is its use in the EU Horizon 2020 project Council of Coaches (grant agreement #769553) that developed two applications for group conversations with multiple embodied conversational agents. The second is the use of the platform in a smartphone application featuring a single embodied conversational agent built for the EU Horizon 2020 project SmartWork (grant agreement #826343).

2.6.1 The Council of Coaches demonstrators

The Council of Coaches project (op den Akker et al., 2018) aimed to develop a virtual group of coaches that can support users in their health behaviour change process. The embodied conversational agents each have their own expertise (e.g., physical activity or nutrition) and personality. Using the WOOL Dialogue Platform, a functional demonstrator was developed that allowed for user-testing of the concept and dialogue content, which then in turn could be incorporated in the project's technical demonstrator

¹<https://www.woolplatform.eu/docs/>

(which is computationally heavier and was initially less suited for conceptual user-testing).

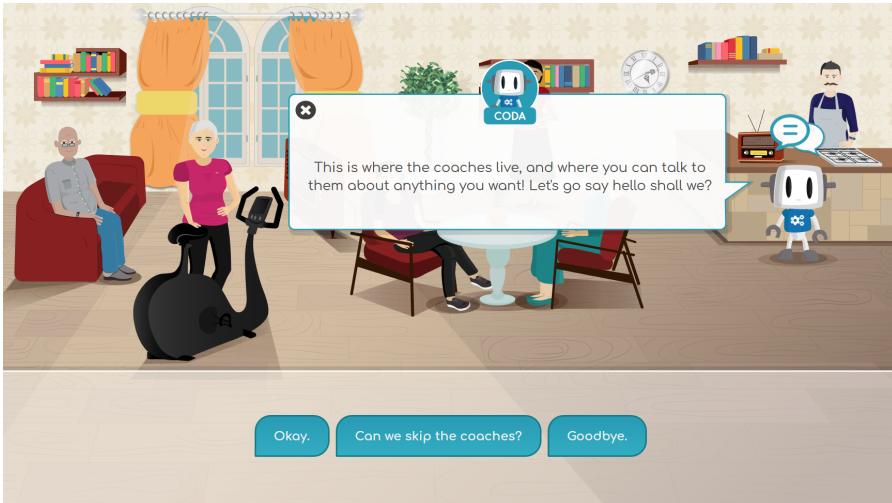


Figure 2.6: A screenshot of the Council of Coaches functional demonstrator.

In the 2D functional demonstrator, participants see a living room with the agents in their personal council (see Figure 2.6). They can start the interaction with an agent by clicking on that agent and are then able to have a conversation with that agent following a speech-bubble and reply-buttons paradigm. Server-side, each agent is a separate software agent, with their own set of dialogues (and translation files; since testing was performed in the Netherlands, Scotland, and Denmark). Conversations are tailored to the user using the features of the dialogue scripts (input fields, variables, set-statements and conditionals) and the interface can be influenced by the actions in the scripts (e.g., to show a recipe book that the coach can refer to). Experts on various domains used the editor to write and test dialogues that were then integrated in the system by the system developers. In this case the experts were not programmers (e.g., a psychologist, a movement scientist), but were used to cooperating on technical projects. They were given a short explanation before using the editor and a link to the language documentation – which in most cases was sufficient information to let them write and test dialogues on their own. Multiple demonstrators were created (in English, Dutch and Danish), with different types of content, but the final demonstrator contained content for 8 different agents (6 coaches, a peer and an assistant). A total of 160 dialogues were written, which contained 1973 dialogue steps and 4116 unique phrases. All dialogues were available in two languages (English and Dutch)².

The WOOL Dialogue Platform was also included in the technical demonstrator, which aimed to integrate state-of-the-art software such as Flipper (van Waterschoot et al., 2018), ASAP (van Welbergen et al., 2014), GRETA (Niewiadomski et al., 2009), and the Platform for Argument and Dialogue (which includes DGEP (Bex et al., 2014) and

²The Council of Coaches Demonstrator is available on: www.council-of-coaches.eu.

the DUG (Snaith et al., 2020)) in one multi-agent system³. The WOOL Dialogue Platform's Web Service fulfils three functions in this system. The first function is that it is used to store and retrieve variables. Second, it provides the system with user management and authentication/authorisation. Its third function is in the use of scripted dialogues. One the one hand, the WOOL editor was used to write dialogues that served as a source for the design of dynamic dialogue games for use in DGEP. On the other hand the WOOL Web Service was used to execute scripted dialogues that are directly used for the statements by the agents and the replies that the user has available (although behaviour instructions for the agents are added onto the executed statements). In this manner, the platform was used as a method for providing both dynamic (indirectly) and static dialogues (directly).

2.6.2 The SmartWork HealthyMe app

In the EU Horizon 2020 research and innovation project SmartWork, (grant agreement #826343), aimed at coaching office workers on various domains, the platform is used to embed an embodied conversational agent in a smartphone application for the health behaviour change service (HealthyMe, see Figure 2.7). The agent, that has the role of a lifestyle coach, provides advice on physical activity, sleep, nutrition and well-being. Through dialogues and a wearable activity sensor the agent gathers information about the user and adjusts its advice accordingly.

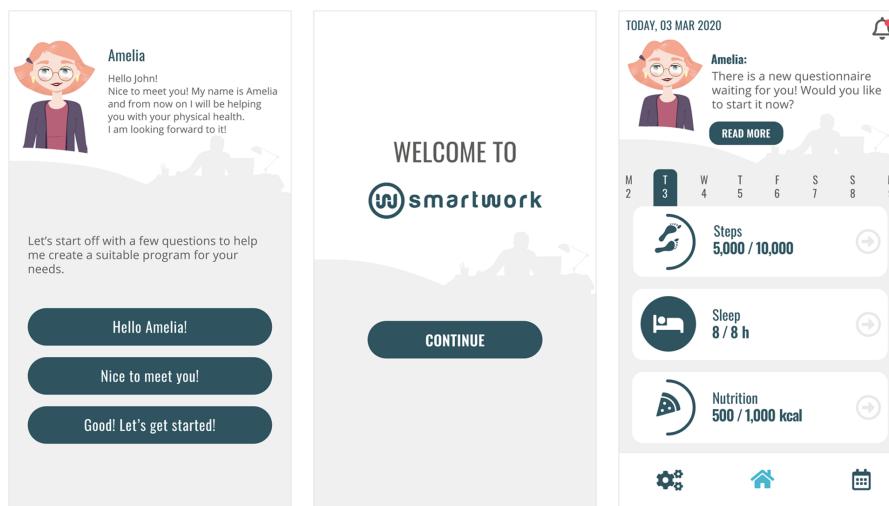


Figure 2.7: Three screenshots of the SmartWork HealthyMe smartphone application.

In the SmartWork project, the editor is used by several domain experts to write dialogues and test these before they are included in the application. The WOOL Core is integrated

³The resulting multi-agent platform has subsequently been released as the Agents United Platform (Beinema et al., 2021) through www.agents-united.org.

in the smartphone application to allow users of the app to have conversations with the coach. Execution of dialogues takes place fully on the smartphone, which means that the application works without an internet connection.

2.6.3 Other use cases

In addition to the projects mentioned above, early versions of the platform have also been used for the assistive agent in the Android app for the Horizon 2020 project GOAL (aimed at older adults) and for the web-based frailty screening agent in the AAL project FRAIL (aimed at frail older adults).

Furthermore, the platform was used for development of the web application in the ZonMW Create Health project PACO (featuring two embodied conversational agents for dietary behaviour change in older adults; grant number 40-44300-98-110), is used in the AAL project LEAVES (featuring an embodied conversational agent for older adults with prolonged grief due to loss of a spouse; grant number AAL-2019-6-168-CP), and it is being included in the smartphone application for the EU Horizon 2020 project Bionic (grant agreement #826304) (which focuses on lifestyle coaching for blue collar workers).

Overall, the WOOL Dialogue Platform is used by an increasing number of domain experts, developers and agent designers, who are actively contributing to the platform's usability and functionalities.

2.7 Conclusions

In this paper, we have presented the WOOL Dialogue Platform (Roessingh Research and Development, 2020). Originally developed to fulfil a need within a single project, it now provides the eHealth and conversational agent communities with an open source and easy to use set of tools that can be used in the development of conversational agent systems and the content for these agents. Whether the aim is to use scripted dialogues written by domain experts for the direct (or indirect) execution of dialogues or as a basis for a demonstrator or end product, the WOOL Dialogue Platform allows system experts and domain experts to work together, and allows for early stage user-testing applications to be developed.

With many ongoing projects building on the WOOL Dialogue Platform, the future for WOOL is looking bright. Future developments focus on improving the platform's usability for non-technical users, localisation support, and releasing additional application components as open-source assets for application developers.

Software availability

Software and documentation available from: <https://www.woolplatform.eu>

CHAPTER 2

Source code available from: <https://github.com/woolplatform>

Archived source code at time of publication: <https://doi.org/10.5281/zenodo.5654838>

License: MIT License

Acknowledgements

We would like to thank Jon Manning and the Yarn community for providing the starting point of the WOOL Dialogue Platform, and all past, present and future contributors to the open source WOOL Dialogue Platform.

Intermezzo

An artificial coaching engine

First steps in tailoring coaching conversations

The tools and dialogue definition language that are presented in **Chapter 2** facilitated dialogue authoring and testing, and provided features that would be useful in tailoring coaching dialogues for conversational agents. For example, the possibility to chain together dialogues from different files, to use conditionals in dialogue steps and to use variables in statements. In addition to the development of the WOOL Dialogue Platform, the first user study for the Council of Coaches had also been conducted. In two focus groups, participants could try out three prototypes for the Council of Coaches application with varying interfaces and levels of interactivity (see Figure I.1). The focus groups showed that participants enjoyed interacting with a group of coaches, and that they appreciated the broader characters and personalities that were defined for the coaches. They also provided further insights into participants' preferences for coaching topics and agent characteristics (as reported in Council of Coaches Deliverable 2.4, detailing that first evaluation (Beinema, op den Akker, Broekhuis et al., 2018)).



Figure I.1: Screenshots of two of the prototypes that participants interacted with during the first user study in the Council of Coaches project.

When investigating tailoring possibilities, it became clear that much of the work on coaching with (embodied) conversational agents in eHealth seemed to focus on the tailored presentation of information that followed a predefined plan (e.g., by adjusting the value of goals that are suggested). Tailoring of coaching dialogues on a higher level, that is, not just *how* something is discussed, but *what* had to be discussed seemed to be a research gap. Thinking about how that type of tailoring could be implemented led to the concept of an artificial coaching engine (Beinema, op den Akker & Hermens, 2018). Such a coaching engine would be a component that was the 'coaching brain' of the conversational agent, that is, it would determine what the coach should say next. In the initial concept, the engine would make decisions on which goals to work towards (or suggest to the user), would select strategies for reaching those goals, and would select actions that supported the execution of those strategies as dialogues. Goals would be structured hierarchically, so that multiple subgoals together contributed to a

larger final goal (e.g., 'user is physically active', as detailed in Council of Coaches Deliverable 3.3 (Beinema & op den Akker, 2019)). Even though the role of strategies later shifted from being templates for actions to being parameters in the selection of tailored content, their function in determining and guiding the long-term coaching approach remained.

In parallel to the technical considerations, background research on behaviour change was conducted to investigate how the content of coaching dialogues and strategies should be defined, and which information had to be included in the user model to support tailoring. This led to an overview of behaviour change theories, behaviour change techniques and persuasive features (as reported in Council of Coaches Deliverable 3.1 (Beinema, Huizing et al., 2018)). The document also included a list of important factors that were extracted from the behaviour change theories. One of these factors, which has an important role in many behaviour change theories was *motivation*. In addition, a new study reported findings that persuasive features could be tailored to motivation to live healthy (van Velsen et al., 2019). The next prototype for the Council of Coaches was therefore implemented with two aims. First, it would be used in a usability study that investigated how users would perceive and interact with the new features (documented in Council of Coaches Deliverable 2.5 (Beinema et al., 2019)). Second, it was used in an online investigation into tailoring coaching strategies to users' motivation to live healthy in the context of coaching dialogues. This second study is the subject of **Chapter 3**.

Chapter 3

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

Based on:

Beinema, T., op den Akker, H., van Velsen, L., and Hermens, H. (2021). Tailoring coaching strategies to users' motivation in a multi-agent health coaching application. Published in *Computers in Human Behavior*, vol. 121, 106787. <https://doi.org/10.1016/j.chb.2021.106787>

Abstract

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

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

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

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

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

3.1 Introduction

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

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

The inclusion of ECAs in health behaviour change applications serves multiple purposes. One is that ECAs can make the use of a health application easier, more satisfying, and less frustrating (André & Pelachaud, 2010; Bickmore et al., 2016). Second, their potential to engage the user can help improve users’ adherence and participation (Bickmore et al., 2010; Scholten et al., 2017) and can potentially increase an application’s effect in changing the user’s behaviour (Ma et al., 2019). ECAs can give the system social ability, which is an important factor for “establishing and maintaining a collaborative relationship between user and system” in such applications (Kamphorst, 2017). Third, an ECA’s credibility, similarity, authority, power, and social attractiveness can all contribute to the persuasion of a user (Pickard, 2012, pp. 73–74), which can be useful when providing behaviour change support. Social attractiveness specifically can influence persuasion by enhancing the positive or negative perception of the message.

In a health behaviour change application, ECAs can take on the role of a virtual coach (Kramer et al., 2020). From counselling literature, we know that the quality of a working alliance between a counsellor and client is a factor in

therapeutic change and adherence (Castonguay et al., 2006) and that it has three key aspects, namely goal agreement, task agreement, and development of a personal bond (Hatcher & Gillaspy, 2006; Horvath & Greenberg, 1989). To ensure that the advice given and topics addressed by an ECA serve such long-term purposes as building up a working alliance and supporting a user's behaviour change, coaching dialogues with an ECA are guided by a *coaching strategy*. These coaching strategies are carefully designed by digitising existing interventions for the target population and domain, and combining those with communication, persuasive design and human-computer interaction principles, such as tailoring to the individual user (Krebs et al., 2010; K. Ryan et al., 2019; Wangberg et al., 2008).

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

3.2 Theoretical background

3.2.1 Tailoring

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

have the same preferences or respond well to the same approach. Therefore, we investigate how ECA's high-level coaching strategies can be tailored.

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

3.2.2 Tailoring of coaching dialogues and strategies

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

Tailoring of coaching dialogues and strategies is a process that involves research on multiple facets of health application development. As Paramythios et al. (2010) describe, each layer of an interactive adaptive system should carefully be designed

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

3.2.3 Coaching strategies and motivation

Motivation is an integral element in changing (health) behaviour and a construct that returns in many theories of behaviour change. Examples range from classic Operant Conditioning (Miltenberger, 2008, p. 141), the Information-Motivation-Behavioural Skills Model (Fisher & Fisher, 1992), Protection-Motivation Theory (Norman et al., 2005; Rogers, 1989), and the Transtheoretical Model (Prochaska & Velicer, 1997) to Self-Determination Theory (R. M. Ryan & Deci, 2000) and the Fogg Behavior Model (Fogg, 2009). To take the Transtheoretical Model as an example, without (a change in) motivation people are less likely to move from one stage to a next stage, and are thus less likely to ultimately end up in the desired maintenance stage.

In health coaching applications, there has been some investigation into how users can be effectively motivated. Examples of approaches include the translation of a human-human approach such as motivational interviewing to the ECA setting (Olafsson et al., 2019) or sending motivational messages that are tailored to a user's personality (de Vries et al., 2016), and a recent scoping review studied which behaviour change techniques and persuasive system design principles have been reported in relation to motivation and adherence in weight loss applications (Asbjørnsen et al., 2019). Tailoring dialogues with ECAs can also have a positive effect on motivation, for example, as found for tailoring explanations to students' personal context and goals in order to change planning-behaviour for stress reduction (Abdulrahman & Richards, 2020). However, most studies either do

not tailor their approach or do not tailor it to a user's motivation specifically. Recently, van Velsen et al. (2019) investigated if persuasive features can be tailored to a user's motivation to live healthy in the context of digital health (eHealth) applications.

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

3.2.4 Multi-agent coaching

Furthermore, most health coaching applications that use ECA technology, provide a single ECA as a coach who then provides coaching on one domain (e.g., for physical activity (King et al., 2017; Watson et al., 2012)). Health, however, is a construct that often includes multiple domains (Huber et al., 2016; World Health Organization, 1946). Thus, changing behaviour to lead a healthy lifestyle often requires a holistic approach. To that end, recent research has been investigating coaching on a combination of domains (e.g., Gardiner et al. (2017) and Klaassen et al. (2018)) and even multiple coaches (Das et al., 2019; Hurmuz et al., 2020; op den Akker et al., 2018). Interaction with multiple agents at the same time provides opportunities for vicarious persuasion (Kantharaju et al., 2018) and engagement (André & Rist, 2001). It also allows for a decentralised presentation of domain information, for example, by casting the agents as coaches that each have their own expertise, thus providing the possibility to include multiple viewpoints without an agent contradicting itself (Kantharaju et al., 2019). We therefore perform our experiment in a setting where multiple ECAs are present

and interact with the user, so that the results can be incorporated in strategies for a broad range of ECA health coaching systems.

3.2.5 Research question and hypotheses

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

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

This leads to the following research question:

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

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

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

As stated above, participants in our experiment participated in a group conversation with multiple ECAs and two of these ECAs will present a strategy to the user. However, differences between agents might cause users to perceive them and the presented strategies differently. We therefore wanted to take possible influences of agent perception on strategy appreciation into account. Previous research has shown that user's perception of agents can influence their perception of a message (e.g. Ruijten et al. (2014) and Schulman and Bickmore

(2009)), the likeliness of them following an agent's advice (ter Stal et al., 2019), and that it can also have an influence on the answers that are given to survey questions (Kim et al., 2019). Research on human-human persuasion has shown that *source likeability* influences persuasiveness of a message (Chaiken, 1980) and for ECAs specifically, Pickard, p. 74 defined likeability to be a dimension of an ECA's social attractiveness that influences the persuasiveness of an ECA's message. This leads to the following hypothesis:

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

In addition to the questions on strategy appreciation and the likeability of the ECAs, we asked participants for their demographics. Various demographics have been shown to have an influence on engagement and appreciation for eHealth applications (Hardiker & Grant, 2011; Perski et al., 2017) and ECAs (e.g. Krämer et al. (2010), Payne et al. (2013) and Pezzullo et al. (2017)). Level of education and living situation specifically were influences for the appreciation of the persuasive features on which we based our strategies (van Velsen et al., 2019). This leads us to our last hypothesis (an overview of our hypotheses can be found in Figure 3.1):

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

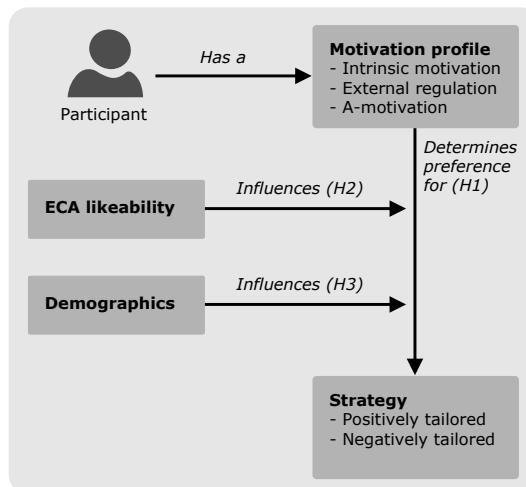


Figure 3.1: Overview of the hypotheses.

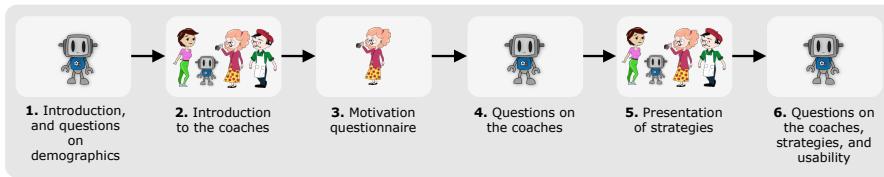


Figure 3.2: Overview of the six steps that participants go through in the experiment.

3.3 Methods

We conducted an online experiment that consists of six steps, in which the participant interacted with our agent application. A schematic overview of these steps is depicted in Figure 3.2, and they can be described as follows:

Step 1 The participant meets a virtual robot (named “Coda”), who in a one-on-one conversation introduces the experiment and poses the questions on demographics.

Step 2 Coda introduces the participant to the three coaches (Alexa, François, and Helen) in a group conversation. The participant gets to interact with the coaches and learn about the coaches’ backgrounds (e.g., their origin and hobbies) and expertise (physical activity, nutrition, and cognition, respectively).

Step 3 The participant has a one-on-one interaction with coach Helen in which she asks them to respond to the statements from the questionnaire on personal motivation to live healthy (van Velsen et al., 2019). The results from this questionnaire are used in the background to classify the participant into one of our four motivation groups. This classification will determine the two strategies that are presented to the participant in step 5.

Step 4 While the coaches are ‘deliberating’, the participant has a one-on-one intermezzo with Coda. Coda asks the participant to rate the likeability of each of the coaches and he asks them to indicate their preference for a coach.

Step 5 After a short introduction by Coda and Helen (who asked the motivation questions), coaches Alexa and François both propose a coaching strategy. One strategy is the positively tailored strategy for the participant’s motivation group, the other strategy is the negatively tailored strategy. The presentation of these two strategies is randomized over Alexa and François, as is the order in which they get to present their strategy.

Step 6 In this last step, Coda asks the participant to again rate the likeability of the coaches and to indicate their preference for a coach (as was done in step 4). He also asks participants to rate both strategies, to choose their preferred strategy from these two, and to rate the usability of the system (a control question to ensure that the system had a high enough usability to conduct the experiment).

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

3

3.3.1 Implementation

The experiment was conducted using a fully-functional multi-party conversational agent system. The system's server provided user and dialogue management. The interface consisted of a webpage that showed the agents, their speech-bubbles, and the reply options. A screenshot of this multi-agent interface can be found in Figure 3.3. The dialogues between the coaches and the user were specified using the WOOL Dialogue Framework (Roessingh Research and Development, 2020).

3.3.2 Motivation group classification

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

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

Third, we used the motivation profile to classify participants into a motivation *group*. Four motivation groups were defined, namely: Intrinsic Motivation, External Regulation, A-motivation and Dual Motivation. We decided to add the Dual Motivation group since previous research shows that people may have combined types of motivation (Gourlan et al., 2016) and that people can be

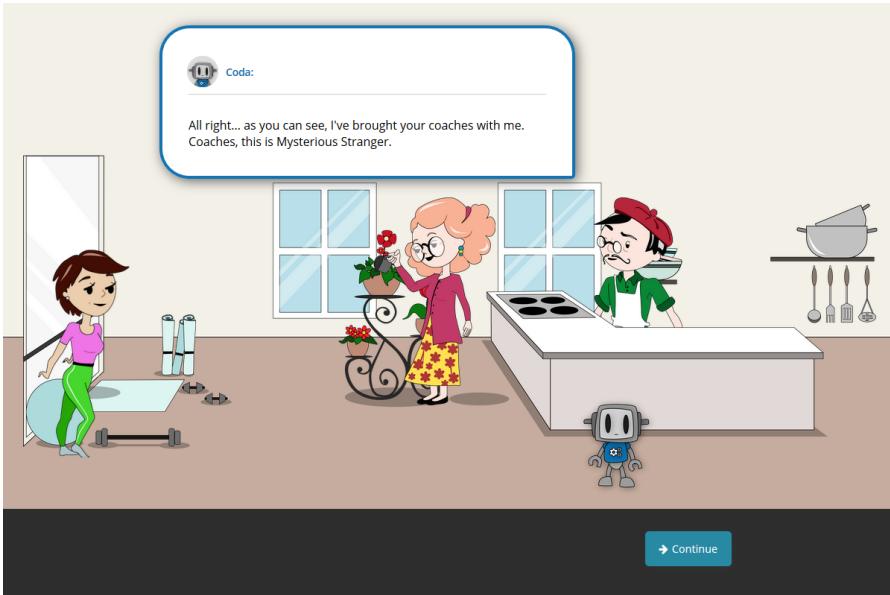


Figure 3.3: A screenshot of the multi-agent interface. From left to right the agents shown are: Alexa (physical activity coach), Helen (cognitive coach), Francois (nutrition coach), and the assistant robot Coda in front of the counter.

both intrinsically motivated and externally regulated (van Velsen et al., 2019). The specific classification rules can be found in Table 3.1. A threshold value of 0.2 was used for the classification into the Intrinsic Motivation and External Regulation groups.

3.3.3 Strategy design

For each of the motivation groups we designed a positively tailored and a negatively tailored strategy. The positively tailored strategies were based on persuasive features for which a motivation type had a large influence on their appreciation in the van Velsen et al. (2019) study. The negatively tailored strategies were based on persuasive features for which a motivation type had a small or no influence on their appreciation. For example, in the dialogue for the coaching strategy based on the 'Self Goal Setting'-feature the coach would propose a coaching approach in which they would help the participant to set a personal goal. The final selection of coaching strategies can be found in Table 3.2. We presented the Dual Motivation group with strategies suitable for both the intrinsic motivation and external regulation types. Since there were no significantly preferred persuasive features for the a-motivation type, but we

Table 3.1: The rules used for classifying participants into motivation groups based on the normalised scores (between 0-1) from their motivation profile. If the prerequisites for the first group (A-motivation) were not met, the prerequisites for the second group were checked, and so on.

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

wanted to include the A-motivation group for completeness, we present the A-motivation group with two random strategies. We selected these from the set of four strategies that were selected for the other three groups.

Table 3.2: The four motivation groups and the positively tailored and negatively tailored strategy presented for each group.

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

Coaching strategy dialogue creation

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

1. Four researchers (with experience in writing motivational content) each wrote a short dialogue in which one of the strategies was presented.
2. One of these four writers examined the resulting four strategy dialogues and created one general dialogue structure suitable for presenting the strategies to a participant. The four dialogues from step 1 were then all adjusted to fit this structure.

3. Each of the four original writers reviewed all four of the resulting dialogues to verify that they provided a good representation of the strategy and were understandable by the study participants).

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

3.3.4 Measurements

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

Demographics

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

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

Coach preference and coach likeability rating

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

Strategy preference and strategy appreciation rating

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

Usability

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

3.3.5 Data analysis

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

We started testing with strategy appreciation ratings after checking the distribution of these ratings for normality (which they were for the four presented strategies, the chosen strategy, and the not-chosen strategy). We then checked the randomisation of the two strategies that were presented to the participant by performing a paired-samples t-tests. There was no significant difference in rating between the scores for the first strategy ($M = 4.92$, $SD = 1.64$) and the scores for the second strategy ($M = 4.85$, $SD = 1.59$) ($t = .368$, $p = .714$).

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

To test if one strategy was chosen more often than the other by participants in the strategy preference question, Binomial tests were conducted. The distribution

of the number of choices for both strategies was compared to a 50-50 chance distribution.

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

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

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

3.3.6 Recruitment

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

3.3.7 Ethics

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

3.4 Results

3.4.1 Demographics

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

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

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

3.4.2 Usability

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

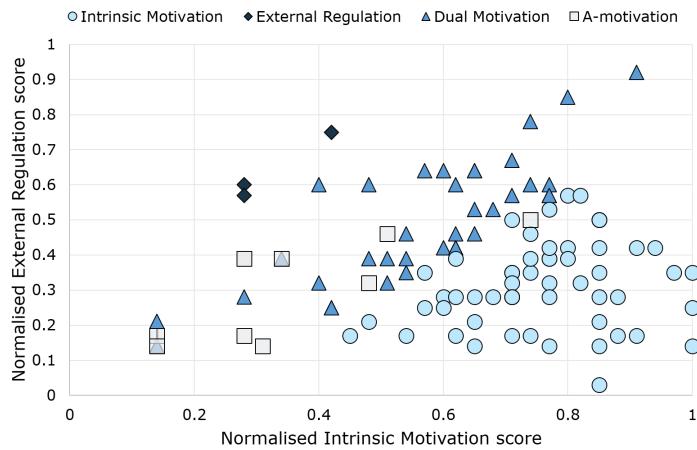


Figure 3.4: A scatter plot showing the normalised score for the intrinsic motivation and external regulation types per participant. The marker shape indicates the motivation group.

3.4.3 Motivation groups

Of the 108 participants, 56 (51.9%) were classified into the Intrinsic Motivation group, 3 (2.8%) into the External Regulation group, 39 (36.1%) into the Dual Motivation group and 10 (9.3%) into the A-motivation group. A scatterplot showing the normalised intrinsic motivation score and external regulation score for all participants can be found in Figure 3.4. As can be seen, there are few participants who solely have a high score on external regulation. Furthermore, there are a few participants in the A-motivation group who have relatively high scores on intrinsic motivation and external regulation. On the other hand, some participants in the Dual Motivation group have quite low scores on intrinsic motivation and external regulation.

3.4.4 Strategy preference

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

Strategy appreciation rating

We tested the appreciation ratings given to the positively tailored and negatively tailored strategies within the Intrinsic Motivation and Dual Motivation groups (see Table 3.3). No significant difference in appreciation rating was found between the two strategies presented to the Intrinsic Motivation group, with the positively tailored Self Goal Setting strategy receiving similar ratings as the negatively tailored Health Education strategy. There was however a significant difference between the appreciation rating for the positively tailored and the negatively tailored strategy within the Dual Motivation group. The positively tailored Showing Progress strategy was appreciated more by participants than the negatively tailored Implementation Intentions strategy.

Table 3.3: Results of the paired-samples t-test comparing the appreciation rating for the positively tailored strategy with the rating for the negatively tailored strategy for the Intrinsic Motivation and Dual Motivation groups.

Group	Positively tailored <i>M (SD)</i>	Negatively tailored <i>M (SD)</i>	t	p
Intrinsic	5.02 (1.58)	5.05 (1.58)	0.152	.880
Dual	4.97 (1.71)	4.28 (1.65)	-2.610	.013

When comparing the appreciation rating for the strategy that participants chose with the appreciation rating for the not chosen strategy, we found a significant difference between the appreciation rating for the strategy that was chosen and the appreciation rating of the not chosen strategy (see Table 3.4). This was also the case for the chosen and not chosen strategy within the Intrinsic Motivation group, and the chosen and not chosen strategy within the Dual Motivation group.

Table 3.4: Results of the paired-samples t-test comparing the appreciation rating for the chosen strategy with the rating for the not chosen strategy within all participants, and the Intrinsic Motivation and the Dual Motivation groups.

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

We did not look into the appreciation ratings given by the A-motivation group in much detail since the group contained just 10 participants and they were presented with two random strategies, but we shortly report the distributions of the appreciation ratings to provide some insight. These were the following: Self

Goal Setting was appreciated with a median of 5.00 ($N = 6$, $IQR = 4.75-6.25$); Health Education was appreciated with a median of 4.00 ($N = 5$, $IQR = 3.00-5.50$); Showing Progress was appreciated with a median of 5.00 ($N = 5$, $IQR = 4.50-6.50$); and Implementation Intentions was appreciated with a median of 5.50 ($N = 4$, $IQR = 5-6$).

Strategy choice

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

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

3.4.5 Coach preference

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

Coach likeability rating

Participants were asked to rate the likeability of the coaches before and after strategy presentation. The mean likeability ratings for all three coaches (Alexa, François, and Helen) can be found in Tables 3.5 and 3.6.

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

there was a significant difference between the ratings for François and Helen ($t = -5.482, p < .001$), and Alexa and Helen ($t = -3.698, p < .001$).

When comparing the ratings for the three coaches given after the strategies were presented, we found that they differed significantly ($F(2, 214) = 7.501, p = .001$). Post hoc tests revealed that there was no significant difference between the ratings for Alexa and Helen ($t = -1.254, p = .212$), but there was a significant difference between the ratings for Alexa and François ($t = 2.517, p = .013$), and François and Helen ($t = -3.734, p < .001$).

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

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

Table 3.5: Likeability rating of the coaches before strategy presentation (the External Regulation group was omitted since N was 3).

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

Table 3.6: Likeability rating of the coaches after strategy presentation (the External Regulation group was omitted since N was 3).

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

Coach choice

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

3.4.6 Strategies and coach influence

To assess a possible influence of presenting coach on strategy appreciation, we compared the appreciation ratings for the strategies presented by coach Alexa with the ratings for the same strategies when presented by coach François. This was done within the Intrinsic Motivation and Dual Motivation groups. The distribution of ratings and results can be found in Table 3.7.

Table 3.7: Results of the independent-samples t-test comparing the appreciation rating for the positively and negatively tailored strategies between participants when Alexa was presenting a strategy with the same strategy being presented by François for the two motivation groups.

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

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

For strategies presented to the Dual Motivation group, tests showed that there was no significant difference between the ratings given to the positively tailored Showing Progress strategy presented by Alexa and the ratings for the same

strategy presented by François. For the negatively tailored Implementation Intentions strategy the ratings when presented by Alexa and the ratings for the strategy when presented by François also did not differ significantly.

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

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

3.4.7 Strategies and demographics

To assess the influence of demographics on the appreciation of strategies we conducted a linear regression for the appreciation rating of the strategy that participants chose and appreciation rating of the strategy that they did not choose. We included the variables age, gender, education, self-reported physical activity, health literacy and living situation. We also conducted this test for each of our four strategies (Self Goal Setting and Health Education for the Intrinsic Motivation group, Showing Progress and Implementation Intentions for the Dual Motivation group). We only report significant results. The results (Table 3.8) show that older participants tend to appreciate their chosen strategy less than younger participants when observing the whole population and the Intrinsically Motivation group. In the Dual Motivation group, females tend to appreciate their chosen strategy more than males. Furthermore, older participants in the Intrinsic Motivation group tend to rate the tailored Self Goal Setting strategy lower than younger participants (Table 3.9).

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

Table 3.8: Results of regression analyses for the chosen strategy.

Group	Demographic	Beta	t (df)	p	R²
All	Age	-.36	-3.97 (106)	<.001	.129
Intrinsic	Age	-.43	-3.51 (54)	.001	.186
Dual	Gender	.38	2.50 (37)	.017	.145

Table 3.9: Results of regression analyses for the Self Goal Setting strategy (positively tailored strategy for Intrinsic Motivation).

Group	Demographic	Beta	t (df)	p	R²
Intrinsic	Age	-.31	-2.42 (54)	.019	0.098

3.5 Discussion

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

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

Furthermore, our experiment resulted in three secondary observations. First, the present study verifies conclusions from previous research that people can

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

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

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

Third and last, we found that there was an influence of participants' age and gender on the appreciation of the strategies. Namely, older participants rated their chosen strategy lower than younger participants in general, and female participants in the Dual Motivation group rated their chosen strategy higher than male participants in that group. These findings are in line with the trend towards more engagement from female participants that Perski et al. (2017) reported for

health coaching applications, but seem to be in contrast with the trend towards more engagement from older participants that they report. The effect found for gender also seems to be in contrast with the reported effect by de Vries et al. (2017) that male participants rated certain types of motivational messages higher than female participants. Combined with the fact that living situation and level of education were found to be relevant factors in the study that inspired our strategy design (van Velsen et al., 2019), the inclusion of user-related parameters remains relevant for future research.

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

3.5.1 Limitations

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

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

Furthermore, to thoroughly test the influence of a participant's impression of an

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

3.6 Conclusions

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

Intermezzo

Structuring coaching conversations

Towards dynamic tailoring during interaction

The results of the study presented in **Chapter 3** indicated that coaching strategies could be tailored in the context of coaching conversations. In addition, that study and the usability evaluation of the prototype also provided valuable insights into the inclusion of elements such as background stories and questionnaire items in dialogues. With the concept of the coaching engine in mind, the next step for the tailoring investigation seemed to be to define the building blocks that could be used to produce dialogues that followed a certain coaching strategy.

For the Council of Coaches project, the next step was to further specify the set of agents, their coaching domains and their content, which would then be evaluated in the third user study (documented in Council of Coaches Deliverable 2.6 (van der Kamp et al., 2019)). It was decided that there would be seven coaches and an assistant agent (see Figure I.1). Furthermore, during the process of defining and structuring the dialogue content for these agents (documented in Council of Coaches Deliverable 3.4 (Beinema et al., 2019)), it became clear that there were several shared dialogue topics between the different domains, such as the coach sharing a background story, or providing information on ‘why’ certain behaviour was recommended.



Figure I.1: Screenshot of a rare moment in the Council of Coaches living room when all of the coaches were present. From left to right: Carlos (peer coach), Olivia (physical activity), Emma (social coach), Katarzyna (type 2 diabetes coach), Helen (cognition coach), Rasmus (chronic pain coach), Coda (assistant), and François (nutrition coach).

Meanwhile, we created a corpus of interactions between multiple health care professionals and actors playing patients as part of the project (published as the *Patient Consultation Corpus* (Snaith et al., 2021)) to get an insight into how multi-human dialogues in a coaching context went. Using an example from that corpus, we investigated how a state-of-the-art dialogue game could be specified

for a goal-setting dialogue (Snaith et al., 2018). Furthermore, we investigated how coaching dialogues could be written to include specific behaviour change techniques (Das et al., 2019). These investigations helped clarify the scope for what would be considered a topic, and what the role of behaviour change literature could be in the content definition and tailoring process.

The experiences described above led to the idea that dialogue topics would be the higher level building blocks for a conversation, and that the strategy would influence which topics would be most relevant to discuss at a certain point in the interaction between coach and user – which was a change from the initial idea that the strategies themselves were templates that could be filled. Thus, a decision was made to investigate tailoring on a topic level. From a technical perspective, the concept of implementing tailoring using a tree of topics arose. A selection algorithm could then traverse that tree to select the next relevant topic to discuss at that point in the interaction. This approach would allow for the selection of relevant dialogues following a coaching strategy, but also for dynamically factoring in other knowledge about users, such as measured behaviour or users' responses in dialogues. A dialogue script could then be written for each topic in the functional demonstrator, but dialogues for topics might also be represented through dialogue games in our technical demonstrator. An important next step towards implementing tailored selection of topics was to define a model for the tree of topics, which is the focus of **Chapter 4**.

Chapter 4

What to discuss? – A blueprint topic model for health coaching dialogues with conversational agents

Based on:

Beinema, T., op den Akker, H., Hermens, H., van Velsen, L. What to discuss? --
A blueprint topic model for health coaching dialogues with conversational agents
(submitted for publication).

Abstract

Introduction: Conversational agents (CAs) are often included as virtual coaches in eHealth applications. Tailoring conversations with these coaches to the individual user can increase the effectiveness of the coaching. An improvement for this tailoring process could be to (automatically) tailor the conversation at the topic level. In this article, we describe the design and evaluation of a blueprint topic model for use in the implementation of such topic selection.

Methods: First, we constructed a topic model by extracting actions from the literature that a CA as coach could perform. We divided these actions in groups and labelled them with topics. We included literature from the behavioural psychology, relational agents and persuasive technology domains. Second, we evaluated this topic model through an online closed card sort study with health coaching experts.

Results: The constructed topic model contains 30 topics and 115 actions. Overall, the sorting of actions into topics was validated by the 11 experts participating in the card sort. Cards with actions that were sorted incorrectly mostly missed an immediacy indicator in their description (e.g., the difference between 'you could plan regular walks' as opposed to 'let's plan a walk') and/or were based on behaviour change techniques that were difficult to translate to a conversation.

Conclusion: The blueprint topic model presented in this article is an important step towards more intelligent virtual coaches. Future research should focus on the implementation of automatic topic selection. Furthermore, tailoring of coaching dialogues with CAs in multiple steps could be further investigated, for example, from the technical or user interaction perspective.

4.1 Introduction

In recent years, research and development in the area of health behaviour change applications has increased (Brinkman, 2016). Where initially, there has been a focus on providing health information via websites or telemedicine services, in later years the focus shifted to building interventions that could also actively send notifications and provide the user with insight into their personal situation (e.g., through the use of sensors and tailored feedback). A major challenge for these health behaviour change applications is user's adherence (Nijland, 2011; Wangberg et al., 2008). Users tend to be engaged with an application at first, but then stop using it or stop following its suggestions after the novelty effect wears off. Potential causes for this lack of adherence are actively being researched. It appears that contributing factors are a lack of direct involvement of a health care professional (*no social incentive*) and content that does not always fit the user's personal situation (*relevance of content*) (Andersson et al., 2009; Buimer et al., 2017). Two possible solutions investigated to tackle these issues are the use of *conversational agents* (CAs) and *tailoring* the application to the user (H. op den Akker et al., 2014).

As Starr (2008) states: "The coaching process can be considered as a series of conversations between two individuals – the coach and the coachee – for the benefit of the coachee in a way that relates to the coachee's learning process". In health behaviour change applications, conversational agents (CAs) can take on the role of a coach and can have coaching conversations with the user (the coachee) (Kramer et al., 2020). These CAs are '*computer systems that imitate natural conversation with human users through images and written or spoken language*' (Laranjo et al., 2018). Non-agent approaches (such as traditional websites or apps) tend to use a one-way method of communication. The possibility for two-way communication – i.e. interactive dialogues – with conversational agents provides a number of advantages. For example, an agent can ask the user questions about their interests, or elaborate on a topic per user request (Bickmore & Giorgino, 2006). While some of these functionalities might also be fulfilled by static text, an agent adds a social element and makes the process more dynamic and interactive. Furthermore, while human coaches may be the preferred way of coaching, CAs are always available, never grow tired of answering (the same) questions and can provide continuous support. These aspects allow for CAs to support users in their daily life by providing support after a diagnosis, but also coaching in preventative care.

4.1.1 Tailoring coaching conversations with CAs

CAs communicate with users through dialogues, where often the dialogue content is carefully designed by domain experts based on existing interventions (e.g., Callejas et al. (2014)). This process requires translation of intervention content to the dialogue domain. In addition, CAs also need to motivate the user to complete the objective of their interaction (Bickmore et al., 2010). From counselling literature, we know that the quality of a working alliance between a counsellor and client is a factor in the therapeutic change and adherence (Castonguay et al., 2006) and that it has three key aspects, namely goal agreement, task agreement and development of a personal bond. Ideally, CAs would build up such a working alliance with their users. However, this does require dialogues between CAs and users to not only contain coaching content, but social elements as well (Bickmore et al., 2005; Schulman & Bickmore, 2009). Furthermore, there should be variation between dialogues in terms of content and structure to keep participants engaged (Bickmore, Silliman et al., 2013). Finally, a key aspect that sets dialogues with health coaching CAs apart from, for example, chatbots for customer support is that their conversations continue over multiple interactions (Bickmore et al., 2018).

Tailoring has been shown to be effective in digital health applications (Krebs et al., 2010; Ryan et al., 2019; Wangberg et al., 2008) and has been researched in CA applications for things like automatic goal selection, switching topics on the sentence level (Glas & Pelachaud, 2018; Smith et al., 2011), and – in the case of agents with an embodiment – non-verbal behaviour adjustments (Krämer et al., 2010). Tailoring can be seen as the adjustment of a communication's timing, intention, content and representation to the user (H. op den Akker et al., 2014). Brinkman (2016) poses that such tailoring is a capability that an advanced eHealth system should be able to perform, that is, it should be able "*to select the most effective and acceptable treatment for the individual and tailor the treatment protocol to optimise potential conflicting values a person holds, e.g., autonomy versus safety.*" (Brinkman, 2016). Literature on conversational systems typically tends to distinguish various levels in a conversation, namely: *domains, topics* (of conversation as a whole and on an utterance level), *dialogue acts* and *utterances* (e.g., McTear et al. (2016, pp. 161–162)). We propose that tailoring a coaching conversation can happen at five levels, namely domain, topic (of conversation), action, dialogue act, and utterance (see Figure 4.1).

Once a domain (e.g., physical activity coaching) has been selected, a discussion topic can be personally decided upon, by using the available information (e.g., the user's profile, their interaction history, available sensor data, and domain knowledge). Tailoring of topics allows for the CA to take the initiative by suggesting a topic to discuss and can provide users with conversations that are relevant and suitable for their specific situation. Selection of topics can also be

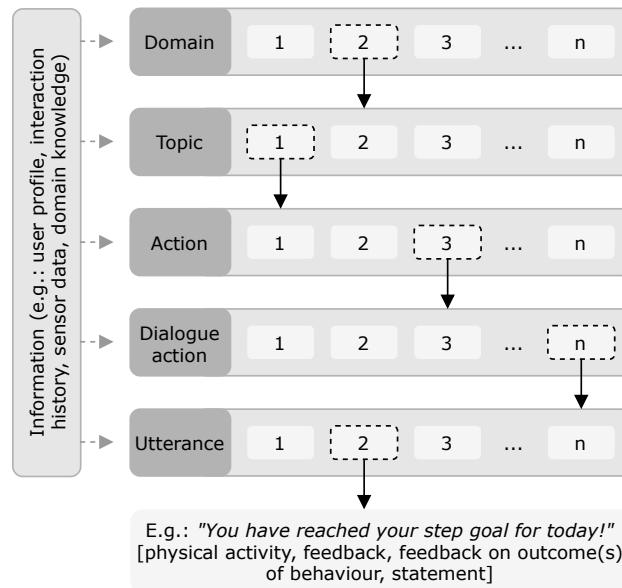


Figure 4.1: Different levels on which coaching conversations can be tailored.

4

influenced by the goals and strategies that are set, either by the system or by the user, or shared. If the CA follows a health education strategy, topics that fit that strategy could be emphasised, such as providing information about healthy behaviour (Zhang & Bickmore, 2018). For an implementation intentions strategy, the action planning topic might be more relevant. Once a conversational topic has been chosen, the actions that are selected to execute the topic can be adapted to the user as well. Finally, once tailored actions are selected, the execution of these actions through (multiple) dialogue acts and utterance selection can, in turn, be tailored too. The combination of all approaches can potentially increase adherence and engagement, by providing users with relevant tools and information only.

However, CAs across application domains tend to apply a one-size-fits-all approach when it comes to their responses (Følstad & Brandtzaeg, 2017). When looking at the literature on health coaching, there are a number of examples where tailoring in coaching conversations with CAs is performed at the lower levels – with participants all being provided with a predefined order of topics that are discussed (e.g., Zhang and Bickmore (2018)), but there are no examples that tailor the higher level content in the manner that we described above. There are some cases in which tailoring of the higher levels is performed by a health care professional who assigns personalised approaches to participants in the system's backend (e.g., Abdullah et al. (2018), Benítez-Guijarro et al. (2019) and Fadhil

et al. (2019)). While this method of personalisation has positive results, the involvement of a health care professional makes it labour-intensive. Bickmore et al. (2011) describe an ontology and task model that are developed to make task execution of health coaching dialogues more modular, dynamic and reusable, but they do not provide such automation for the higher level topics.

4.1.2 Objectives

To automatically tailor health coaching conversations with CAs on a topic level, we need a blueprint topic model that can be used as the basis for a topic model in a practical implementation. That is, researchers can use it to select and structure topics that they would like to include in their agent applications and can extend it with new topics if that is desired for specific domains or coaching approaches.

In this article we focus on the development of that blueprint topic model. We structure the topics in our model as a hierarchical tree, so that in a future implementation a topic selection *algorithm* can work from the tree's root and can chose the most relevant subtopic at each split, until a topic has been reached that has no further subtopics. This topic can then be returned as the topic that should be discussed in the conversation that will be started. We therefore address the following research question:

RQ: *Which topics should be included in a blueprint topic model for health coaching conversations with CAs, and how should they be structured?*

In order to answer this question, we first present a background and review of related research to illustrate the context of this work. We will then report on the two steps we took towards the resulting topic model and its practical implementation in CA systems. First, we investigated which topics are relevant to include in the blueprint topic model based on literature. Second, we evaluated the model that resulted from the first stage by performing a card sort study with experts. We conclude this paper with a discussion of the process, our findings, potential implementation and suggestions for future research.

4.2 Background and related research

Health behaviour change applications with CAs have been researched in the past fifteen years as a method to assist people in obtaining a healthy lifestyle. These agents offer assistance on one or more domains, such as physical activity, nutrition or well-being. In some cases, agents are just a component of a broader eHealth application (e.g., ter Stal et al. (2020) and van Velsen et al. (2020)),

while in other cases they are the intervention's main component (e.g., H. op den Akker et al. (2018) and Sebastian and Richards (2017)). In all cases, agents are designed to communicate with the user and in this manner hopefully increase application usage and potential effect.

Coaching sessions with CAs in health behaviour change systems generally follow a certain structure. While there are, to our knowledge, no topic models for coaching, there are papers that discuss the structure of a coaching session with a virtual coach as a series of phases. For example, de Kok et al. (2014) present a structure for coaching with a virtual agent during squat exercise sessions. Such a session consists of an introduction, initial assessment, coaching cycle and closing. The coaching cycle itself is a series of explanation, demonstration, instruction, performance assessment and provision of feedback. For insomnia therapy, Beun et al. (2014) describe three phases for the long-term, which are an opening phase, an intervention phase, and a closure phase. The opening phase involves an introduction between coach and coachee, introduction to the therapy, inclusion/exclusion advice, and planning and committing. The intervention phase involves four exercise steps, which consist of an introduction, plan and commit, task execution and evaluation. The closing phase involves closing the therapy.

In addition to interventions having a certain general structure, a number of researchers also motivate the combined use of both a CA and assistive tools. Beun et al. (2017) motivate the combined use of both an agent and tools such as diaries and graphs for self-monitoring in a mobile application. Another example is the agent designed by Bickmore, Silliman et al. (2013) who augmented the dialogues that could be held with their agent with various images (e.g., characters demonstrating exercises and proper pedometer use) and 'dynamically generated self-monitoring chart showing the participant's step counts relative to goals over time'.

While we have found no instances of automatic topic selection as is the subject of this article, there are some examples of tailoring and automatic dialogue or content generation in health behaviour change applications with CAs that are relevant to discuss. For example, the application developed by Fitrianie et al. (2015), which follows the session structure described by Beun et al. (2014), automatically generates the dialogue for the subtopics. In the paper by Beun et al. (2017) a more general approach to instantiating the dialogues using persuasive features is reported. The manner in which Fitrianie et al. (2015) structure the dialogue using interaction recipes, that refer to scripted dialogue actions, allows them to do this in a mobile phone CA application, without the natural language generation process becoming too computationally heavy for the phone. This approach is similar to the use of the task model described by Bickmore et al. (2011), but where Beun et al. (2016) focus on incorporating persuasive features, Bickmore et al. (2011) focus on behaviour change models

and techniques. Another virtual coach discusses diet one day and physical activity the other day (Bickmore, Schulman & Sidner, 2013), which can be seen as an example of a holistic coaching approach that addresses multiple domains. Both the approaches by Fitrianie et al. (2015) and Bickmore et al. (2011) seem to focus on automatically generating the content to instantiate scripted sets of actions. Montenegro et al. (2019) on the other hand, define a dialogue act taxonomy for a virtual coach, which they use to classify the dialogue acts in coaching dialogues. They do define a hierarchical model for topics, but this is a classification that contains topics such as 'Sport and Leisure' with subtopics 'Demotivation' (subtopics 'Free time', 'Loneliness', 'Fear'), 'Hobbies' and 'Sport', which have no coaching aspect to them. That is, they do not include intents to, for example, inform or give feedback. Similarly, their 'Intent' model covers relevant intents, but these are aimed at the level of dialogue actions.

Examples of tailoring communication to users on a more specific level include the generation of brief interventions for excessive alcohol consumption using Markov decision processes and reinforcement learning (Yasavur et al., 2013). Smith, Cavazza et al. (2008) include a recommender system that is used in conversations with an agent that suggests and discusses the activities for the day with its user. Gupta et al. (2018) worked on the extraction of health goals and topic boundary detection as research towards building a coach that can have SMS conversations. Beinema, op den Akker et al. (2021) investigated the adjustment of coaching strategies to users' motivation to live healthy in a CA application, and Gross et al. (2021) investigated personalisation of interaction styles (e.g., deliberative or paternalistic) for CAs in chronic disease management. Furthermore, even though this is a one-way manner of communication mostly aimed at an utterance level, the process of tailoring messages that are sent to the user has been researched in relationship with users' stage of change (de Vries et al., 2016; Uribe et al., 2011), stage of change and personality and gender (de Vries et al., 2017), activity and self-efficacy (Achterkamp et al., 2013), score on the susceptibility to persuasion scale (Kaptein et al., 2012), and the state of the COMBI model (Klein et al., 2013). Different tailoring techniques can be used for this purpose, as described in H. op den Akker et al. (2014). While not all of these approaches were developed for dialogue-based interactions specifically, we do feel that these tailoring principles can be applied when tailoring the dialogue generation and execution phase that comes after the topic selection.

4.3 Methods: Topic model construction

The intervention structures discussed in the background section already provide some insight into which topics are prevalent in health behaviour change applications. In this section, we will elaborate on our methods for constructing a topic

model for use in a coaching conversation with a conversational coach.

We started the process of constructing a blueprint topic model by clearly defining the concepts for the higher levels of tailoring coaching conversations. What do we understand to be a topic? What do we mean with actions? How are they related to the concept of tailoring (as, for example, defined for motivational messages)?

Following the definition phase, we extracted a set of actions from literature. Each of these actions was given a name and description. We took a multidisciplinary approach, looking at literature from the behavioural psychology, relational agents and persuasive technology fields.

Once we had extracted the set of actions from literature, we began by splitting this set into groups of actions that contributed to the same general topic of conversation (following the concept of a hierarchical tree of topics). We then defined a topic name to each of these groups. This resulted in a topic model, with actions being grouped as contributing to the discussion of a shared topic.

4.4 Results: Topic model construction

4.4.1 Defining concepts

The first step in building the topic model was to define what we understood a topic and action to be and how this related to the existing literature on tailoring. In this article, we use the term *topic*, following the definition for a ‘discourse topic’ by Riou (2015) (which is on a higher level than the topic of a sentence). The term topic is used frequently in literature, but as Riou (2015) states, it is rarely defined beyond an intuitive understanding and interpretations vary. She states that a topic is what a portion of the interaction is about and that it must be the center of shared attention. Furthermore, a topic is participant- and interaction-specific; it is jointly determined during an interaction by its participants.

An *action* is something a coach can do or say during the coaching process. For example, during a conversation on how a user could be more active, a coach could suggest to add a specific activity to their daily schedule (e.g., go on lunch walks), but they could also suggest to replace one activity with a healthier option (e.g., take the stairs instead of the elevator). While these are different actions, they both contribute to a conversation on how someone could be more active. Inclusion of actions in a coaching conversation can require multiple dialogue moves to be executed, but a dialogue that discusses a specific topic might include a combination of actions.

4.4.2 Extracting actions from literature

The next step was to extract a set of actions from the literature. As stated in the method, the literature we included in our process came from the following three research fields:

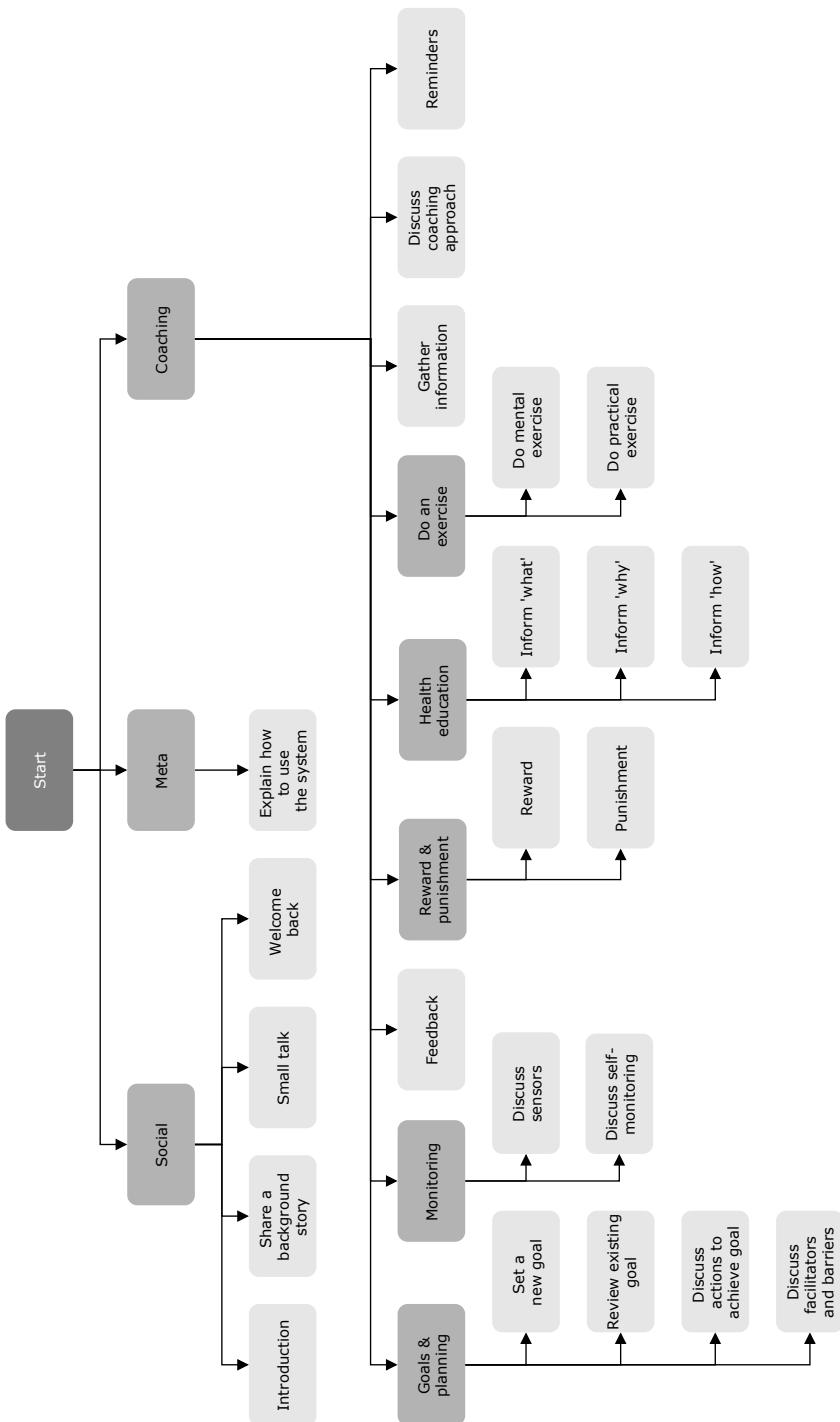
- The behavioural psychology field. The behaviour change technique taxonomy (BCT Taxonomy) by Michie et al. (2013) provides an overview of 'techniques a coach may apply', when it comes to human-human coaching. We therefore included all 93 techniques as actions in our set of actions.
- The relational agents field. Research on relational agents teaches us that social actions are important to include if we want a relationship to develop between a coach (the CA) and a user. We therefore defined actions that contribute to this social aspect and included them in our set of actions.
- The persuasive technology field. We defined additional actions by reasoning about the actions that are involved when executing persuasive strategies. Examples include the persuasive principles listed by Fogg (2002), which are included in the Persuasive Systems Design model (Oinas-Kukkonen & Harjumaa, 2009). We also include a listing of often used persuasive features in the article by van Velsen et al. (2019), which has a specific focus on health coaching.

All three types of literature provided actions that we feel can be important to include when building a health coaching agent. First, the BCT Taxonomy focuses on the core techniques for coaching, but its origin in human-human interaction means that social processes are not addressed. The literature on relational agents allowed us to include important actions for the discussion of social topics. Furthermore, the inclusion of actions based on persuasive technology literature allowed us to include actions that are relevant to build a persuasive agent application in an eHealth context. Finally, we also defined some basic actions that are relevant for human-computer interaction, such as 'explain how to interact with the system' and 'assist with sensor connection'.

The resulting set of actions included 115 items. The full list of all 93 behaviour change techniques can be found in the article by Michie et al. (2013). The list of 22 additional actions that we defined can be found in Appendix B.

4.4.3 The topic model

As described in Section 4.3, we constructed a topic model from the set of actions we extracted from literature by splitting the set of actions into groups that seemed to contribute to the same dialogue and defining a topic to cover each



4

Figure 4.2: The full topic model on which the setup of the card sort pilot was based.

subset. An overview of the resulting topic model can be found in Figure 4.2. An overview of topic descriptions and contributing actions can be found in Appendix C.

In the following subsections we will discuss the topics in the topic model step by step¹.

Split one: the social, meta, and coaching topics

The first distinction that we make is a split between *social* and *coaching* topics (see Figure 4.3). We know that if we want to build a working alliance, goal agreement, task agreement and development of a personal bond are important (Castonguay et al., 2006; Horvath & Greenberg, 1989). From literature on relational and social agents, we also know that it is important to include social behaviours in dialogues (Bickmore et al., 2005).

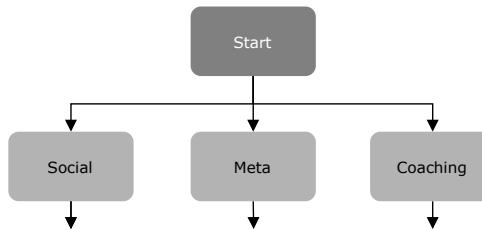


Figure 4.3: The first split in topics.

We also include a *meta* topic at this level, which contains actions in which the agent explains the application (their application) and how to interact with it (and the agent), as such topics can neither be considered *social*, nor *coaching*.

Split two: social subtopics

Social communication between an agent and a user is suggested to contribute to the trust and working alliance between an agent and the user (Bickmore, 2010). Oinas-Kukkonen and Harjumaa (2009) also list principles such as 'similarity', 'liking' and 'social role' as important for dialogue support. In their papers on establishing the computer-human working alliance, Bickmore et al. (2005) provide

¹Please note that we discuss the topic model that resulted from the first study and that some topics were moved based on the pilot in the evaluation study. The final model can be found in Figure 4.16. The listing of topics in Appendix C corresponds to that final model.

a number of specific examples of verbal relation behaviours. These verbal relation behaviours are: Expressing empathy for a user, social dialogue, reciprocal self-disclosure, humor, meta-relational communication (talk about the relationship), expressing happiness to see the user, talking about the past and future together, continuity behaviours, and reference to mutual knowledge. Similar behaviours are named by Kowatsch et al. (2018).

Bickmore and Picard (2005) explain that they focus on behaviours for relational agents that can be employed by a computer and provide literature that motivates their effects in human-human interaction. Schulman and Bickmore (2009) state that social dialogue may only be relevant in long-term interventions, emphasizing to carefully consider the difference between short-term compliance and long-term adherence. While some of the behaviours listed in these papers are typically intertwined in dialogues that are primarily about another topic (e.g., the use of humor), others can be seen as full dialogues (e.g., getting acquainted).

We split the four actions that we extracted from this literature, and we define four subtopics for social dialogues that cover these four actions contributing to 'social' conversations: *introduction*, *share a background story*, *small talk*, and *welcome back* (see Figure 4.4).

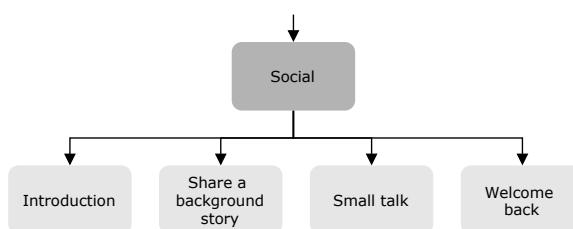


Figure 4.4: The second split in topics, different subtopics for the *social* topic.

Introduction. A topic that covers the 'getting acquainted talk' (Bickmore & Picard, 2005) action. A typical example would be an introduction between user and agent when they meet for the first time. The *introduction* topic would be most relevant at the start of the relationship between agent and user. Furthermore, an introduction is also a starting point for the information that the coach collects about the user, which can then be used for tailoring the topic selection or lower level personalization (e.g., using the user's name in a sentence).

Share a background story. Discussing the *share a background story* topic could result in dialogues in which the agent shares a background story about themselves with the user. This type of dialogue is intended to help the process of reciprocal self-disclosure (Bickmore et al., 2005) and it can provide a short break in the

discussion of coaching content. Furthermore, these types of dialogues could contribute to a user's engagement by, for example, sharing increasingly more personal parts of the coach's story, thus keeping the user curious to learn more.

Small talk. A topic that covers the 'social dialogue' action. Typical dialogues could involve the discussion of general small talk topics, such as the weather. While not of obvious relevance to the coaching process, answers given by the user in these conversations might provide input for tailoring the system's content to the user, and these dialogue can provide a welcome break in otherwise serious coaching conversations. Furthermore, sometimes interacting with an application is enough to build a social bond (following the computers as social actors (CASA) paradigm (Nass et al., 1994)) even when not performing a task (Bickmore & Picard, 2005).

Welcome back. This topic covers an often short but important dialogue between coach and user in which the user is welcomed back to the application (Bickmore & Picard, 2005). As discussed by Bickmore and Picard (2005) for the 'good morning' sentence, the sentence itself has lost much of its semantic meaning, but whether you say it and how you say can influence the development of a relationship.

Split three: a meta subtopic

The meta topic is a topic that covers actions that lead to conversations about the application. We defined one² subtopic for the meta topic, namely *explain how to use the system* (see Figure 4.5).

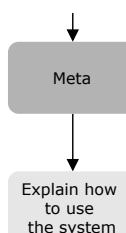


Figure 4.5: The subtopic for the *meta* topic.

²Please note that following the pilot card sort study, we added the *assist with sensor connection* and *discuss coaching approach* subtopics, which were originally a part of 'discuss sensors' and a separate topic.

Explain how to use the system. A topic that covers the ‘explain interaction paradigm’ and ‘explain user interface’ actions. Where in human-human interaction it is rarely necessary for someone to explain that you can talk to them to communicate, in an agent application a user might need some explanation for the use of reply-buttons or where to find certain features of the application (profile page, etc.).

Split four: coaching subtopics

The next split we made is a split in subtopics for coaching (see Figure 4.6).

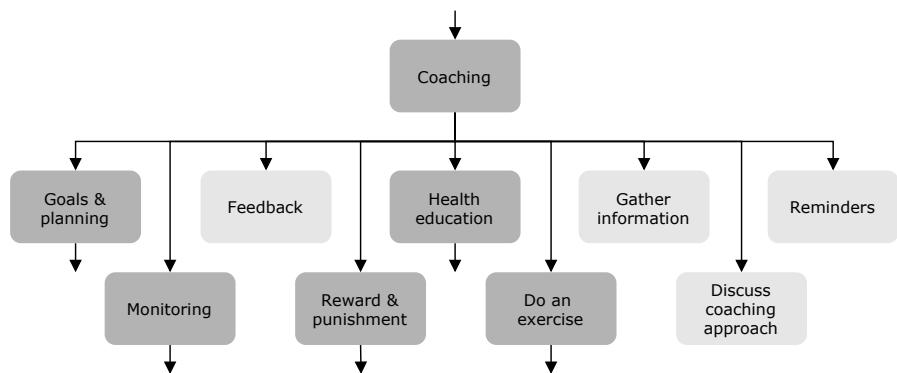


Figure 4.6: The subtopics for the *coaching* topic.

Michie et al. (2013) define 93 behaviour change techniques grouped in 16 categories in their taxonomy. These techniques and their categories provided us with a basis for defining the topics for dialogues that a coach might have when it comes to coaching. We also took frequently defined persuasive features in eHealth applications (van Velsen et al., 2019) into account when defining these topics (e.g., health education).

Goals & planning. The first subtopic is defined for the discussion of goals and planning. In turn, this topic has four subtopics in itself: *set a new goal*, *review existing goal*, *discuss actions to achieve goal* and *discuss facilitators and barriers* (see Figure 4.7).

One of the important elements in changing behaviour is having a goal to work towards. In their Goal-Setting Theory, Locke and Latham (2002) describe that the difficulty and specificity of a goal are the two core factors that influence performance. They describe that four mechanisms influence performance towards



Figure 4.7: The subtopics for the *goals & planning* topic.

a goal: people will be focused on the goal and direct effort towards it, goals have an energizing function, persistence is higher for more difficult goals, and people will act towards it and use task-relevant knowledge and strategies.

Goals and planning is also the first category that Michie et al. (2013) define for their taxonomy. The category features behaviour change techniques for setting goals in terms of behaviour or outcomes of behaviour. It also features techniques regarding reviewing those goals. Within the persuasive technology field, automatic goal setting and self-goal setting are persuasive features that are often used, as well as implementation intentions, for which planning is a key step (van Velsen et al., 2019).

Set a new goal. A topic that covers actions for setting a new goal. There are two typical dialogues for setting a new goal. A long-term goal can be set as a dot on the horizon, while setting short-term goals can help with working towards the long-term goal in smaller and more achievable steps (the reduction principle (Oinas-Kukkonen & Harjumaa, 2009)). These dialogues can include a shared decision-making process or a negotiation (Snaith et al., 2018). Even if the goal is set automatically by the system, the system can still explicitly ask the user if they accept it (R. op den Akker et al., 2016).

Review existing goal. A topic that covers actions for reviewing an existing goal. In a typical dialogue the agent would discuss with the user if the existing goal (long-term or short-term) should be adjusted, and they would adjust the goal if so (e.g., as mentioned by King et al. (2017)). Alternatively, the user can indicate they would like to adjust their goal.

Discuss actions to achieve goal. A topic covering actions that help the user to take action. In a typical dialogue the coach helps the user to plan actions for achieving their goal, for example by discussing today's plan (Smith, Charlton et al., 2008), co-creating of a weekly activity plan (King et al., 2017), or agreeing to plan a new appointment and that the user will work towards the goal in the

meantime (King et al., 2017).

Discuss facilitators and barriers. A topic that covers actions that can help a user overcome barriers or leverage facilitators. In a typical dialogue, a coach could discuss potential problems or opportunities with the user, and they could argue against any self-doubts the user might have (e.g., Bickmore et al. (2005), Bickmore, Silliman et al. (2013), Gardiner et al. (2017), King et al. (2017) and Watson et al. (2012)).

Monitoring. Monitoring is an important aspect for tailored coaching and it is strongly connected with another important aspect, namely feedback. For monitoring, we distinguish two subtopics: *discuss sensors* and *discuss self-monitoring* (see Figure 4.8).

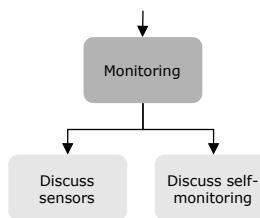


Figure 4.8: The subtopics for the *monitoring* topic.

Michie et al. (2013) defined 'feedback and monitoring' as the second category in their taxonomy and distinguish between self-monitoring and monitoring by others. They also distinguish among monitoring of behaviour and behaviour outcomes, and monitoring with and without feedback. Monitoring is also a key element in persuasive technology, since systems need the information about a user's behaviour to implement persuasive features such as showing progress, social competition, automatic goal setting and rewards (either through compliments or monetary).

Discuss sensors. A topic that covers actions to do with monitoring through sensors. It also covers the 'explain use for sensors' action. In a typical dialogue, the user will get an explanation about the sensor and will then start using the sensor (e.g., for a pedometer (King et al., 2017)). Users can also receive help with connecting the sensor to the system if needed (the 'connect sensor', 'provide assistance with sensor problems', and 'confirm measured values' actions)³.

³Please note that actions covered by the *assist with sensor connection* subtopic were moved to *meta* after the pilot of the card sort study.

Discuss self-monitoring. A topic that covers actions that relate to the user monitoring themselves, e.g., by using a (digital) diary. It also covers the 'explain use for self-monitoring' action.

Feedback. Being able to provide feedback to the user is an important skill for a coach (see Figure 4.9). Feedback can be given verbally, but the coach could also show a graph of the measured behaviour.



Figure 4.9: The feedback topic (no further subtopics).

As mentioned under monitoring, 'feedback and monitoring' is the second category that Michie et al. (2013) define in their taxonomy. Furthermore, persuasive features such as showing progress and rewards are forms of feedback, and features such as social competition can also be seen as a form of feedback.

Reward & punishment. The reward & punishment topic covers actions that have to do with consequences for the user's behaviour. We define two subtopics for the reward & punishment topic, namely: *reward* and *punishment* (see Figure 4.10).

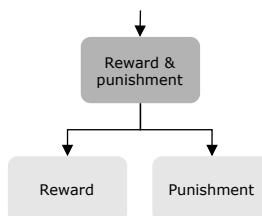


Figure 4.10: The subtopics for the *reward & punishment* topic.

Reward and punishment are key elements of operant conditioning (Miltenberger, 2008), and are essential in human behaviour change as external factors that have an influence on people's motivation. For example, the Health Belief Model states that the perceived benefits of an action and perceived threat influence a person's likelihood of taking action (Janz & Becker, 1984, p. 4) and Protection Motivation Theory (Norman et al., 2005) and the Health Action Process Approach (Schwarzer et al., 2011) model threat appraisal and risk perception to have a similar effect. Michie et al. (2013) define a 'reward and threat' and a 'scheduled consequences' category in their taxonomy, which contain BCTs related to incentives, rewards and punishment; and the 'reward' persuasive prin-

ciple (Oinas-Kukkonen & Harjumaa, 2009) is reflected in often used persuasive features such as compliments or monetary rewards (van Velsen et al., 2019). A specific example included in the agent literature is shaping (Bickmore et al., 2005) (Watson et al., 2012).

Reward. A topic that covers actions that have to do with rewarding the user and discussing the terms for rewards.

Punishment. A topic that covers actions that have a punishing effect on the user and discussing the terms for punishment.

Health education. This topic covers actions through which the coach educates the user about healthy behaviour in the coach's domain. We define three subtopics for health education, namely: *inform 'what'*, *inform 'why'* and *inform 'how'* (see Figure 4.11).

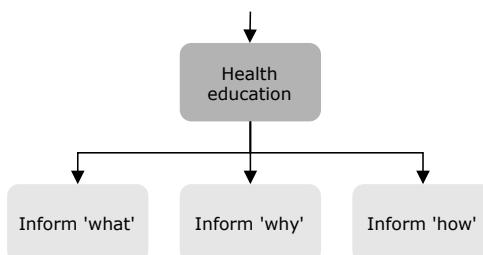


Figure 4.11: The subtopics for the *health education* topic.

Health education is an educative feature often applied in eHealth applications (van Velsen et al., 2019). It is also a topic that is covered in the 'shaping knowledge' category of the BCT taxonomy by Michie et al. (2013) and is relevant for some techniques in the 'natural consequences', 'comparison of behaviour', and 'antecedents' categories. Health education is important for ensuring that the user is informed about their health problem and the options available to them (King et al., 2017; Zhang & Bickmore, 2018).

Inform 'what'. A topic that covers actions that inform on what healthy behaviour is. A typical dialogue provides information based on domain standards and guidelines, but could also involve a short quiz to be more interactive and test the user's knowledge.

Inform 'why'. Another important aspect of health education is informing why

healthy behaviour in the domain is important. This can involve fear appeals or listing benefits. The design of inform ‘why’ dialogues could benefit from the work of Chalaguine et al. (2019) who list six types of arguments for health coaching with chatbots and investigate their effects.

Inform ‘how’. Actions that inform on how to perform or facilitate healthy behaviour are covered by this topic. A typical dialogue would involve suggestions or clear instructions, but other options are to give a demonstration (Jofré et al., 2017) or a tip of the day (Bickmore, Silliman et al., 2013; Gardiner et al., 2017).

Do an exercise. A coach could let the user perform exercises to, for example, practice behaviour or gather new insights. In a typical dialogue, the coach would explain an exercise to the user and lets the user do that exercise. Depending on the system, the coach could potentially provide suggestions or feedback during execution. We distinguish *do mental exercise* or *do practical exercise* (see Figure 4.12).

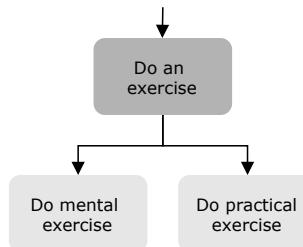


Figure 4.12: The subtopics for the *do an exercise* topic.

Do mental exercise. A topic that covers actions related to the coach instructing the user to perform mental exercise. Examples could be to have dialogues for exercises aimed at enhancing motivation and using a change ruler (Olafsson et al., 2019)) or making a list of pros and cons for changing behaviour in the coach’s domain.

Do practical exercise. Actions related to the coach instructing the user to perform practical exercises are covered by this topic. An example of a typical dialogue could be to introduce and explain a physical activity exercise (Ruttkay & van Welbergen, 2008).

Gather information. Gathering additional information about the user that can be used for tailoring the provided coaching can be an important topic for a virtual

coach (see Figure 4.13).



Figure 4.13: The *gather information* topic (no further subtopics).

A typical dialogue could involve questions on previous experience with an exercise (de Kok et al., 2014), questionnaires turned into dialogues (e.g., to determine stage of change, motivation to live healthy (van Velsen et al., 2019), frailty parameters (ter Stal et al., 2020), or readiness to change (Olafsson et al., 2019)), questions about the user's living environment that might be relevant for activity suggestions (e.g., 'Do you live near a park or forest?'). Furthermore, this dialogue can be used to check for important health changes (King et al., 2017) or to perform well-being checks (Bickmore, Silliman et al., 2013). When discussed regularly between coach and user, it can also be used for monitoring user's general physical and mental state (Ruttkay & van Welbergen, 2008); essentially using the coach as a sensor.

4

Discuss coaching approach. Another topic that can be important for tailoring coaching conversations is this topic that covers actions aimed at explaining and discussing how the user will be coached⁴ (see Figure 4.14).



Figure 4.14: The *discuss coaching approach* topic (no further subtopics).

A typical dialogue could be a dialogue in which a coach would explain that they can coach the user with a strict or understanding tone of voice. Depending on the complexity of the system, the user could then indicate what their preference is. As with the gather information topic, this topic is another example of explicit personalization as defined by Fan and Poole (2006).

Reminders. A topic that covers actions related to reminding the user to perform behaviour, take measurements, or to use the system (see Figure 4.15).

The potential 24/7 availability of agents as coaches makes that they can also support their user through reminders and following its status as one of the

⁴Please note that this topic was moved to *meta* following the results of the pilot for the card sort study.



Figure 4.15: The *reminders* topic (no further subtopics).

persuasive principles (Oinas-Kukkonen & Harjumaa, 2009) 'reminders' is an often-included persuasive feature in mobile apps. Reminders could be used for compliance with guidelines/goals (Benítez-Guijarro et al., 2019), taking a break (Bickmore et al., 2008), or starting a coaching session (Fitrianie et al., 2015).

4.5 Methods: Card sort study

To evaluate the topic model that was constructed in the first stage, as a second step, a closed card sort study with experts was conducted to verify the classification of actions as relevant for the discussion of the topics that were defined.

4.5.1 Design

In the closed card sort study, experts were asked to sort cards into predefined categories using an online card sorting tool⁵. Each category in the card sort represented a topic. Each card included the name and description of an action (as defined and extracted from literature in the first study). Since our topic model has a hierarchy, we let the experts perform multiple card sorts. Every round of the card sort represented a split (or 'layer') in the topic model (see Figure 4.2). The first round of the card sort started at the top of the topic model with the split into 'social', 'meta' and 'coaching'. Participants were therefore asked to sort all cards (with actions) into those three categories. The following rounds each covered further splits into subtopics. For each round, cards were included with the actions that contributed to the discussion of the topics that were included (or their subtopics). In this manner, eight rounds were defined. An overview of round numbers, number of cards (actions), topics (categories), and category labels (topic names) can be found in Table 4.1.

⁵We used the Proven By Users tool: www.provenbyusers.com

Table 4.1: Number of cards and categories per round for the initial set-up of the card sort study as tested in the pilot. Please note that adjustments were made to the setup after the pilot. The final setup can be found in Table 4.3.

Round	N cards	N categories	Category labels
1	115	3	Social, meta, coaching
2	109	9	Goals & planning, monitoring, feedback, reward & punishment, health education, do an exercise, gather information, discuss coaching approach, reminders
3	4	4	Introduction, share a background story, small talk, welcome back
4	10	4	Set a new goal, review existing goal, discuss actions to achieve goal, discuss facilitators and barriers
5	9	2	Discuss sensors, discuss self-monitoring
6	15	2	Reward, punishment
7	31	3	Inform 'what', inform 'why', inform 'how'
8	19	2	Do mental exercise, do practical exercise

4.5.2 Participants and recruitment

The target group for the study were experts on coaching, conversational agents and eHealth. Potential participants with the relevant expertise were identified through personal connections and recommendations, and sent a recruitment email. This email contained short descriptions of the general research aim, the card sort principle and an estimate for the time it would take to participate.

4.5.3 Procedure and collected data

If participants agreed to participate, they received an email with the details for participating in the study. This email started with an introductory text with a task description, and two links to the privacy statements of the online tool and our research facility. Furthermore, it explained that they should sort the cards as they thought was right and that cards were not necessarily equally divided over groups. Then, the email asked them to fill in a participant code for each of the rounds (with a short note explaining that this would allow us to anonymously connect rounds to the same participant) and provided links to the different rounds of the card sort.

When entering the first card sort round, participants were asked to complete an

informed consent form that explained that they could stop their participation whenever they wanted without providing a reason, and that the anonymized data would be used for scientific publications. This was followed by three statements to measure their self-perceived expertise on eHealth, coaching, and conversational agents ('I am an expert on [...]') with a five-point Likert scale ranging from 'Completely disagree' to 'Completely agree'.

After the intake step, participants were shown the first round of the card sort with a short instruction (to drag cards from the list on the left to the categories on the right). Upon completing each card sort round, they were shown a 'thank you'-message with a button that would refer them to the next round. The next round then started again with entering the participant code, a short welcome message, and a short instruction. During all rounds of the card sort, participants had the possibility to save and continue at a later moment. For each round of the card sort, the completion time and final sorting of the cards were stored.

4.5.4 Data analysis

In each round of the card sort, each participant in the card sort study performed a multi-class classification task (assigning one label to each card). In order to analyze this, we first created confusion matrices for all participants individually. Second, using our original classification of actions as the ground truth, we computed the overall percentage of correctly sorted cards per participant (which in this case is equal to the micro F_1 -score and accuracy). Third, we computed agreement between participants using Krippendorff's Alpha. Fourth, we computed the class-wise F_1 -score for all categories in a round (per participant). Finally, taking the agreement and class-wise F_1 -scores into account, we looked at the confusion matrices and the raw classification data in order to investigate which cards were classified differently from our labeling and what could be the underlying reasons for those classifications.

4.6 Results: Card sort study

4.6.1 Pilot

Prior to the card sort study, a pilot sort was conducted with two experts in June of 2020 to test the study's procedure and setup (e.g., for clarity of instructions, and to test the cognitive load). The experts were asked to perform the sorts for the first two rounds of the card sort.

The procedure in general was clear, but the cognitive load for completing the two card sort rounds was too high. Furthermore, it was suggested to add a

description to the topics. The first round took them 47 minutes on average and the second round 45 minutes. For the first round, the experts classified 95.22% (on average) of the cards into the 'correct' group. They agreed on the classification of 103 cards, and did not agree on 12. The average precision and recall values can be found in Table 4.2. As can be seen, the recall is high for all categories, while the precision of the 'meta' category is quite low. This is because the experts sorted a relatively high number of cards from the 'coaching' category (4 for Expert 1 and 13 for Expert 2) into the 'meta' category as opposed to the 2 cards that were supposed to be labeled as 'meta'.

Table 4.2: The precision and recall values for both experts for the three categories in the first round of the pilot study.

Group	N cards	Expert 1		Expert 2	
		Precision	Recall	Precision	Recall
Social	4	1.00	1.00	0.50	1.00
Meta	2	0.33	1.00	0.13	1.00
Coaching	109	1.00	0.96	1.00	0.84

4.6.2 Adjustments to study setup and model

Since the number of correct classifications and the agreement was high for the first round in the pilot, it was decided to remove this round from the card sort to lower the cognitive load and time to complete. Cards on which the experts did not agree or that they classified incorrectly were discussed between the first author and the experts. This lead to a reclassification of the 'discuss coaching approach' topic as a subtopic for 'meta'. Actions that discussed procedural aspects from the 'discuss sensors' topic were also reclassified as a second new 'meta' subtopic: 'Assist with sensor connection'. We also decided to remove the third round (on social subtopics), since participants were asked in that round to sort four cards into four categories with very similar names. These changes led to the updated topic structure as shown in Figure 4.16 and a new setup of the card sort rounds as presented in Table 4.3. Furthermore, descriptions were added to clarify the topics (these are included as Appendix C).

4.6.3 Card sort results

The card sort study was conducted in July and August of 2020 in the Netherlands. A total of 11 experts participated in the first round of the card sort study, and 10 of them completed all rounds. The participants rated their expertise on coaching relatively high ($Mdn = 4$, $IQR = 3-4$). They also indicated a

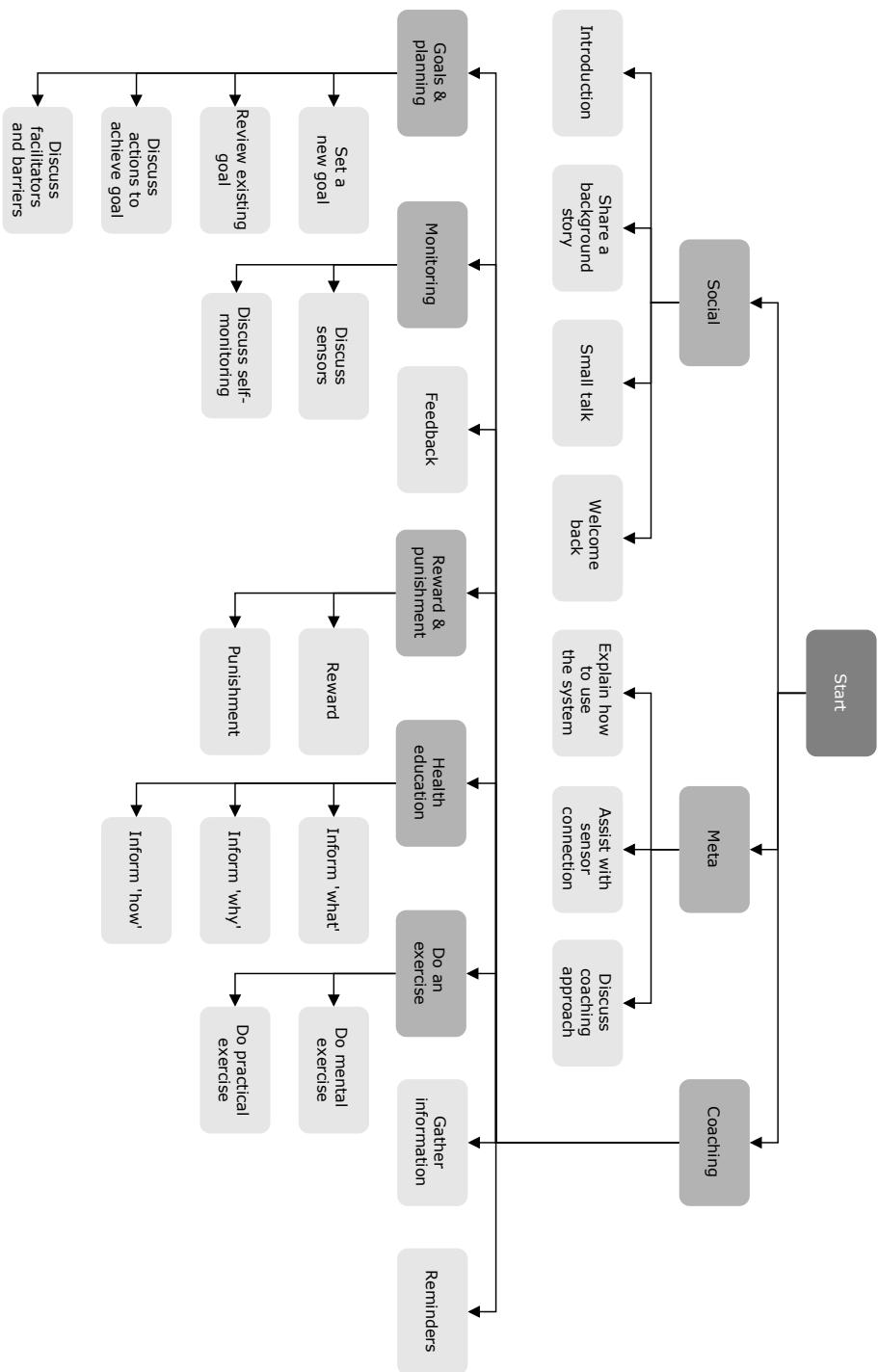


Figure 4.16: The updated topic structure (with pilot-based updates included).

Table 4.3: Number of cards and categories per round for the final set-up of the card sort study.

Round	N cards	N categories	Category labels
1	104	8	Goals & planning, monitoring, feedback, reward & punishment, health education, do an exercise, gather information, reminders
2	10	4	Set a new goal, review existing goal, discuss actions to achieve goal, discuss facilitators and barriers
3	7	2	Discuss sensors, discuss self-monitoring
4	23	2	Reward, punishment
5	29	3	Inform 'what', inform 'why', inform 'how'
6	21	2	Do mental exercise, do practical exercise

high expertise on eHealth ($Mdn = 5$, $IQR = 4-5$). They were less convinced about their expertise on conversational agents ($Mdn = 2$, $IQR = 1-3.5$). On average a participant spent about one hour to complete all rounds of the card sort.

Table 4.4: Distribution of the percentage of correctly sorted cards per round according to the intended labelling, and the inter-sorter agreement (Krippendorff's alpha).

Round	N cards	N categories	% sorted correctly <i>Mdn (IQR)</i>	K's α
1	104	8	69.23 (63.94-71.63)	0.51
2	10	4	80.00 (70.00-90.00)	0.65
3	7	2	71.43 (71.43-71.43)	0.80
4	23	2	100.00 (95.65-100.00)	0.86
5	29	3	79.31 (72.41-89.66)	0.47
6	21	2	86.96 (86.96-91.30)	0.80

Table 4.4 shows the percentage of correctly sorted cards and the agreement between participants for each round of the card sort. The mean class-wise F_1 -scores for all categories and rounds of the card sort can be found in Figure 4.17.

The first round, which had the most cards and categories, had the lowest percentage of correctly sorted cards, with participants not fully agreeing on where cards had to go. A frequently occurring switch was sorting 'health education'

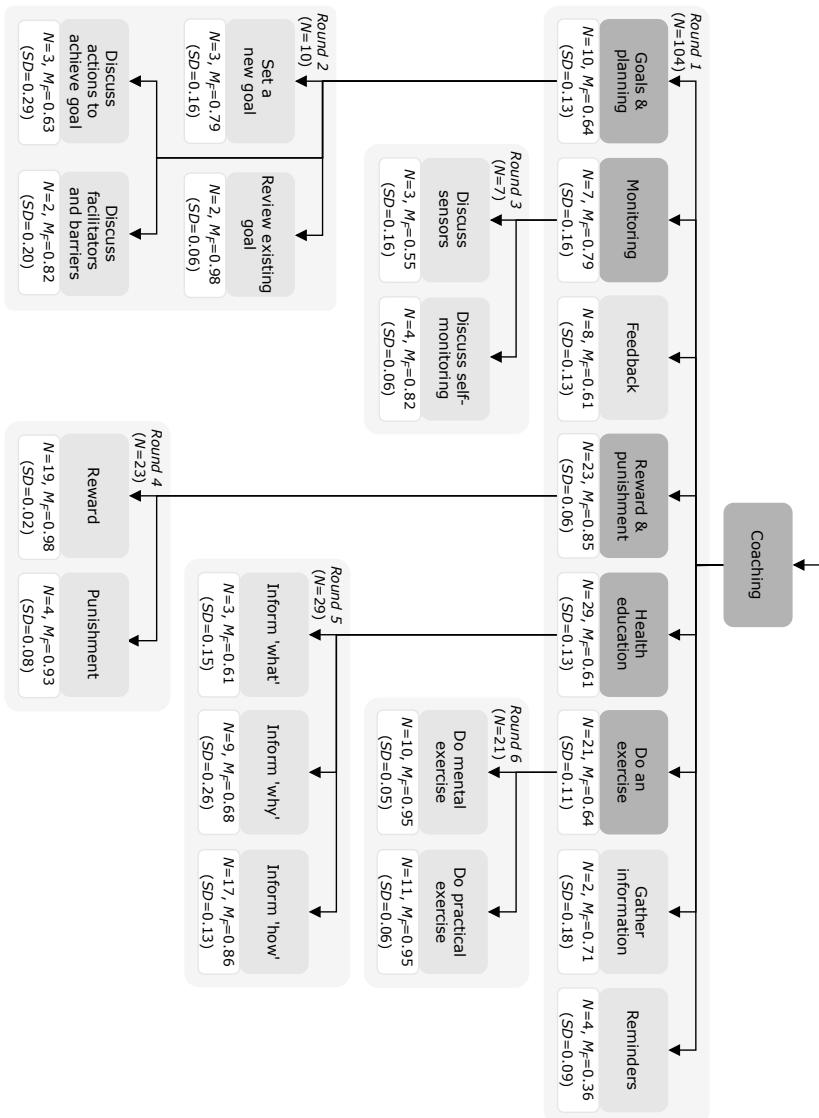


Figure 4.17: An overview of the mean class-wise F_1 -scores for the six card sort rounds. N indicates the number of cards.

cards as 'do an exercise', and vice versa. Two participants also showed a clear preference for a class ('goals & planning' and 'reward & punishment'). For the four cards in the 'reminders' category, most participants assigned the cards related to cues into other categories.

A frequent occurrence in the second round was that cards from the 'discuss actions to achieve goal' category were assigned to the 'set a new goal' category. For the third round there was a clear interpretation of 'self-monitoring' to also include monitoring through sensors, while it was intended to be 'as opposed to sensors'. Therefore, the two monitoring cards that were labeled as 'discuss sensors' were consequently sorted into the 'discuss self-monitoring' category by all but one participants.

The fourth round had a high percentage of correctly sorted cards, but some participants mislabeled the 'remove punishment' and 'remove reward' cards. 'Reduce reward' was also found to be challenging. In the fifth round, the 'inform 'what'' cards were mostly sorted correctly, but some cards from the 'inform 'why'' and 'inform 'how'' categories were assigned to either of the three. An example was the sorting of the card 'demonstration of the behaviour' into the 'inform 'what'' category instead of into the 'inform 'how'' category. Participants did not have much trouble sorting the cards for the sixth round.

Overall, the results of the card sort showed that most of the topics we defined were clear with the 'core' actions that defined the topics being sorted correctly and actions for which it was more difficult to imagine a specific dialogue being sorted differently than intended.

4

4.7 Discussion

In this article we defined a topic model for health coaching conversations with CAs, and evaluated it with experts. The defined model is a blueprint model, which can be used as a basis for developing tailored coaching content and can be extended to support specific types of coaching. While it might be necessary to update the model in the future, to integrate new insights or technological developments, it provides a practical and intuitive starting point for implementing systems to automatically select a conversational topic that is tailored to the user.

The resulting topic model provides a set of topics that are relevant to include in CAs for health coaching. These topics include a set of 'social' topics, for which the actions were deduced from literature on relational agents (e.g., Bickmore et al. (2005), Bickmore and Picard (2005) and Kowatsch et al. (2018)), as social interaction is essential for building up a personal bond and beneficial for interactions spanning a longer period of time (Bickmore, 2010; Bickmore et al.,

2018; Schulman & Bickmore, 2009). They also include a set of 'meta' topics, since it may be useful to let the CA (as the coach or expert) explain how to use various functionalities of the system. We also included a 'discuss coaching approach' topic, which can be used to discuss users' explicit preferences for coaching, since such control features can be beneficial for engagement (Perski et al., 2017) and facilitate the tailoring process. Finally, a set of 'coaching' topics is also included, which are deduced from behaviour change techniques (Michie et al., 2013) and persuasive strategies (Oinas-Kukkonen & Harjumaa, 2009; van Velsen et al., 2019). The topics that are included in the model can be used to generate dialogues with similar resulting structures as described in previous research (for example, the structure by de Kok et al. (2014)), but the hierarchical setup of topics in the model allows for a dynamic implementation with tailored selection. Furthermore, dialogues matching the topics in this model can be combined with tools that support communication and coaching, such as, the diary and graphs discussed by Beun et al. (2017) or the images used by Bickmore, Silliman et al. (2013).

In addition to its primary use, namely implementing automatic topic selection, the topic model can be used as a starting point when designing dialogue content for a new coaching agent. The model can be used in discussions with domain experts, for example, to identify which topics are important for a specific domain and application. When requesting domain experts to author dialogues, the actions that are classified as contributing to the topics can help clarify what the content for those dialogues could be. Using the topic model in this manner in the content design process also facilitates clear and concise description of the coaching content in scientific publications. Such need for description of coaching content is an issue reported by Michie et al. (2013) for health coaching, but that is still relevant when it comes to coaching with CAs (Kramer et al., 2020).

On a larger scale, tailoring coaching dialogues on different levels (domains, topics, actions, dialogue acts, utterances) in a step-wise process has practical benefits. That is, the task goes from deciding on what to say next (in general) to a series of decisions (Which domain? Which topic? Which action? Which dialogue act? Which utterance?). This means that each decision can be made by an expert module, where each module further specifies within the bounds of the previous decision. For example, once the topic has been specified to goal-setting within the physical activity domain, the dialogue could be modelled as a dialogue game (as in Snaith et al. (2018)) or a dialogue script for that specific topic. This limits the size and complexity of the models or scripts involved, which in turn not only simplifies their construction, but also helps to evaluate and check the resulting coaching dialogues from a safety perspective.

4.7.1 Implementation of topic selection

We started the process of developing the topic model with a practical application in mind: improving coaching conversations by tailoring the selection of topics that are discussed with individual users. Implementing this topic selection has two key steps⁶. First, a representation of topics and a selection algorithm would need to be implemented. Second, the calculation of a topic's relevance by the selection algorithm requires parameters to be added to topics that can be used for that purpose.

When it comes to the representation of topics, we made a deliberate choice to structure our topic model as a hierarchical tree. This means that, intuitively, the selection of topics can be implemented as making a choice at each split (e.g., ‘Would it be more relevant to have a social, meta or coaching conversation now?’). An implementation of such a tree structure would require the topics to be implemented as nodes. Each of these nodes can then be assigned a set of selection parameters that can be used to calculate a topic's relevance.

An algorithm that traverses the topic tree and outputs which topic to discuss starts at the top of the structure ('start') and uses the calculated relevance (e.g., a weighted average of selection parameters) for the nodes in the first split to select a subtopic. It then continues this process until it selects a topic that has no more subtopics. Selection of a topic in each step could follow an exploitation strategy by always selecting the node with the highest relevance, but an exploration element could also be included (e.g., by randomly selecting from a distribution based on the topics' calculated relevance).

Selection parameters that are added to the topic nodes can represent various important elements for a topic's relevance. Examples that come to mind are: *availability* (e.g., are there still dialogues available for this topic?), *prerequisites* (e.g., is there sensor data to give feedback on?), *strategy* (e.g., the inform topic could be more relevant when following a health education strategy) and *time* (how long has it been since this topic has been discussed previously?). These can be supplemented with parameters that represent other connections between relevant topics and user characteristics as investigated in previous research on tailoring coaching content.

⁶A proof-of-concept implementation of a topic selection component that uses the process described in this section has been included in the Agents United Platform (Beinema, Davison et al., 2021), where dialogue execution can be performed using either WOOL scripts (Roessingh Research and Development, 2020) or dialogue games (Snaith et al., 2020). We also evaluated topic selection in a micro-randomized trial as part of the final evaluation of the Council of Coaches project (paper submitted, protocol published as Hurmuz et al. (2020)).

4.7.2 Limitations

In this article we focused on a blueprint model that includes a set of commonly included topics for health coaching conversations with CAs. A first aspect of our study which might be seen as a limitation could be the use of a closed-card sort study to verify the constructed model instead of constructing the model itself using an open card-sort. However, the large number of actions (115 items) that we extracted from literature would have made an open card sort a task with an enormous cognitive load – as was also confirmed by the feedback from the two participants in the pilot for the closed-card sort.

A second limitation is that there were some actions based on behaviour change techniques for which the description in the card sort was ambiguous. For example, these lacked an immediacy indicator which means that there was no clear distinction between ‘explain to the user how they might perform healthy behaviour’ and ‘give the user a specific exercise to perform now’. We made the decision to stay true to the original BCT descriptions when defining actions (and on the cards for the card-sort), but such ambiguity might have made some of the cards in the card sort more difficult to sort. However, this does indicate that some behaviour change techniques might simply be more difficult to translate to the digital domain and conversations with an agent than others – a notion that is also supported by reports of occurrence of BCTs in the literature (DeSmet et al., 2019; Kramer et al., 2020).

4.8 Conclusion

The inclusion of automatic topic selection in conversational agent applications for health coaching has the potential to improve user’s interactions with these applications by presenting them with engaging and relevant conversations. The concept of a topic-tree was introduced both as a framework for the automatic topic selection, but also as a human understandable structure of the possible topics of conversation. While there are still steps to be taken, the definition of a blueprint topic model as described in this article is a step in the right direction.

Now that we have a blueprint topic model and initial implementations of the topic selection component, the further development of the topic selection component provides plenty of directions for future research. Future work could include the development of tools that allow domain experts to construct topic models and define selection parameters for the embedded topics. A generic variable-driven topic selection tool would allow domain experts to easily influence the artificial intelligence algorithms that automatically select conversation topics based on human understandable variable input. Furthermore, insights from the existing literature on tailoring coaching content within eHealth applications in general

might be transferred to the topic selection step in the tailoring process. Finally, the general development of systems that tailor dialogues with conversational coaches in multiple steps can be further investigated, for example, from the technical and user interaction perspectives.

Intermezzo

From concept to component

A proof of concept implementation of automatic topic selection

With the five-step tailoring process and the blueprint topic model (**Chapter 4**) in mind, the next step was to implement a proof of concept component that performed automatic topic selection and to evaluate it. This evaluation is discussed in **Chapter 5**. In that micro-randomized trial, a condition that uses automatic topic selection (coach-initiative) is compared with one in which the user decides the conversational topic (user-initiative). This intermezzo provides an insight into the implementation of the topic selection component.

The new proof of concept component (the 'topic selection engine') was integrated in the version of the Council of Coaches application used in the project's final user study⁷. If the experimental condition required the agent to suggest a tailored topic, the topic selection engine would provide a relevant topic to discuss based on available information about the user and their interaction history. To that end, the component was developed to include a topic structure based on the blueprint model in **Chapter 4**, sets of topic selection parameters that could be used to compute a topic's current relevance, and a topic selection algorithm.

In the topic structure, which is implemented as a tree, each topic is represented as a node. The leaves of this tree are the topics for which the execution of a dialogue script can be started. Each topic node has a topic, a list of child nodes, and a list of selection parameters that can be used to calculate its relevance. Furthermore, each node is also assigned an a-priori value and weight. The top of the tree is the start node, which is also the starting point for the selection algorithm. This topic selection algorithm (see Code Block 1) navigates the topic structure by selecting a topic at each split. For example, the first choice in the **Chapter 5** study is between the 'Social' and 'Coaching' topics. It repeats this process for subtopics of the chosen topic until there are no more subtopics. It then returns the chosen topic and a dialogue can be started.

The topic selection algorithm also has a parameter that can be used to set the balance between exploitation and exploration. This setting can be used to indicate if the algorithm should always choose the node with the highest relevance (exploitation) or if it selects the next node from a distribution in which each candidate topic is represented based on their relevance (exploration). In this manner, topics with a higher relevance will have a larger chance of being chosen, but the algorithm could also still select a less relevant topic. Such exploration can be a useful feature, for example, when the agent system would be extended to learn if topics are effective (a case in which it sometimes needs to try a new option) or if it is desired to add a slight unpredictability to the conversations to keep them interesting.

⁷It has also been integrated in the project's more complex technical demonstrator, which has been released as the Agents United Platform (Beinema et al., 2021).

```
TopicNode navigateTopicStructure(node, nodeRelevances)
    chosenNode = node
    if node.hasChildren()
        childNodesWithRelevance = this.calculateRelevanceForChildren(node.getChildren())
        randomForExploration = random.nextInt(100)
        if randomForExploration > explorationThreshold
            chosenNode = selectNodeWithHighestRelevance(childNodesWithRelevance)
            return navigateTopicStructure(chosenNode, childNodesWithRelevance)
        else
            chosenNode = selectNodeWithRandomFactor(childNodesWithRelevance)
            return navigateTopicStructure(chosenNode, childNodesWithRelevance)
    return node
```

Code Block 1: Simplified code for the topic selection algorithm.

A node's *relevance* is calculated by taking the weighted average of a node's selection parameters and a-priori weight and value (see Equation 1, with p being the number of selection parameters for a node).

$$\text{Relevance} = \frac{\text{weight}_{a-priori} \cdot \text{value}_{a-priori} + \sum_{i=1}^p \text{weight}_i \cdot \text{value}_i}{\text{weight}_{a-priori} + \sum_{i=1}^p \text{weight}_i} \quad (1)$$

The selection parameters can encode relevance for categorical or continuous information, and have weights and values between 0 and 1. The specific value of selection parameters is determined using the information that is stored about the user and their previous interactions. For example, a parameter encoding if a topic has already been discussed can be given value 0 (decreasing relevance) or 1 (increasing relevance). A parameter in itself can encode a simple fact or piece of information, but combination of multiple parameters for a topic in the selection between multiple topics can result in seemingly intelligent behaviour. Examples of relevance properties that can be represented include availability, prerequisites, strategies or even time. That is, topics can be more relevant if a new dialogue is available (e.g., for background stories), if certain prerequisites have been fulfilled (e.g., sensor data is available to give feedback on), if it is relevant for the strategy (e.g., the inform topics for the health education strategy), or if the time since a topic was discussed increases (e.g., the number of days since the last background story).

Once a topic is selected by the topic selection engine, a dialogue is started in which the agent suggests to discuss that topic (e.g., 'Let's talk about setting a new goal!'). The effect of the agent taking the initiative in this manner on the interaction with users is the subject of **Chapter 5**.

Chapter 5

Automatic topic selection for embodied conversational agents in health coaching: A micro-randomized trial

Based on:

Beinema, T., op den Akker, H., Hormuz, M., Jansen-Kosterink, S., Hermens, H. Automatic topic selection for long-term interaction with embodied conversational agents in health coaching: A micro-randomized trial (*submitted for publication*).

Abstract

Introduction: Embodied Conversational Agents (ECAs) can be included in health coaching applications as virtual coaches. The engagement with these virtual coaches could be improved by presenting users with tailored coaching dialogues. In this article we investigate if the suggestion of an automatically tailored topic by an ECA leads to higher engagement by the user and thus longer sessions of interaction.

Methods: A Micro-Randomized Trial (MRT) was conducted in which two types of interaction with an ECA were compared: (a) the coach suggests a relevant topic to discuss, and (b) the coach asks the user to select a topic from a set of options. Every time the user would interact with the ECA, one of those conditions would be randomly selected. Participants interacted in their daily life with the ECA that was part of a multi-agent health coaching application for a period of 4-8 weeks.

Results: In two rounds, 82 participants interacted with the micro-randomized coach a total of 1,011 times. Interactions in which the coach took the initiative were found to be of equal length as interactions in which the user was allowed to choose the topic, and the acceptance of topic suggestions was high (71.1% overall, 75.8% for coaching topics).

Conclusion: Tailoring coaching conversations with ECAs by letting the coach automatically suggest a topic that is tailored to the user is perceived as a natural interaction that is accepted by users. Future research could focus on improving the novel coaching engine component that supports the topic selection process for these suggestions or on investigating how the amount of initiative and coaching approach by the ECA could be tailored.

5.1 Introduction

Digital behaviour change interventions (DBCIs) are increasingly investigated (Brinkman, 2016) as tools to support people in their health behaviour change process, both as a means of treating health conditions and in preventative contexts. These applications can provide users with support as needed and are always available. However, they face challenges in terms of adherence (Crutzen et al., 2011; Kohl et al., 2013; Nijland, 2011; Wangberg et al., 2008; Yardley et al., 2016). Potential causes for this lack of adherence are actively being researched. It appears that contributing factors are the lack of direct involvement of a health care professional (*no social incentive*) and content that does not always fit the user's personal situation (*relevance of content*) (e.g., Andersson et al. (2009) and Buimer et al. (2017)). Two directions of research aimed at improving the interaction and engagement with these applications are therefore *Embodied Conversational Agents* (ECAs) and *tailoring*.

ECAs are “*more or less autonomous and intelligent software entities with an embodiment used to communicate with the user*” (Ruttkay et al., 2004). In DBCIs these agents can take on the role of a coach (Kramer et al., 2020) and they give a system social ability, which is important for maintaining a collaborative relationship (Bickmore et al., 2010; Bickmore et al., 2018; Kamphorst, 2017). ECAs make the use of health applications easier, more satisfying and less frustrating (André & Pelachaud, 2010; Bickmore et al., 2016) and potentially more effective (Ma et al., 2019). Furthermore, they are always available and their dialogues can be tailored dynamically to the user, for example by elaborating on certain topics of discussion when needed (Bickmore & Giorgino, 2006). In health coaching applications, ECAs can be the main component (e.g., (op den Akker et al., 2018; Sebastian & Richards, 2017)) or they can be part of a broader application (e.g., van Velsen et al. (2020)).

Ultimately, regardless of their specific features and capabilities, health applications need to be *engaging* for a longer period of time to impact users' behaviour change (Cole-Lewis et al., 2019; Perski et al., 2017; Yardley et al., 2016). Or, as stated by Bickmore et al. (2010) in the context of ECAs: “*Engagement is a prerequisite for other system objectives: If a user stops interacting with a system, then it cannot have any further impact.*” For DBCIs, Cole-Lewis et al. (2019) distinguish 'Big E' (health behaviour engagement) and 'Little e' (DBCI engagement), with a subdivision for the latter in user interaction with a) features that encourage frequency of use and b) behaviour change intervention components. They also emphasise that 'Big E' is dependent on 'Little e'. In general for interaction with applications, short-term engagement tends to be characterised as flow (Hamari et al., 2016; Nakamura & Csikszentmihalyi, 2002) or enjoyment (O'Brien & Toms, 2013), while long-term engagement can be seen as the duration and

depth of usage of a system over time (Couper et al., 2010). There are several objective measures for long-term engagement, such as the number of voluntary interactions users choose to have, the number of logins, and amount and type of content used (Perski et al., 2017; Trinh et al., 2018).

Tailoring (op den Akker et al., 2014) is investigated to make a digital health application's content, such as the conversation with an ECA, more personally relevant for the specific user (e.g., Krebs et al. (2010), Ryan et al. (2019) and Wangberg et al. (2008)). However, most research on tailoring coaching content has its focus on the last steps of presenting information (e.g., inserting tailored goals or adjusting wording of sentences). In general, in health coaching applications with ECAs, participants either all follow the same coaching programme or a human expert manually defines the high level structure of content to be presented (e.g., Abdullah et al. (2018), Benitez-Guijarro et al. (2019) and Fadhil et al. (2019)). Where the first approach presents all participants with the same dialogues, the second approach requires continuous involvement of health professionals. Alternatively, participants could be allowed to select what they want to discuss, which can lead to more engagement because the user feels that they are in control (Perski et al., 2017). However, that approach does put the initiative with the user, while the coach should be the expert – or as Kamphorst (2017) states: “*an e-coaching system should be credible and proactive*”. Therefore, we think that combining the first and second approach by letting an ECA automatically suggest a relevant topic to discuss (both immediately and in the long-term), while allowing users to make the final decision, could be an improvement in the interaction with virtual coaches.

5.1.1 Research aims

In this article, we present a study in which participants interact with an ECA over a longer period of time in a daily life setting. The study investigates the influence of automatically tailoring coaching dialogues at the topic level on users' interaction with the application. Specifically, we let a coaching ECA take the initiative by suggesting a relevant topic to discuss, and we compare this with a more conventional approach in which the user selects a topic themselves. If such suggestion of relevant topics leads to longer interactions between users and the ECA, this would be a step towards extending tailoring methods for effectively coaching people to lead a healthy lifestyle. Our research question therefore is:

RQ: *What is the influence of automatically tailored topic suggestions on the length of interactions with an ECA?*

The use of ECAs, personal relevance and tailoring have all been found to have an effect on engagement with DBCIs (e.g., Krämer et al. (2010), Krebs et al.

(2010), Perski et al. (2017), Ryan et al. (2019) and Wangberg et al. (2008)). Furthermore, manual tailoring of coaching topics or modules by human experts is also appreciated (Abdullah et al., 2018; Benítez-Guijarro et al., 2019; Fadhill et al., 2019). We therefore hypothesise that suggestion of a relevant topic will lead to increased engagement with the ECA, and thus longer interactions:

H1: *Suggestion of automatically tailored topics will lead to longer interactions with the ECA.*

In addition, we take a closer look at users' acceptance of topics that are suggested when the ECA takes the initiative, which gives an indication of the quality or relevance of those suggestions. We also perform an initial exploration of potential demographics that might be of influence on the acceptance of these suggestions, since various demographics have been found to influence engagement and appreciation of DBCIs (e.g., as reported in Beinema et al. (2021), Hardiker and Grant (2011), Perski et al. (2017) and van Velsen et al. (2019)) and ECAs (e.g., Payne et al. (2013) and Pezzullo et al. (2017)). Both these investigations can provide starting points for future work on automatically tailoring topics of conversation. Therefore, our second and third hypothesis are the following:

H2: *Because topic suggestions are tailored to the individual, participants will accept suggested topics more often than not.*

H3: *A participant's demographics affect their acceptance of a suggested topic.*

In the following sections, we will first provide some background on multi-agent health coaching applications, proactiveness in virtual coaches, and the micro-randomized trial method. Then we will provide details on the design and implementation for both conditions of the micro-randomized trial, and our methods for conducting the trial and analysis. Finally, we will present the results and discuss our findings and conclusions.

5

5.2 Background

5.2.1 Multi-agent health coaching

Changing behaviour can prevent or relieve health conditions, but change in the long-term tends to be difficult (Bouton, 2014). As stated previously, health coaching applications with ECAs can support users in this process. Most ECA applications feature a single ECA as a coach who has a specific expertise (e.g., physical activity (King et al., 2017; Watson et al., 2012)). However, health

often requires support on multiple domains (Huber et al., 2016; World Health Organization, 1946). This has led to single ECAs coaching on multiple domains (Gardiner et al., 2017; Klaassen et al., 2018) and recently multiple coaches coaching on multiple domains (Das et al., 2019; Kramer et al., 2021; op den Akker et al., 2018). Having multiple coaches available at the same time provides opportunities for vicarious persuasion (Kanthalraju et al., 2018) and engagement (André & Rist, 2001). Each agent can have a specific expertise and role – for example, a dietitian or personal trainer – and multiple viewpoints can be presented without an ECA contradicting itself (Hayashi & Ogawa, 2012; Kanthalraju et al., 2019). We perform our experiment in a setting where multiple ECAs are present and interact with the user, so that the results can be incorporated in tailoring approaches for a broad range of ECA health coaching applications.

5.2.2 Proactiveness in virtual coaches

Taking the initiative or being proactive, is an important property for a virtual coach. As previously stated by Kamphorst (2017), an e-coaching system needs to, for example, invite the user to reflect on their commitment to a goal or warn them at suspected moments of weakness. This requires that the system is flexible enough to respond to new developments and is able to start a communication about those topics.

When it comes to starting interactions, ECAs that are proactive were found to be better in providing support (e.g., on loneliness to older adults (Ring et al., 2013)). Agents that are proactive can also be perceived to be more helpful, even if their proactiveness does not immediately improve task performance (Xiao et al., 2002). It is however important to use the right tone of voice when being proactive, for example, when giving reminders to users during working hours (Bickmore et al., 2007). Furthermore, while a proactive coach could provide the right coaching at the right moment, in the end, a virtual coach should support the user and not just dictate how they should behave (Brinkman, 2016). Thus, research on proactiveness of virtual coaches could learn from developments in shared-decision making research – both between humans (e.g., on facilitators and barriers (Joseph-Williams et al., 2014)) as well as between humans and ECAs (Zhang & Bickmore, 2018).

5.2.3 Micro-randomized trial

Properly assessing the effectiveness of technology-supported health services in real-world settings is challenging and there is a need for pragmatic study designs (Ekkeland et al., 2010, 2012; Kairy et al., 2009; LaPlante & Peng, 2011). The *Micro-Randomized Trial* (MRT) is a method of evaluating interventions

originally proposed by Klasnja et al. (2015) for the evaluation of *Just-in-Time Adaptive Interventions* (JITAl). They found that conventional methods such as randomized trials were not suitable for evaluating these JITAls. In a micro-randomized trial, an intervention option is randomly selected at every relevant decision point (e.g., whether or not to send a notification). Furthermore, in a MRT, effect is measured after each intervention through a short-term parameter that resembles the intended long-term effect. In our evaluation, the initiative for choosing a topic is randomized every time the user interacts with a coach, and we measure the length of the interaction that immediately follows.

5.3 Methods

We performed a MRT to compare users' responses to the coach suggesting a topic (coach-initiative) with a more conventional implementation of coaching dialogues in which users could select a topic themselves (user-initiative). The micro-randomized trial was embedded in the final evaluation of the Council of Coaches application (op den Akker et al., 2018), which consisted of two separate rounds with participants. The full protocol for that evaluation is described in an article by Hurmuz et al. (2020). Since the MRT shared the same participants and setup as the full evaluation, we will only summarise the important aspects of the overall procedure that are relevant for the MRT, while elaborating on the design and implementation of the MRT itself.

5.3.1 The multi-agent eHealth application

In the Council of Coaches application (Hurmuz et al., 2020; op den Akker et al., 2018), users can interact with multiple ECAs. Each of these ECAs has their own role, expertise and backstory. There are six coaches with expertise on the following domains: physical activity, nutrition, social activity, cognition, chronic pain and diabetes. In addition, there was an agent that provided peer support and an agent that guided the user through the application (the assistant). After an intake with the assistant, users could select their council of coaches. The physical activity coach and nutrition coach were obligatory, and the diabetes and chronic pain coaches were only available to those who had indicated in the intake to have those conditions.

An example interaction with the application can be found in Figure 5.1. Users could start an interaction with one of the coaches in the application by clicking on a coach of their choice. The main participants in such an interaction are the user and a specific coach, but the other coaches can also join in to provide their own viewpoint on the ongoing conversation. The interactions followed a

speech-bubble and reply-button paradigm, as depicted in Figure 5.1.

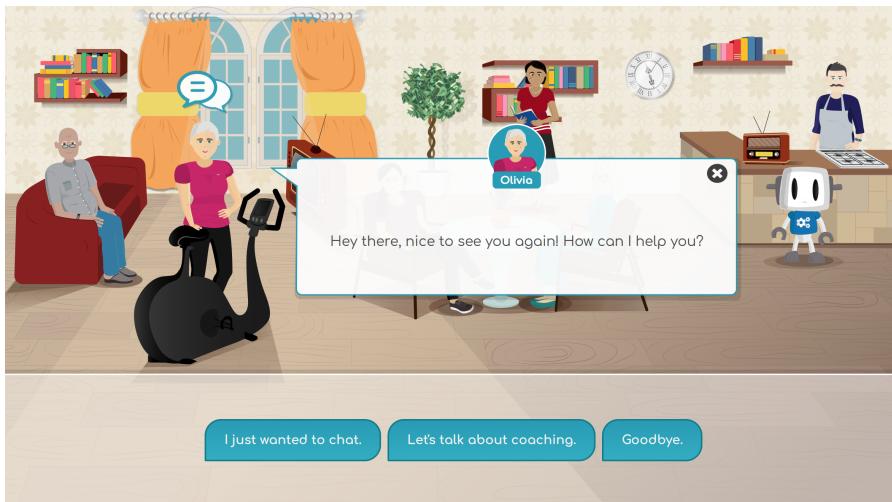


Figure 5.1: An example interaction with the physical activity coach (Olivia) in the multi-agent eHealth application.

5.3.2 Study design

The MRT was specifically set up for one of the two obligatory coaches, the physical activity coach (Olivia). In the MRT, we randomized two types of interaction. The first is an interaction in which the user decides what they would like to discuss. The second is an interaction in which the coach suggests a topic to discuss. Thus, there were two conditions:

Condition 1: The user gets the initiative and chooses the topic of conversation (*user-initiative*).

Condition 2: The ECA takes the initiative and suggests the topic of conversation (*coach-initiative*).

Every time the user would click on the physical activity coach (to start the interaction), the system would micro-randomly select one of the two conditions; both with a 50% chance (see Figure 5.2). The difference between the two conditions was in the start of the interaction, while the dialogues that followed were the same.

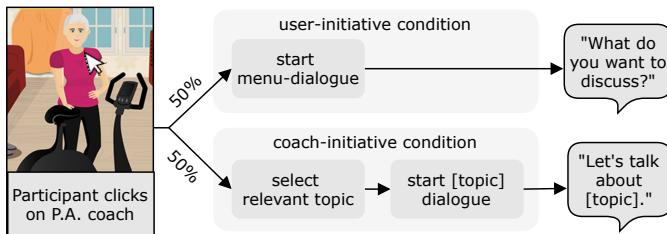


Figure 5.2: A schematic representation of the procedure in the MRT.

5.3.3 Content

Dialogue content was created for eight topics, namely:

Introduction. A conversation between the coach and the user in which the coaches introduce themselves to the user and provide some information on their background and the type of coaching content the user might expect from them.

Background story. Social dialogues in which the coach shares a part of their background story with the user. For example, a short story about Olivia's dog ('Brian') and how she likes to go running with him.

Discuss sensors. Dialogues that cover subtopics such as 'Connecting your activity tracker.', 'Why should I use an activity tracker?', and 'My sensor is not working, why is that?'.

Goal-setting. Dialogues in which the user can set a new long- and short-term goal, or change their current goal.

Feedback. Dialogues that allow the user to view their measured activity data. In these dialogues, the coach can also show the user their 'activity book' widget, which provides an overview of their physical activity (steps taken) over the past week.

Gather information. Dialogues in which the coach asks the user for information that can be used to tailor or personalise the coaching provided. For example, the question 'Do you have a dog?' can be used to suggest that the user walks with their dog more often as a form of increased activity.

Inform 'why'. Dialogues in which the coach explains why it is good to be physically active. For example, 'Being active increases blood flow, which is healthy for your brain'.

Inform 'how'. Dialogues in which the coach gives advice on how to be more physically active. For example, 'Take the stairs instead of the elevator'.

For every topic, dialogue scripts were written, which were available in both Dutch and English.

5.3.4 Implementation

The health coaching application was implemented for use on a tablet, laptop or desktop using a client-server setup. The WOOL Dialogue Platform (Roessingh Research and Development, 2020) was used for dialogue authoring and execution. To facilitate the implementation of both conditions in the MRT, we structured the topics as an hierarchical tree. This tree can be found in Figure 5.3.

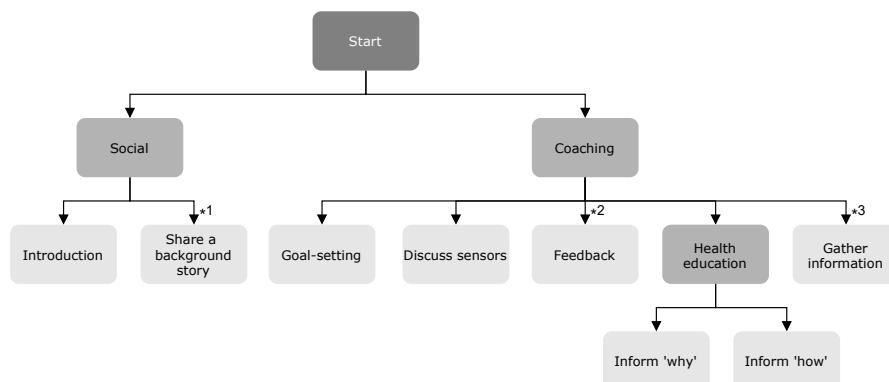


Figure 5.3: The topic model featuring the eight topics for which dialogues could be held with the physical activity coach in both conditions. *¹ The *Background Story* topic was added for the second round of the study, and *² the *Feedback* topic was extended. *³ The *Gather Information* topic was only available in the coach-initiative condition.

User-initiative

In the user-initiative condition, the user could select a topic to discuss. We defined a 'menu'-dialogue that facilitated this. When the user would click on the physical activity coach for the first time, they would get the *Introduction*-dialogue, but for each subsequent interaction the coach would state 'Hey there, nice to see you again! How can I help you?' (resembling *Start* in the topic model). The user could then respond with: 'I just wanted to chat.' (leading to *Social*), 'Let's talk about coaching.' (leading to *Coaching*) or 'Goodbye' (ending the interaction). If the user would then click on 'Let's talk about coaching', the coach would say 'Let's get down to business and talk about some coaching. What do you want to discuss today?' and the user could select: 'My goals' (*Goal-Setting*),

'My activity tracker' (*Discuss Sensors*), 'Tips and info' (*Health Education*), 'My progress' (*Feedback*), or 'Goodbye.' (ending the interaction). In this manner, the menu-dialogue would allow the user to navigate towards a dialogue on the topic of their preference in line with the topic structure as depicted in Figure 5.3.

Coach-initiative

In the coach-initiative condition, the coach would suggest a topic to discuss. This suggested topic was selected by our coaching engine component (original concept by Beinema et al. (2018)). This coaching engine applied a topic selection algorithm that took into account parameters such as dialogue availability, completion dates and data prerequisites.

In this condition, the dialogues for the different topics would be preceded by a short 1-step dialogue in which the coach suggests the topic. Such a statement could be, for example, 'Would you like me to tell you something about how you can be more active?'. The user could then respond with (a variation of) 'Yes, that would be nice.' (accepting the suggestion) or 'Goodbye.' (ending the interaction). In the second round of the study, users would additionally have the option to reply 'I would like to discuss something else.' (rejecting the suggestion and being forwarded to the menu-dialogue that was also used in the user-initiative condition).

The topic selection algorithm

For the coach-initiative condition, the topic model was implemented as a tree (with the topics as nodes). Each topic was assigned a set of selection parameters that resembled aspects that contributed to that topic's relevance (positively or negatively). These parameters had a weight assigned to them and their value was dependent on the information that was stored for a specific user (e.g., their age, previously completed dialogues, or available data). Topics could also be assigned an a-priori weight and value.

The *relevance* for a topic was computed by taking the weighted average of the parameters' weights and values using the following formula (with p being the number of selection parameters for a topic):

$$\text{Relevance} = \frac{w_{a-priori} \cdot v_{a-priori} + \sum_{i=1}^p w_i \cdot v_i}{w_{a-priori} + \sum_{i=1}^p w_i} \quad (5.1)$$

Whenever a topic has to be suggested, the selection algorithm starts at the top of the tree ('Start'), computes the relevance for all the direct subtopics and then select one of those subtopics (in this case 'Social' or 'Coaching'). Depending on the balance between exploration and exploitation that is set for the algorithm,

the selection is made from a weighted distribution (more relevant topics have a higher chance of being selected) or by selecting the topic with the highest relevance. Between the first and second round of this study, the exploration probability was increased from 0% to 25%. The selection process continues until a topic is selected that has no further subtopics.

5.3.5 Participants

The study's population consisted of older adults (55 years or older). Inclusion criteria were that they had to be able to read and speak Dutch or English, had a WiFi connection at home, were able to provide informed consent, and were able to see a smartphone/tablet screen clearly. Participants were recruited in the Netherlands and Scotland in two rounds, each preceding the corresponding round in the study.

5.3.6 Procedure

The MRT was conducted in two rounds, which both followed the procedure for the overall evaluation (Hurmuz et al., 2020) (see Figure 5.4).

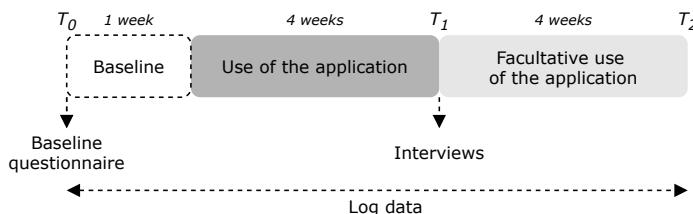


Figure 5.4: A timeline showing the procedure for the micro-randomized trial.

At T_0 , a researcher met with participants to provide them with the technology, create an account and complete the intake with the assistant agent, and let the participant complete the baseline questionnaire (on demographics and health status). In the one-week baseline phase that followed, participants wore the activity tracker, but they did not yet use the coaching application. This phase was included to ensure activity data would be present when they would start using the application and allowed for the novelty effect of the tracker to wear off. After this week, the participants were asked to use the application for four weeks as they wanted (they received no instructions for frequency). At T_1 participants were interviewed (additional questions for the MRT were added in round 2) and they could indicate whether they wanted to use the application for an additional four weeks.

Due to Covid-19 and our target population of older adults, for the second round, the same overall procedure was followed, but direct contact between researchers and participants was limited. Materials were sent to participants by post, explanations given over the phone, and interviews were also conducted by phone.

5.3.7 Measurements

Three types of data were collected for the MRT. The first type of data came from the demographic questions that were included in the baseline questionnaire at T_0 . Participants were asked for their gender (male/female), age, highest level of education (primary education, secondary education, further education or higher education) and living situation (alone, married/living together, living together with my caregiver, other). They were also asked the three items from the health literacy scale by Chew et al. (2004) (on a five-point Likert scale), four items on attitude towards technology (Agarwal & Prasad, 1998) (on a seven-point Likert scale), questions on their motivation to live healthy (van Velsen et al., 2019) (on a seven-point Likert scale) and for their self-reported physical activity level (not at all, not at all but thinking about beginning, less than 2.5 hours a week, more than 2.5 hours a week in the last six months, more than 2.5 hours a week for more than six months).

The second type of data came from the system's interaction logs. These logs included a record for every dialogue that was started with a coach. Stored information per dialogue included:

- *Dialogue steps.* Each statement by a coach or reply by a user is counted as one dialogue step.
- *'Cancellation' boolean.* Whether the dialogue was cancelled by clicking on 'X'-button of the coach's speech-bubble.
- *'Completed' boolean.* Whether the dialogue was completed, that is, the dialogue ended because it was finished or the user ended the dialogue by responding with a 'Goodbye'-reply.
- *Condition.* Indicating experimental condition: user-initiative or coach-initiative.
- *Referrer.* If the dialogue was started because the user was referred to it from another dialogue (and if so, which one).

The number of dialogue steps was the parameter that we used to quantify *interaction length*, as it provided a clear short-term parameter for the amount of

interaction with the agent. The ‘Cancellation’ and ‘Completed’ booleans and the first user-response to a suggestion were used to analyse the acceptance of suggestions.

The third type of data were participants’ responses to interview questions. We added these questions to the interviews that were conducted at T_1 for the second round. These questions were the following: ‘Did you interact with the physical activity coach?’, and (if ‘yes’) ‘Did you notice that the physical activity coach sometimes came with a suggestion for a topic?’ and (if ‘yes’) ‘What did you think of this?’.

5.3.8 Pre-processing of log data

The outcome parameter that we used to determine the length of an interaction was *the number of dialogue steps*. An *interaction* started when the user clicked on a coach, and could then involve a chain of several dialogues and ended when the last dialogue was completed (e.g., the user choosing ‘Goodbye’), cancelled (closing the speech-bubble with the ‘X’-button) or when it was not completed or cancelled, but a new dialogue was logged (e.g., when the browser was refreshed). A user-initiative interaction started with a menu-dialogue that had no referrer and a coach-initiative interaction started with one of the ‘Shall we discuss X’-dialogues.

We applied the following pre-processing steps:

1. We excluded logs for dialogues that were the result of a ‘double click’ error. That is, two dialogues are started within 1 second of each other, with the first log only including the agent’s first statement.
2. The ‘sensor connection completed’ dialogue that was initiated by the system after successful connection of a sensor was not automatically labelled with the correct condition. We manually relabelled dialogues following this event with the same label as given to the preceding ‘connect sensor’ dialogue.
3. The ‘sensor connected’ dialogue was triggered when the system registered that it could retrieve data. In some cases, there was a delay for the start of that dialogue after participants returned from the external connection page. Sometimes, the participant had started a new dialogue with the coach themselves, which would then be interrupted. We removed such interrupted dialogues if they were only one dialogue step long. If the new dialogue was longer than one step, we removed both the interfering and the interfered dialogue, since neither could reach their full number of dialogue steps (and in some cases both dialogues belonged to different conditions).

To conclude pre-processing, we aggregated the information for dialogues that were part of one interaction. The resulting data set contained one row per interaction with information such as a participant identifier, experimental condition, number of dialogue steps, whether the last dialogue in the interaction was completed or cancelled, and a transcript of statements and replies.

5.3.9 Analysis

IBM SPSS Statistics version 25 was used for data analysis. Since we made adjustments to the available content and the settings of the topic selection algorithm between the two evaluation rounds, we analysed both rounds separately, except for the analysis about investigating the influence of demographics on acceptance.

To test our first hypothesis that interactions in the coach-initiative condition would be longer (in terms of dialogue steps) than interactions in the user-initiative condition, a generalised estimating equations (GEE) analysis was performed. Participant numbers were added as the subject variable. An exchangeable structure was selected for the working correlation matrix. We checked our intended dependent variable (number of steps) for normality and decided to include it with a natural log transformation applied. Condition (user-initiative or coach-initiative) was included as the main predicting factor in our model. All other settings were set to the standard options.

The interview data from the added questions in the second round were analysed to gather insights into users' experience of the change in initiative. The full interviews were recorded and transcribed for the larger evaluation (Hurmuz et al., 2020). The specific answers to the three interview questions that we added for this study were listed in an Excel file. In that file, we counted the number of 'yes' and 'no' responses for the 'Did you interact with the coach?' and 'Did you notice the coach suggesting a topic?' questions. We categorised participants responses to the 'What did you think of this?' follow-up question on a five-step scale from negative to positive.

To assess the acceptance of topics that were suggested by the coach in the coach-initiative condition (as is relevant for our second hypothesis), we created an overview table that listed the number of accepted and rejected suggestions per topic for each participant. From that table we then computed the overall number of suggestions that were accepted (user agreed to discuss the topic) or rejected (user wanted to change the subject, chose the 'goodbye' response, closed the speech-bubble, or did not respond at all). We also computed the percentage of user-initiative interactions in which the coach only made one statement to get an idea of the rejection rate in that condition.

Finally, to explore if certain user characteristics might be linked to acceptance or rejection of suggested topics (our third hypothesis) we observed individual users' responses. We included participants from both rounds that had had more than 10 total interactions in the coach-initiative condition. A non-parametric Kendall's Tau correlation was performed between ordinal and continuous demographics and the percentage of accepted suggestions. Mann-Whitney U-tests were performed to test if there was a difference in acceptance for country (NL, SC) and gender (male, female). We did not include the living situation demographic in our tests, since almost all participants were married and/or living together.

5.3.10 Ethical approval

As previously stated, the micro-randomized trial was included in a larger evaluation (Hurmuz et al., 2020). That evaluation was conducted according to the principles of the Declaration of Helsinki (64th WMA General Assembly, Fortaleza, Brazil, October 2013) and in accordance with the Medical Research Involving Human Subjects Act (Dutch law: Wet medisch-wetenschappelijk onderzoek met mensen (WMO)). According to the WMO, the study did not require formal medical ethical approval to carry this out in the Netherlands. This was checked by the MREC CMO Arnhem-Nijmegen (file number: 2019-5555). For Scotland, the ethical approval was given by the School of Science and Engineering Research Ethics Committee (SSEREC) at the University of Dundee. Each participant gave his/her informed consent on paper beforehand.

5.4 Results

5.4.1 Participants

In the first round, 44 participants created an account and 40 interacted with the ECA that embodied the MRT (23 NL, 17 SC). A full overview of all demographics can be found in Table 5.1. The mean age of these MRT participants was 65.35 ($SD = 7.35$). Most of them were female (67.5%). They had a good health literacy ($M = 4.35$, $SD = 0.67$) and had a slightly positive attitude towards technology ($M=4.46$, $SD = 1.17$). Their levels of intrinsic motivation were high and their levels of external regulation or a-motivation relatively low when it came to living healthy. Most of them (57.5%) had completed higher vocational education or university-level education. Most were married or living together (75%) and they were quite active with 52.5% being active for more than 2.5 hours per week.

In the second round, 46 participants created an account and 42 interacted with

Table 5.1: Demographics of participants in the MRT.

Demographic	Category	Round 1 <i>M (SD) or N (%)</i>	Round 2 <i>M (SD) or N (%)</i>
Age		65.35 (<i>SD</i> = 7.35)	62.12 (<i>SD</i> = 8.68)
Gender	Male	13 (32.5%)	12 (28.6%)
	Female	27 (67.5%)	30 (71.4%)
Country	Netherlands	23 (57.5%)	24 (57.1%)
	Scotland	17 (42.5%)	18 (42.9%)
Health literacy		4.35 (<i>SD</i> = 0.67)	4.32 (<i>SD</i> = 0.62)
Attitude towards technology		4.46 (<i>SD</i> = 1.17)	4.57 (<i>SD</i> = 1.53)
Motivation to live healthy	Intrinsic motivation	5.19 (<i>SD</i> = 1.11)	5.07 (<i>SD</i> = 0.92)
	External regulation	2.82 (<i>SD</i> = 1.26)	3.14 (<i>SD</i> = 1.19)
	A-motivation	2.28 (<i>SD</i> = 1.45)	2.19 (<i>SD</i> = 1.06)
Level of education	Preparatory secondary vocational education	8 (20.0%)	3 (7.1%)
	Higher general secondary education, pre-university education	9 (22.5%)	13 (31.0%)
	Higher vocational education, university	23 (57.5%)	26 (61.9%)
Living situation	Married/living together	30 (75%)	32 (76.2%)
	Alone	9 (22.5%)	10 (23.8%)
	Other	1 (2.5%)	0 (0.0%)
Self reported physical activity	Not at all	4 (10.0%)	1 (2.4%)
	Not at all, but thinking about beginning	1 (2.5%)	3 (7.1%)
	< 2.5 hours a week	14 (35.0%)	13 (31.0%)
	> 2.5 hours a week in the last six months	12 (30.0%)	14 (33.3%)
	> 2.5 hours a week for more than six months	9 (22.5%)	11 (26.2%)

the ECA that embodied the MRT (24 NL, 18 SC). The mean age of participants was 62.12 ($SD = 8.68$). Although the target population was adults ageing 55 years and older, 4 participants were included that were 40, 42, 47 and 54. We decided to keep these participants included. Most of the participants were female (71.4%). They had a good health literacy ($M = 4.32$, $SD = 0.62$) and a slightly positive attitude towards technology ($M = 4.57$, $SD = 1.53$). This second group also had high levels of intrinsic motivation and relatively low levels of external regulation or a-motivation when it came to living healthy. The level of education was high (61.9% higher vocational or university), they were mostly married or living together (76.2%) and they were quite active with 59.5% being active for more than 2.5 hours per week.

5.4.2 Collected log data

In the first round, 6,077 logged dialogues were collected for all participants, and for the second round 6,222 dialogues were collected for all participants. Pre-processing of those logged dialogues led to 2,384 and 2,210 dialogues with the ECA who embodied the MRT (see Table 5.2). These dialogues amounted to 568 interactions in the first round and 443 in the second round (see Table 5.3).

Table 5.2: Logdata after pre-processing (only from the ECA that embodied the MRT).

Round	N participants	N logged dialogues
	Total (NL, SC)	Total (NL, SC)
1	40 (23, 17)	2381 (1604, 777)
2	42 (24, 18)	2210 (1075, 1135)

Table 5.3: Interactions after pre-processing (only from the ECA that embodied the MRT). Note that multiple dialogues chained together form one interaction.

Round	Condition	N interactions
		Total (NL, SC)
1	User-initiative	295 (207, 88)
	Coach-initiative	273 (186, 87)
	Total	568 (393, 175)
2	User-initiative	238 (125, 113)
	Coach-initiative	205 (108, 97)
	Total	443 (233, 210)

5.4.3 Comparing interaction length

To test the hypothesis that the coach taking the initiative would lead to more engagement and thus longer interactions (H1), a generalised estimating equations (GEE) analysis was conducted for both rounds. The distribution of dialogue steps for both conditions in both rounds can be found in Table 5.4.

Table 5.4: Distribution of the number of dialogue steps in interactions for both conditions in both rounds, and the natural log transform of the number of dialogue steps (which was used in the GEE).

Round	Initiative	N steps	Ln(steps)
		<i>M (SD)</i>	<i>M (SD)</i>
1	User	25.29 (22.53)	2.70 (1.21)
	Coach	22.09 (19.25)	2.47 (1.36)
2	User	28.47 (28.47)	2.90 (1.14)
	Coach	24.91 (20.29)	2.72 (1.20)

Table 5.5: Results for the generalised estimating equations (GEE) analysis for both rounds.

Round	Beta	Std. Error	p	Wald χ^2
1	.239	.1272	.060	3.531 (1)
2	.186	.1640	.256	1.290 (1)

The GEE showed that there was no significant difference in length between interactions in the user-initiative and coach-initiative conditions for either of the two rounds (see Table 5.5). Therefore, we *cannot accept our hypothesis that the coach taking the initiative will lead to longer interactions (H1)*. User-initiative and coach-initiative interactions were of equal length.

5.4.4 Interview results

Three interview questions were included in the interviews for the second round. Of the 42 participants, five participants did not have an interview (11.9%). Two participants indicated that they did not interact with the ECA that embodied the MRT (4.8%), two said they only completed the introduction dialogue with her (4.8%), and one said they did not interact with the coach, but actually did interact twice according to the logs (2.4%).

The other 32 participants had interacted with the physical activity coach. Almost all of them had been exposed to the coach-initiative dialogue condition ($N = 30$), and most of them also had accepted and discussed a suggested topic with

the coach ($N = 28$). However, only 16 of them stated that they had noticed that the coach sometimes took the initiative.

Of the 16 participants that noticed an initiative change, five did not have a preference for either user- or coach initiative and five seemed to like suggestions by the coach (e.g., '*Thought that was good. Also directed me to the other coaches on a few occasions.*' [P13]), motivating that it provided the option to not have to think about something and that the coach was pleasantly surprising. One participant stated that it gave a stimulus to discuss something and indicated that some people need that (himself included).

Six participants were not interested in suggestions (e.g., '*I was more interested in my progress.*' [P19]) or thought they were not personalised enough (e.g., '*Not sure the character knew enough about me to do that. Seemed a bit random.*' [P09]). Their opinion also seemed to be reflected in their responses to the suggestions, for example, the person who thought it was pleasantly surprising and enjoyed the interaction accepted 13 out of 14 suggestions, while the person interested in their progress accepted 5 of the 12 suggestions, chose the '*I want to discuss something else*' reply 6 times, and cancelled the dialogue after 1 suggestion.

In two cases there was a clear discrepancy between logged response to suggestions and participants' expressed opinion. For example, one did not like suggestions by the coach, and motivated this by stating that she wanted direct and specific coaching advice and not social conversations. However, her log data did show that she agreed to discuss 75% of the suggested topics. Another stated that she always had a topic in mind that she wanted to discuss and that she therefore did not accept suggestions by the coach. She however had agreed to discuss all suggested topics about coaching, and only rejected all social topics.

To summarise: about half of the participants did not notice the coach taking the initiative, which suggests that the change in initiative was perceived as a *natural variation in the interaction with the coach*. The other half of the participants did notice it, but most of them thought it was fine or liked it. This indicates that they were *not bothered by the suggestions of the coach*.

5.4.5 Acceptance of suggested topics

After finding that the coach suggesting a topic did not lead to longer or shorter conversations and that the coach taking the initiative was either not noticed or appreciated, we investigated how often the suggestion of a topic by the coach was accepted by users.

Table 5.6: The acceptance for topic suggestions by a coach in the coach-initiative condition. We present acceptance and rejection numbers for both rounds separately, and both combined. Furthermore, for the two separate rounds we also provide details on the manner in which a topic was rejected, namely: *changed* when the user selected 'I want to discuss something else', *goodbye* if they choose to end the interaction with 'goodbye', *cancelled* if they closed the coach's speech-bubble, and *nothing* if a user did not respond at all.

Round	Topics	N	N accepted	N rejected	N rejected reason			
					Changed	Goodbye	Cancelled	Nothing
1	Social	38	26 (68.4%)	12 (31.6%)	n.a.	5 (13.2%)	7 (18.4%)	0 (0.0%)
	Coaching	235	187 (79.6%)	48 (20.4%)	n.a.	16 (6.8%)	27 (11.5%)	5 (2.1%)
	Overall	273	213 (78.0%)	60 (22.0%)	n.a.	21 (7.7%)	34 (12.5%)	5 (1.8%)
2	Social	72	35 (48.6%)	37 (51.4%)	25 (34.7%)	7 (9.7%)	4 (5.6%)	1 (1.4%)
	Coaching	133	92 (69.2%)	41 (30.8%)	23 (17.3%)	3 (2.3%)	14 (10.5%)	1 (0.7%)
	Overall	205	127 (62.0%)	78 (38.0%)	48 (23.4%)	10 (4.9%)	18 (8.8%)	2 (0.9%)
Both	Social	110	61 (55.5%)	49 (44.5%)				
	Coaching	368	279 (75.8%)	89 (24.2%)				
	Overall	478	340 (71.1%)	138 (28.9%)				

In the first round, the coach took the initiative by suggesting a topic to discuss in 273 interactions. As can be seen in Table 5.6, the overall acceptance rate for these suggestions was high with 213 accepts (78.0%). In 60 cases (22.0%), the suggested topic was not accepted by the user. We must note, however, from the 16 'goodbye' rejections, 8 were in response to a dialogue where 'no additional information was available', and 4 of the 27 cancellations were in response to that statement.

Overall, most suggested topics were well received in the first round. The acceptance rate was high for the topic *Introduction* (26 out of 38; 68.4%), *Goal-Setting* (32 out of 35; 91.4%), *Gather Information* (20 out of 22; 90.9%), and *Inform* (69 out of 79; 87.4%). The *Sensors* topic was accepted less by participants (66 out of 99; 66.7%). Of those 33 rejections, 15 were cancellations in response to the suggestion to connect a sensor; mostly by participants that on a later prompt did agree to discuss that topic.

In the second round, the coach took the initiative by suggesting a topic to discuss in 205 interactions. As can be seen in Table 5.6, the overall acceptance rate for suggestions in this round was lower than that of the first round (62.0%). There was also a large difference in acceptance between suggestions for social and coaching topics (48.6% and 69.2%, respectively). Furthermore, in this round, participants had the option to indicate that they wanted to discuss something else. This meant that even though the suggestion by the coach was not accepted, the conversation still continued.

Overall, suggestions for coaching topics were again well received. *Goal-Setting* was accepted 2 out of 3 times (66.7%), and the *Feedback* topic 17 out of 18 (94.4%). While the *Gather Information* topic was accepted 18 out of 38 times (47.4%), its suggestion also resulted in 9 changes of topic (23.7%) and 9 cancellations (23.7%). The *Inform* topics were accepted 31 out of 42 times (73.8%), and 7 times the topic was changed (16.7%).

Suggestions for social topics were not always welcomed. Participants accepted the *Introduction* topic 13 out of 23 times (56.5%), and the suggestion for the coach to *Share a Background Story* about herself was only accepted 22 out of 49 times (44.9%), while a change of topic was requested 19 times (38.7%).

To get a sense of rejections in the user-initiative condition, the percentage of cases in which the coach only got to make one statement in that condition was also computed. For the first round, this was the case in 11.2% of interactions (33 out of 295), and for the second round 9.7% (23 out of 238) of the interactions were rejected, which amounted to 10.5% of user-initiative interactions over both rounds.

The *high acceptance rates* for the coach-initiative condition indicate that even

though suggestions by the coach did not lead to longer interactions, the suggestions that the coach made were suitable. We therefore accept our second hypothesis (*H2: Because topic suggestions are tailored to the individual, participants will accept suggested topics more often than not.*).

5.4.6 Demographics and individual acceptance

Finally, a possible influence of demographics on acceptance of coach-suggestions was explored. From both rounds, 18 participants had had 10 or more coach-initiative interactions with the MRT coach. The correlations between ordinal or continuous demographics and the percentage of accepted suggestions can be found in Table 5.7. As can be seen, there was a moderately strong correlation between attitude towards technology and percentage of accepted suggestions ($r_\tau = 0.48, p = .007$).

Table 5.7: Kendall's Tau correlation of users' demographics with percentage of accepted suggestions.

Demographic	Correlation	
	r_τ	<i>p</i>
Age	.09	.620
Self-reported physical activity	-.06	.743
Health literacy	-.11	.561
Education	-.12	.551
Attitude towards technology	.48	.007
Intrinsic motivation	.12	.492
External regulation	.06	.732
A-motivation	.21	.255

The Mann-Whitney U-test for country showed that there was no significant difference ($U=30.0, p=.574$) between the acceptance of Dutch ($Mdn = 0.78, IQR = 0.55-0.82$) and Scottish ($Mdn = 0.59, IQR = 0.39-0.90$) participants. For gender there also was no significant difference ($U = 34.0, p = .851$) between the acceptance of Male ($Mdn = 0.75, IQR = 0.55-0.84$) and Female ($Mdn = 0.70, IQR = 0.46-0.89$) participants.

Overall, these findings show that participants with a more positive attitude towards technology had a *higher percentage of accepted suggestions*, which might mean that acceptance of topic suggestions by an ECA is dependent on how open participants were to using a health coaching application to begin with. We therefore accept our hypothesis that there is an influence of demographics on topic acceptance (H3).

5.5 Discussion

In this article we investigated the influence of a virtual coach taking the initiative and suggesting a relevant (tailored) topic to discuss on users' interaction with that coach. To that end, we implemented two versions of a coach's dialogue (coach-initiative and user-initiative) and compared these in a micro-randomized trial where participants interacted with the coach in their daily life over a longer period of time.

Our first hypothesis was that the coach taking the initiative by suggesting a relevant topic to discuss would lead to longer interactions than when a user had to select a topic themselves. However, our results show that there is no difference in length between interactions in the user-initiative and coach-initiative conditions. While this does not support that hypothesis, our findings do show that a coach taking the initiative was equally engaging as asking the user to indicate what they would like to discuss. Likewise, the results from the interviews taught us that participants either did not notice that the coach took the initiative or were fine with it. While this lack of awareness of the manipulation seems to stand out, a similar finding was reported by Olafsson et al. (2019). They performed a manipulation in which they removed the possibility of participants to be able to respond negatively to suggestions by a health coaching ECA, and found that participants did not notice that lack of choice. Furthermore, while there might not have been a difference in dialogue length, the suggestions by the coach might have had an influence on relational parameters that we were not able to measure after every interaction (e.g., perceived helpfulness as found by Xiao et al. (2002), or preference for an ECA (Olafsson et al., 2019)).

The equality between the coach-initiative and user-initiative conditions, both objectively (interaction length) and subjectively (user experience), is interesting, since they are actually quite unequal when it comes to 'freedom of choice'. That is, where the algorithm in the coach-initiative condition only suggests one specific topic, in the user-initiative condition users have the full set of topics to choose from. Having a coach suggest a relevant topic to discuss could lead to higher engagement with DBCIs because of personal relevance and tailoring of content, and perhaps also because of novelty, a sense of narrative and guidance (Perski et al., 2017). However, suggesting a specific topic has to be done right. That is, if that single topic you suggest is not relevant for the participant, they will rightfully reject your suggestion. The high acceptance rate (71.1%) for topic choices by the coach, suggests that our algorithm did select topics that were mostly relevant and suitable for users. This also supports our second hypothesis. One might wonder if users would just accept all suggestions, but the difference between the acceptance of social (55.5%) and coaching topics (75.8%) indicates that users did care which topics were suggested. Overall, we conclude that our

underlying coaching engine component seemed to have selected topics with a high enough relevance for users. Potential factors that could have influenced topic acceptance could be the task-mindedness and independence of users, since some indicated in the interviews that they already had a clear purpose in mind when starting and interaction. We therefore advise to include these in future research.

We see at least two possible directions for future research that relate to the improvement of the coaching engine's topic selection and the more general content design processes. First, the balance between social and coaching topics could be improved. While our results show that suggestions for social topics were not always welcomed, previous research found that background stories and other relational behaviours are important for enjoyment and engagement with an ECA (Bickmore et al., 2010; Trinh et al., 2018). We therefore suggest that instead of simply adjusting the frequency of social topic suggestions, a system could learn a user's interest in social dialogue by taking into account their responses to social comments. These responses can be measured when a topic is suggested, but also for all similar remarks or social sidesteps in dialogues about coaching topics. Furthermore, other predictors may be used in modelling a participant's interest in social interactions. One example of such a predictor in the context of social robots was whether a participant greeted a robot before interacting (Lee et al., 2010).

Second, a similar approach could be applied to improving the suggestion of coaching subtopics. The ECA and the user could have an explicit discussion on preferences for coaching style or strategy (e.g., as investigated in **Chapter 3**). For example, when it comes to deciding what to discuss or the balance between coach- and user-initiative. These investigations could benefit from including lessons from research on shared decision making (Joseph-Williams et al., 2014; Zhang & Bickmore, 2018). Another option could be to implement classification functionality or self-learning mechanisms to determine different types of users based on previous digital health coaching research (e.g., type of motivation to live healthy (van Velsen et al., 2019) or stages of change (de Vries et al., 2016)). This would provide support for further tailoring of initiative and strategies. Such models could benefit from our finding that participants with a more positive attitude towards technology could have a higher acceptance rate for suggested topics, which supported our third hypothesis (influence of demographics on acceptance).

5.5.1 Strengths and limitations

A strength of this study was that it was a long-term study conducted in users' daily life. Participants were asked to use the application at will over a period of at least

four weeks, which could be extended with another four weeks. This meant that every recorded interaction with the system was a) voluntary and without possible influence of a researcher's presence, and b) these interactions extended past the first two weeks in which a novelty effect can still be present (Nijland, 2011). Furthermore, we evaluated a novel implementation of an interaction condition in which coaching dialogues were automatically tailored at the topic level (by introducing a coaching engine), and we conducted the (to our knowledge) first micro-randomized trial in the context of embodied conversational coaches.

There were also some limitations. First, participants were recruited through advertisements in local newspapers and on social media, which meant that selection bias was an issue as discussed in the published study protocol (Hurmuz et al., 2020). This probably caused our participants to have relatively high scores on intrinsic motivation and health literacy, and relatively high levels of education. These are all factors that are associated with participants being more active in managing their own health, which might have influenced the way they interacted with the system. Second, during the study, the COVID-19 outbreak reached the Netherlands and Scotland. From that point on, the study was performed remotely, that is, materials were sent to participants by post and interviews were conducted by phone. Nevertheless, we have no indication that these procedural changes affected participants' interactions with the application.

5.6 Conclusion

Tailoring coaching conversations with embodied conversational agents (ECAs) has the potential to increase the engagement of users with those coaches, which is deemed a prerequisite for desired behaviour change. The main finding from this micro-randomized trial is that coaching conversations with ECAs can be automatically tailored on a topic level, and that the resulting interaction is perceived as a natural with a high user acceptance of suggestions by the coach. This is an important step towards more intelligent and engaging health coaching applications. Future work should investigate how to further improve the automatic topic suggestion process, and how the amount of initiative, the coaching strategies and the coaching style applied by the coach could be adjusted to specific types of user.

Acknowledgements

The authors would like to thank Dominic De Franco, Ellis Oude Kempers, and Katrien Fischer for their part in participant recruitment and conducting and transcribing the interviews.

Chapter 6

General discussion

Overview

Adopting a healthy lifestyle can help prevent or mediate illnesses. However, changing behaviour can be difficult. Digital health applications with conversational agents in the role of a coach might provide support when needed, but the conversations with these virtual coaches are sometimes too general and not personally relevant enough. Tailoring content is regarded to be a solution for this problem in eHealth applications, but has not yet been extensively investigated for interactive two-way communication over a longer period of time, such as in the context of coaching conversations with conversational agents.

The aim of this thesis was to investigate **how we can tailor users' coaching conversations with conversational agents to improve engagement**. This chapter places the results from the previous chapters in a broader perspective. First, a short overview is given that summarises the chapters. Second, the main findings are discussed on three themes, namely: 1) content and system design; 2) tailoring content and conversation; and 3) evaluation of tailored content. Third, we discuss open directions for future research and considerations for the use of conversational agents in eHealth.

The work in this thesis began with the development of specialised tools that facilitate dialogue authors in the creation of dialogue content for conversational agents and that supported tailoring of those dialogues (**Chapter 2**). With these tools, the investigation into tailoring of coaching conversations then continues in **Chapter 3**. In general, tailoring of coaching content in eHealth applications is performed by adjusting the communication and content to certain information about the user. The first study described in **Chapter 3** investigated if a tailoring approach from the field of eHealth applications can be applied in the context of coaching dialogues.

Where **Chapter 3** focused on tailoring the coaching strategies that are underlying to a coaching dialogue, the last two chapters focus specifically on tailoring coaching conversations during interaction. **Chapter 4** began with the introduction of a five-step process that describes five levels of detail on which coaching dialogues can be tailored: domain, topic, action, dialogue act and utterance. It then focused on one of these levels specifically, namely the topic level. The chapter describes the construction of a blueprint topic model that is based on literature and evaluated through a card-sort with experts. This blueprint model can be used to define a topic structure for use in implementing automatic selection of tailored topics. Finally, **Chapter 5** evaluated the influence of such automatic topic selection on interaction length as a measure of engagement and investigated the acceptance of suggested topics by participants.

How can we tailor users' coaching conversations with conversational agents to improve engagement?

As stated throughout this thesis, tailoring coaching conversations with conversational agents has the potential to improve the interaction with those agents. Presenting users with coaching dialogues that match users and their interests in terms of content could lead to increased engagement of the user with the application. This type of engagement is sometimes referred to as "*Little e*" as defined by Cole-Lewis et al. (2019) or *micro engagement* as defined by Yardley et al. (2016). Such engagement with the application in turn is a prerequisite for engagement with the health behaviour, or "*Big E*" (Cole-Lewis et al., 2019), and intervention goals, or *macro engagement* (Yardley et al., 2016). This engagement with health behaviour hopefully leads to adherence and behaviour change in the long-term.

However, while tailoring is actively investigated when it comes to eHealth applications and one-way communication (e.g., sending messages or using apps), much is still unknown when it comes to tailoring interactive two-way coaching dialogues. Therefore, the aim of this thesis was to investigate how this could be done for coaching conversations. The following three subsections discuss the main insights following three essential aspects of the tailoring process, namely: 1) content and system design; 2) tailoring content and conversations; and 3) evaluation of tailored content.

Content and system design

Tailoring of coaching dialogues requires dialogue content that can be tailored. In health coaching, the provided advice should have an impact on users' health. Thus, it is important that dialogue content is correct, and follows certain standards (e.g., for physical activity or nutrition), that it is designed using intervention approaches that are known to be effective, and that the possible resulting dialogues can be verified and validated beforehand to avoid adverse effects. That is, various suggestions could be included in dialogues and potential combination of these suggestions should not negatively impact users' health. In both the Council of Coaches project and this thesis, the choice for dialogue scripts as a representation ensured that it was possible to have such a clear overview of dialogue outcomes. While the restriction of user responses to predefined reply-options in these dialogue scripts could be perceived as a limitation of the system, it ensures that users can only select responses to which the agent can provide an answer. The use of reply options was also important since it removed the risk of incorrect interpretation that is associated with free text or speech input (e.g., as discussed by Bickmore, Trinh, Asadi et al. (2018)). The developed

tools (**Chapter 2**) supported the process of creating scripted dialogue content by allowing anyone (not just technical developers, but also non-technical domain experts) to author and test-run dialogues. These tools were developed and used in the Council of Coaches project, but the use cases in **Chapter 2** show that they are now also applied in a number of other European and national projects.

As discussed in **Chapter 2**, developing conversational agents for health coaching is a process which does not only require expertise on the coaching domain, but also a certain amount of creativity. Often multiple iterations per dialogue are needed. Furthermore, it is important for coaching dialogues that the advice, content and the underlying approach (or strategy) are proven and based on principles from the literature (Michie et al., 2011; Scholten et al., 2017). For many domains, existing guidelines and interventions in the human-human context can be found. There is also a plethora of behaviour change theories from the behavioural psychology field that can be chosen as a basis for content development, such as the Transtheoretical Model (Stages of Change model) (Prochaska & Velicer, 1997), Social-Cognitive Theory (Bandura, 1991), Goal-Setting Theory (Locke & Latham, 2002), and the Health Action Process Approach (HAPA) (Schwarzer et al., 2011). The more specific Behaviour Change Technique (BCT) taxonomy (Michie et al., 2013) includes a set of 93 techniques. Then there are approaches from the persuasive technology field, such as the Persuasive Systems Design (PSD) model (Oinas-Kukkonen & Harjumaa, 2009), the Computerized Behavior Intervention (Combi) model (Klein et al., 2013), and specific persuasive features, such as the set listed by van Velsen et al. (2019), which was used in **Chapter 3**.

Previous research has shown that applications that include BCT techniques are associated with higher engagement (Perski et al., 2017), and there is ongoing work in the behavioural psychology field to gather evidence for the mechanisms of action that make sure that these techniques are effective (e.g., Carey et al. (2018)). However, much is still unknown when it comes to what works in health coaching and why. Even though many eHealth interventions incorporate principles from the literature, learning from reported studies remains a challenge due to lack of detail when it comes to content description. Where Michie et al. (2013) reported this as one of the motivators for their development of the BCT taxonomy, the same issue is still being reported in recent reviews on conversational agents (El Kamali et al., 2020; Kramer et al., 2020). Studies tend to list techniques and theories on which content was based, but it is often difficult to gauge the exact (conversational) execution. For example, there are many ways in which a dialogue for a step goal could be written. It could involve a negotiation about the number of steps for a physical activity goal (e.g., as in Snaith et al. (2018)) or the agent could also provide background information and insight into why a specific goal is suggested and should be accepted. While these are just two variations, they show that ‘a goal-setting dialogue’ or ‘based on goal-setting theory’ are insufficient descriptions for replication.

Furthermore, incorporation of proven principles from previous interventions in health coaching dialogues requires additional careful implementation and evaluation – in a sense, they need to be ‘translated’ to the dialogue domain as demonstrated with the strategies in **Chapter 3**. In addition to the underlying coaching principles and content, coaching dialogues with virtual agents need to be more than simply a set of dry instructions or questions that are turned into a monologue. In a sense, they need to be like a good book, movie, or video game. Structure is important ('building up the plot'), as well as interesting characters and a good flow. They also need to be dynamic and coherent over multiple sessions (Bickmore, Trinh, Asadi et al., 2018) and should allow for adjustment to individual users. For example, a dialogue in which the agent explains why being physically active is beneficial, could take the form of a simple explanation, with the user being able to ask the agent to elaborate on or repeat information (e.g., as in Bickmore and Giorgino (2006)). But, if a user indicates that they are already familiar with the topic, a quiz like format might be more suitable (e.g., “How many steps do you think it takes to burn the calories of a pizza?”).

There are many different domains for which ‘translation to the dialogue format’ can be investigated and common principles and differences between domains should be researched for the conversational agent context. For example, in **Chapter 5** the topic selection algorithm was only implemented for the physical activity coach, but in the Council of Coaches project, the dialogue content was for all coaches was structured similarly. The common topics are all included in the blueprint topic model presented in **Chapter 4**. Furthermore, literature from the persuasive technology, eHealth, relational agents and behavioural psychology fields shows that social interaction, a working alliance, and a relationship in general between coach and coachee are important for engagement, adherence and retention (Bickmore, Trinh, Asadi et al., 2018). This means that a character, personality and background story need to be defined for a coach and that this has to be incorporated consistently in all dialogues. Furthermore, social dialogues such as those about background stories can offer a short break from the serious coaching topics. For example, the agents in the Council of Coaches application (used in **Chapter 3** and **Chapter 5**) could share information about their shoe collection, grandkids, or even an obsession with cheese. The blueprint topic model presented in **Chapter 4**, thus not only provides a structure for coaching topics, but also includes social and meta topics.

Finally, an observation on the amount of content that is needed. Users will go through available dialogue content at an incredible speed when compared to the time needed for development. That is, a dialogue can take multiple days to write, fine-tune, and verify, especially with tailoring elements included, but it can be completed by a user in a few minutes. In addition, users will only interact with a small number of possible dialogue steps because they only use one of the potential paths through a dialogue. To cope with this challenge, there are some

system elements that can help ensure that enough content stays available in a long-term study. An example of such a principle that was included in the Council of Coaches application is ‘time gating’ (often used in video games), which is the principle of content becoming available after a certain amount of time has passed. For example, the chapters of the background story dialogues in **Chapter 5** became available with one week intervals. Another applied principle was the inclusion of informative widgets that ‘keep their value’, such as a recipe book that was combined with an agent assisting with recipe selection (“Would you like a fish, meat or vegetarian recipe?”), or an activity book that showed the steps per day based on Fitbit data.

Tailoring content and conversations

Various approaches exist for tailoring coaching content when it comes to eHealth applications. Tailored persuasive features can be used to adjust the content of websites or apps (e.g., van Velsen et al. (2019)), motivational messages can be generated that are completely adjusted to the user in terms of timing, intent, content, and representation (Achterkamp et al., 2013; de Vries et al., 2017; op den Akker et al., 2014), and users can be presented with personalised goals and feedback (Lentferink et al., 2017). These adjustments can be made using a dynamic tailoring approach (based on an evaluation of variables before message generation) or a static tailoring approach (using information collected at baseline) (Krebs et al., 2010), and tailoring can be implicit (deduced) or explicit (through direct questions) (Fan & Poole, 2006). However, in contrast to static tailoring approaches, dynamic tailoring approaches in interventions have been found to have greater long-term effects which remained significant at follow-up (Krebs et al., 2010).

When it comes to coaching conversations, tailoring tends to often be limited to the suggestions of personalised goals or the adjustment of communication on a sentence level. This thesis therefore focused on higher level tailoring of coaching conversations. The five-step tailoring process that is presented in the introduction of **Chapter 4** defines that conversations can be tailored on a *domain, topic, action, dialogue act and utterance* level. The strategies in the experiment in **Chapter 3** can guide the selection of tailored content on the higher levels of that five-step process, such as making some topics more relevant to discuss for the strategy than others. The blueprint model presented in **Chapter 4** and automatic topic selection approach evaluated in **Chapter 5** focus on the topic level. Meanwhile, the scripted dialogues used to discuss the various topics still made use of conventional tailoring approaches, such as adjusting step goal suggestions based on users’ age (following guidelines of the WHO (World Health Organization, 2011a, 2011b)) and the inclusion of users’ first name.

The division of tailoring coaching dialogues into a five-step process is not just a theoretical exercise, but also a practical setup for implementing tailored systems. Layering a series of simple choices can result in emergent behaviour that seems intelligent, but which is interpretable and relatively simple to explain. Both of these aspects are highly important. Intelligent behaviour makes the conversations with the agent seem more like human-human interaction, which is essential when it comes to social incentive – the main reason for using conversational agents in the first place. Interpretability is essential to have an insight into why certain statements are made by the agent – both from a safety perspective and to potentially be able to explain such reasons to the user. Meanwhile, the tailoring process on each layer can be as simple or complex as desired. For example, in the first step a choice for a domain can be made by a healthcare professional that recommends a certain coach (explicit static tailoring). But, domain selection can also be the result of a set of intake questions (as it was in the Council of Coaches application), or it could be an ongoing discussion of priorities between a coach and a user if the coach is able to give advice on both physical activity and nutrition (making it more dynamic).

While the investigations in this thesis open up a range of possibilities for future research, the increase in likeability that we saw for agents presenting serious content in **Chapter 3**, the preference for tailored coaching strategies (**Chapter 3**), and the high acceptance of suggested topics in **Chapter 5** — with users being critical of suggestions about background stories — shows that providing the right content at the right moment matters. The five-step model gives structure to the tailoring process of coaching dialogues and the blueprint topic model and topic selection algorithms are a first step to fill the gap of dynamic tailoring on higher levels, but there are still many opportunities to improve the processes in each of these steps and to tailor interaction with conversational agents in general. Tailoring of coaching conversations should also be investigated for combinations between tailoring approaches in these steps. For example, in the first step, when it comes to selecting a domain, the balance between coaching on multiple domains could be investigated since changing all behaviour at the same time tends to be too much of a challenge. This also requires careful attention to the setup of domains and models, since multi-domain coaching tends to be implemented through separate activities for different domains instead of integrated models (El Kamali et al., 2020). When it comes to tailoring topics, the automatic selection of topics could be improved. For example, the system could learn users' preference for social topics from their response to suggestions for such topics or social statements in other dialogues (as discussed in the discussion of **Chapter 5**). Finally, the discussion of each individual topic in itself can be further tailored on the action, dialogue act and utterance steps. In the tailoring of all these steps, insights from the content development process and literature, such as suitable communication styles and intervention techniques could (and should)

be integrated. In addition to the content of the communication, the underlying delivery styles, coaching approaches, strategies and other aspects of interaction could also be tailored.

Evaluation of tailored content

While evaluation approaches were not the main focus of this thesis, evaluation is a challenge when it comes to interactive adaptive systems (Paramythis et al., 2010), and evaluation of tailored coaching conversations is no different. That is, during evaluation, the tailoring process is not the only factor that can influence users' perceptions and behaviour. Users can be influenced by an agent's looks (e.g., through first impressions (ter Stal et al., 2019)), errors in the user interface or even the mismatch between their own expectations of the system and what the agent is actually capable of (the adaptation gap (Komatsu et al., 2012)). Paramythis et al. (2010) distinguished five layers of adaptation, namely: collection of input data, interpretation of the collected data, modelling of the current state of the world, deciding upon adaption, and applying adaption. While each of these layers should be evaluated explicitly, they also note that it is not always possible to isolate them. When it comes to conversational agents, there is no established method of evaluation (Diederich et al., 2021; Laranjo et al., 2018; Milne-Ives et al., 2020; Tudor Car et al., 2020). This is partially due to the large variation in implementations (e.g., in input and output modalities, roles, underlying algorithms, and presence and realism of embodiment), which also makes comparisons between different studies complicated (Diederich et al., 2021; Scholten et al., 2017). This shows that there is a need for replication studies and validated outcome scales (e.g., when it comes to design features for relationship quality (Loveys et al., 2020)). Such efforts are currently ongoing. For example, when it comes to engagement, Short et al. (2018), Oertel et al. (2020) and Trinh et al. (2018) created overviews and defined measures of engagement in eHealth and human-agent interaction, respectively. For eHealth a number of engagement scales were defined, such as the eHealth Engagement Scale (Lefebvre et al., 2010), the TWente Engagement with Ehealth Technologies Scale (TWEETS) (Kelders et al., 2020) and the DBCI Engagement Scale (Perski et al., 2018). Within the agent community there is an ongoing initiative to come to questionnaire items for the evaluation of artificial social agents by the 'Artificial Social Agent Evaluation Instrument' workgroup. This workgroup showed the wide variance of evaluation metrics used in agent studies in a literature survey (Fitrianie et al., 2019) (189 constructs in 89 questionnaires for 81 studies), from which they created a set of 19 unifying constructs (Fitrianie et al., 2020), and recently presented the first results of creating questionnaire items (Fitrianie et al., 2021) (131 expert-generated and content-validated items).

The two user studies (**Chapter 3** and **Chapter 5**) in this thesis focus on different

aspects of the tailoring process and in effect also evaluated these differently. In the study in **Chapter 3**, a validated relationship between motivation to live healthy and persuasive features (van Velsen et al., 2019) was used as a basis for strategy design. The strategy dialogues were then written and validated by a team of four authors. Furthermore, control questions were included for the likeability of the presenting agent (taken from Acosta and Ward (2011)) and the usability of the system (Tedesco & Tullis, 2006). In the evaluation in **Chapter 5** on the other hand, a single set of dialogues with a specific topic structure (following the blueprint model in **Chapter 4**) was developed for both conditions, and only the start of the dialogues for each topic was changed between conditions. In this last study, interview questions were added to the second round, which provided valuable insights into users' reasons as to *why* they might accept a suggestion to discuss a topic or not. Furthermore, they showed that some users did not notice the change of initiative between conditions, and that there can be a difference between actual interaction and users' opinions, since some statements by users did not fully match their interaction data. Such addition of qualitative elements to quantitative studies is also suggested for the evaluation of conversational agents for health by Milne-Ives et al. (2020). Furthermore, the influence of individual parameters such as demographics on interaction should be included in evaluations, since this information might provide valuable insights for further tailoring (Diederich et al., 2021).

Many studies with conversational agents, especially when they have a physical embodiment or are computationally heavy, still take place in the controlled settings of a lab (Griffin et al., 2020; Laranjo et al., 2018). Such evaluations in controlled conditions provide valuable information and should be performed during development, for example, to investigate relations between personalisation of conversational agents and user experience (Kocaballi et al., 2019). The user studies for the Council of Coaches application that were conducted previous to and in parallel with the studies in this thesis were also conducted in such a controlled environment. However, behaviour change is a process that takes place in users' daily life, over a longer period of time and many factors can influence this process. In addition, the engaging effects of tailored interactions might also be more clear over a longer period of time as use of eHealth applications often becomes less when the novelty effect wears off (Nijland, 2011). Tailoring of interactions with conversational agents for health should therefore also be evaluated in realistic (daily life) conditions with use over a longer period of time. Both user studies in this thesis were performed outside of the controlled conditions of a lab. While the study in **Chapter 3** still featured a short interaction, in the study described in **Chapter 5** users interacted with the system at will for a period of 4-8 weeks. However, such evaluations do come with challenges in addition to the setting being less controlled. A first major challenge is that they take a lot of time and resources to conduct. That is, the system needs to

have a high enough technology readiness level (TRL) (Mankins, 2009), sufficient dialogue content needs to be available, participants need to be recruited that are willing to participate for a longer period of time, and support should be available (e.g., for performing the intakes, handing out materials, answering questions, and conducting the interviews).

A second major challenge with long-term daily life evaluations is the study design. As discussed throughout this thesis, many aspects of a conversational agent's dialogue can be tailored in a variety of ways. However, classical evaluation designs tend to be suitable for comparing a limited set of conditions and parameters. A classic example of such an evaluation is the comparison of a group that receives a tailored intervention versus a control group (e.g., in randomized controlled trials (RCTs)). This often introduces the need for two versions of an application to be developed (tailored versus non-tailored). Alternatively, the control group can be placed on a 'waiting list', but in some cases starting treatment is time-sensitive (e.g., when it comes to conversational agents for mental health (Callejas & Griol, 2021)). A micro-randomized trial (Klasnja et al., 2015), as conducted in **Chapter 5**, allows for a more specific evaluation of intervention features, and can take into account context parameters and a time effect. Still, the evaluation of tailored components in applications remains challenging and the number of variations that can be tested at once limited. All in all, choices for evaluation of tailored content should be adjusted to the phase of development and the maturity of the system, and should additionally involve short-term and long-term studies that include qualitative and quantitative measures.

Considerations for future work

The use of conversational agents in healthcare is still in its infancy (Tudor Car et al., 2020). Literature suggests that conversational agents could be a complimentary element to check-ins with healthcare professionals (Callejas & Griol, 2021), and in that manner can form a hybrid system where technology supplements healthcare services (Tudor Car et al., 2020), self-guided interventions (Scholten et al., 2017) and other technologies (Montenegro et al., 2019). Such a hybrid collaboration between healthcare professionals and conversational agents is important in curative applications of conversational agents, and might be the best option with the capacities of the current generation of agents.

However, conversational agents should still be able to function independently in between sessions with health care professionals – especially when the aim is to one day apply them as stand-alone solutions in prevention scenarios. Applying conversational agents as such stand-alone technology does require careful consideration and provides some additional challenges for the development process, especially if the focus is on improved tailoring, increased intelligence and influencing users

to follow suggestions and advice. The following five subsections address some of these considerations and challenges for future work.

Safety, verification and validation

As previously stated, a conversational health coach provides suggestions and advice that aims to have an effect on users' behaviour, and consequently their health. This means that the dialogues, their content, and the potential resulting advice to users should be verified and validated. The use of scripted dialogues that are tailored on a high level, as described in this thesis, has the advantage that it is possible to perform such verification and validation of resulting dialogues (as also reported as an advantage for rule-based approaches by Callejas and Griol (2021)). This means that combinations of different dialogue scripts can be made in a relatively safe manner since the direct context of utterances does not change. While conversational agents based on black box machine learning approaches can produce suitable responses, these suggestions can be less predictable and it is more difficult to verify the reasoning behind statements. This could potentially lead to unintended consequences (Laranjo et al., 2018). Furthermore, a large amount of training data is often required to train such models, which often is not available for dialogues in a health context due to privacy concerns.

However, the desire for improved tailoring of coaching conversations with health coaches also suggests that even scripted dialogues will have to become more and more dynamic. With an increase in such dynamic tailoring, it will inevitably become more challenging to check, verify and validate the resulting dialogues. The dialogue authoring tools, five-step process of tailoring and blueprint topic model introduced in this thesis could help in that process by supporting the gradual introduction of these dynamic components. For example, at the start of the design process, the topics to be addressed in a system could be defined in a session where experts use the blueprint topic model to design the dialogue structure and topics. Once a plan has been formed, domain experts can write scripted dialogues that are highly controllable, and can be validated and used for initial user testing. Or, if available, existing examples of coaching dialogues between healthcare experts and patients could be used to write these scripts, as were collected in the Council of Coaches project as the *Patient Consultation Corpus* (Snaith, Conway et al., 2021).

In a next phase, the scripted dialogues can then be used to develop and test the tailoring processes for domains and topics, while simultaneously serving as examples for development of more dynamic dialogue generation methods that are still interpretable (e.g., as demonstrated for a goal-setting dialogue in Snaith et al., 2020). Limiting the scope of these dynamic dialogues to a dialogue model for each topic makes them easier to validate than a model that represents

"all possible things an agent should be able to say". Dynamic models can be developed for all topics, but an application could also use a mix of scripted and dynamic dialogues. Of course, a prerequisite for such a development approach would be that the coaching system (or platform) has various modules that can be used interchangeably. This is however possible, as demonstrated in the Agents United Platform (Beinema et al., 2021), which features a single topic selection engine, two different dialogue engines that can be used interchangeably, and even two methods for generating agents' 3D embodiments that can be used simultaneously if desired.

The process of gradually making conversational agent applications more dynamic with intermediate evaluations could also support the formal process of proving that an application is safe. In addition to moral and ethical considerations on user safety, this is relevant when it comes to medical ethical approval for studies and CE marking. With the new European Medical Device Regulations (MDR) (The European Parliament and Council, 2017), the definition of medical device has been broadened and many software applications have been reclassified to constitute higher risks. For conversational agents specifically, this means that they can now be considered a Class II medical device.

Autonomy and trust

Developing coaching applications that are designed to optimally persuade a user of performing healthy behaviour and designing conversational agents to be trustworthy is one thing, but it might – in an extreme situation – increase the risk of persuading a user to act against their own interests. Trust in an agent's capabilities, data privacy and confidentiality all have an influence on the reliance on and cooperation with that agent (Rheu et al., 2020). An important point of attention for trust in a system is that it should match the system's actual capabilities. However, there are still many applications that overly focus on enhancing trust instead of matching it to actual capabilities (Rheu et al., 2020). It is therefore important that a conversational agent raises an appropriate amount of trust (Snaith, Nielsen et al., 2021) and that users' autonomy is ensured (Kamphorst, 2020).

An essential element in safeguarding autonomy and trust is clear communication about an agent's processes and aims. For ensuring autonomy in e-coaching systems, Kamphorst (2020) makes five recommendations, namely: ensure ongoing consent, reveal the reasoning behind, increase user awareness of system fallibility, offer reassessment opportunities and promote reflection on suggestions. For calibrating trust to actual capabilities, Rheu et al. (2020) list being clear about what a system does with data, how it makes its decisions, and what its limits are. Recent reviews and literature on conversational agents for health make similar

recommendations (Bickmore, Trinh, Asadi et al., 2018; Scholten et al., 2017). These recommendations are also in line with the European Union's general data protection regulation (GDPR) (The European Parliament and Council, 2018), medical device regulations (MDR) (The European Parliament and Council, 2017), and the proposed artificial intelligence act (European Commission, 2021). In 2020, one example that emphasised the need for regulations on and clear communication from conversational agents was that of Babylon Health's triage chatbot, which doctors in the United Kingdom were concerned could miss red flags for serious conditions (Iacobucci, 2020). Regulations are especially important if conversational agents are to be used as a stand-alone service and for them to be integrated in healthcare. The long-term effects of conversational agent use are yet to be determined and their impact on society should be carefully considered, for example, when it comes effects on under-represented groups (Laranjo et al., 2018).

To tailor conversations to a user and to have coherence across sessions, information needs to be stored about the user (Bickmore, Trinh, Asadi et al., 2018). Such information can come from multiple sources, such as sensors or user replies. Known considerations for systems that use such data are those with respect to data collection, accuracy of data, use of data and security of data storage. Informing users about data that is collected, stored and used, asking for consent, and updating consent are important elements in eHealth applications in general. But, with conversational agents there are also interaction considerations. For example, what information should an agent remember and how much should it recall in later interactions? Richards and Bransky (2014) found that an agent that remembered previously shared information drew the user into the interaction and that unnatural showing of memory had a negative effect on believability, enjoyment and trust. They also found that characters who partially or completely forgot information were found to be more natural than an agent that incorrectly recalled.

When it comes to tailored coaching conversations, the design of communication about capabilities and processes could benefit from principles in other research domains. Two of such domains are Shared Decision Making (SDM) and Explainable Artificial Intelligence (XAI). Shared decision making is an approach from the medical domain in which patients are encouraged to consider their options, communicate their preferences and make a decision, all in partnership with their clinician (Godolphin, 2009; Stiggelbout et al., 2012). A conversational agent could use strategies for shared decision making in their coaching approach. This principle is gaining attention from the agent community, for example, in the study by Zhang and Bickmore (2018). Another source for coaching interaction with conversational agents might be lessons from research into facilitators and barriers for shared decision making between patients and clinicians (e.g., as discussed in the review by Joseph-Williams et al. (2014)).

Furthermore, even though this thesis suggests to stay within the domain of interpretable models by using scripted or rule-based dialogues, insights from explainable artificial intelligence (Barredo Arrieta et al., 2020) can still serve as a source for designing dialogues. Such use of ‘white box models’ is also supported by, for example, Rudin (2019) who argues that for high stakes decisions such as those in the health domain, interpretable models should be used instead of black box models with post-hoc explanations. Little work has been done on explainable conversational agents in a health context, but one example is the work of Abdulrahman et al. (2021) who investigated different types of explanations for a virtual advisor with the aim of encouraging behaviour change intentions for health behaviours correlated with stress in university students. They investigated explanations based on users’ cognition, and found no difference between belief-based and goal-based explanations, but did find that a longer explanation that included both could hinder motivation. Furthermore, Sheh (2017) lists five types of explanations and relate these to the complexity of potential underlying machine learning algorithms. These five are: teaching, introspective tracing, introspective informative, post-hoc and execution. While they indicate that teaching explanations are only possible with a limited number of machine learning and AI techniques, execution explanations can almost always be provided. With the approach suggested in this thesis, all of these explanation strategies should be possible.

Social interaction

As originally argued by Nass et al. (1994), the human-computer relationship is fundamentally social. Users tend to mindlessly apply social rules to computers if enough social cues are present (Nass & Moon, 2000) (the computers as social actors paradigm or CASA). Nass and Moon (2000) explicitly rejected the assumption of anthropomorphism as a cause (consciously attributing human characteristics), since users consciously know that computers have no feelings, but still treat the computer as a human. However, e-coaching systems in general should have social ability (Kamphorst, 2017) and conversational agents are developed specifically to have anthropomorphic features, which inevitably leads to an increased presence of social cues. Recently, Gambino et al. (2020) proposed that the CASA paradigm should be extended. They argued that technology and interactions with technology have advanced and that there is evidence that users can develop specific scripts for interaction with, for example, conversational agents and that these scripts might differ from those for human-human interaction (Gambino et al., 2020).

In any case, as argued in the literature on relational agents and throughout this thesis, careful attention should be paid to the design of relational and meta content when it comes to conversational agents for health coaching. Agents

should not just adapt the content and strategy of their communication to users, but the social cues that they exhibit – both verbally and nonverbally – should be suitable for building rapport (Gratch et al., 2007), a working alliance (Abdulrahman & Richards, 2020; Bickmore et al., 2005), and a relationship in general. There are many types of social cues (Feine et al., 2019) and design features (Loveys et al., 2020) that can be included and adjusted when it comes to relational behaviour in conversational agents. Examples include social intelligence, empathy, social-oriented language, voice characteristics, communication styles, appearance, nonverbal communication cues, and performance quality (Creed et al., 2015; Loveys et al., 2020; Milne-Ives et al., 2020; Rheu et al., 2020). Adding these elements to an agent can improve the relation between agents and users, for example, when it comes to trust (Rheu et al., 2020). However, they need to be suitable for the situation or they might negatively impact the relationship (Nass & Moon, 2000). For example, Kirakowski et al. (2007) discuss various responses that could make a system seem less human-like, such as responding too quickly to a difficult question.

While many social cues can be leveraged to increase the use of conversational agents, the effects of such design choices on the user need to be carefully considered. Similar to the issues discussed with a focus on coaching advice in the autonomy and trust subsection, there are ethical challenges when it comes to the design of agents to be social and to adopt human-like roles. For example, there is a risk of emotional dependence, isolation, loss of freedom, and amplification of stereotypes (Devillers, 2020). Mensio et al. (2018) discuss similar social and psychological threats that might arise from the perception and use of emotion-aware conversational agents such as addiction, isolation, a change in users' personality (similar to the 'filter bubble' issue on social media) and societal consequences (e.g., attention theft and thus less time to meet new people). When it comes to stereotypes specifically, results show that higher physical attractiveness is reported to increase trust (Rheu et al., 2020), respondents prefer young female agents (ter Stal et al., 2019), most embodied conversational agents are white females (Loveys et al., 2020), and most commercial voice assistants are designed to have female personalities and often give a dismissive or submissive response to insults (Lee et al., 2021). All in all, this shows that while to an extent including social cues and design features might be an effective manner to increase the use of an application, the potential effects on human-human interaction and society do need to be taken into account.

Finally, an important area of future investigation when it comes to interaction is that of initiative. On the one hand, this is about initiative within conversations. In this thesis, the tailored conversations that were implemented (**Chapter 3** and **Chapter 5**) always featured the agent suggesting the topic of conversation. However, in the interviews in **Chapter 5** some participants indicated that there were cases in which they already had a specific topic in mind that they wanted to

discuss. Therefore, the balance between the agent suggesting a topic to discuss or the user doing so should be investigated. This balance could potentially be tailored to factors such as users' stage of change or intrinsic motivation. On the other hand, there is the initiative to initiate a conversation (imagine receiving a message from your coach, which then starts an interaction). The latter is not addressed in this thesis, but could be an area for future research where lessons about timing could be learned from the research area related to *just-in-time interventions* (JITs) (e.g., as discussed by Schembre et al. (2018)) and *just-in-time adaptive interventions* (JITAdIs) (e.g., as discussed by Hardeman et al. (2019) and Nahum-Shani et al. (2018)).

Input and output modalities

The conversational agents in the studies in this thesis used predefined answers options as means of input, that were presented to the user in the form of reply-buttons. As previously discussed in this chapter, major advantages of this approach are that it ensures that an agent is always able to understand the reply and that it has a response for it. Reply-buttons or fixed text input are often used in conversational agents in the health domain. Kramer et al. (2020) report that they were the most common means of input for the embodied conversational agents (for health) in their review, and other reviews on conversational agents also report this (Laranjo et al., 2018; Tudor Car et al., 2020). Bickmore, Trinh, Asadi et al. (2018) argued that unconstrained natural language input (speech, free text) should never be used with a system that provides health advice, unless the system can provide the user with a complete insight into its interpretation and processing of that input. They also elaborate on a range of potential risks. These risks were also shown in studies with well-known commercial conversational assistants (Bickmore, Trinh, Olafsson et al., 2018; Miner et al., 2016). Milne-Ives et al. (2020) report that the most frequently raised issue in studies with conversational agents in their review was poor understanding of speech input. They also reported that agents often had to ask questions more than once to be able to process the response. In addition to interpretation challenges, free text input also raises concerns with respect to privacy and safety (Griffin et al., 2020; Tudor Car et al., 2020). In contrast, it is argued that speech input might make applications more accessible and intuitive (Dingler et al., 2021), for example, for those who have difficulty reading or typing (Tudor Car et al., 2020).

Where free speech can be a difficult modality for input, output modalities are not necessarily limited to speech-bubbles that should be read, but they should be carefully selected to suit the type of agent, user, device and environment of use. For example, speech output could be more natural for an embodied conversational agent that is human-like, while text output might be a more natural choice if the agent does not have an embodiment and communicates

through a text messaging interface (e.g., as is the case with chatbots). Speech output can also make agents more accessible for groups such as older adults (El Kamali et al., 2020), or those have difficulty understanding written text (low literacy) (Bickmore et al., 2010; Bickmore, Trinh, Asadi et al., 2018). On the other hand, the addition of nonverbal behaviour might assist those who have impaired hearing (Montenegro et al., 2019). In any case, non-speech output modalities should always be offered and the agent should ask for permission before discussing sensitive topics (Bickmore, Trinh, Asadi et al., 2018). Furthermore, integration in existing and often used text-messaging platforms might make conversational agents easier to access (Tudor Car et al., 2020), although this does require considerations about (health) data being stored in those platforms.

All in all, interaction with conversational agents might benefit from different interfaces and interaction modalities for different situations and environments. For example, a user could interact with a state-of-the-art 3D-coach when at home, but will have their coach available through a text-messaging interface when on the move. An important notion in this is that conversational strategies will have to be adjusted to the situation and device used. For example, people are often easier distracted and more interrupted when using a mobile phone, which might require reminders of the current topic and shorter dialogues (Bickmore, Trinh, Asadi et al., 2018).

Open science

As a final consideration, a note about open science. Open science aims to make science accessible to all members of society, and features concepts such as findable accessible, interoperable and re-usable (FAIR) data, open access publishing, and open source software. When it comes to conversational agents for health, there is a positive trend in publishing papers open access.

However, sharing source code and dialogue content is less common. Publishing agents' source code could contribute to the safety, verifiability and validation of applications. Making source code publicly available is an approach that is being applied more often, for example, the source code for the Dutch Corona notification app is also publicly accessible. In addition, the use of a common open source research platform could make studies easier to compare, since underlying software would be the same (Scholten et al., 2017). The open source WOOL Dialogue Platform (presented in **Chapter 2**) and the Agents United Platform (Beinema et al., 2021) could fulfil such a role.

In addition to accessibility of source code and open access publishing of papers, a clear description of dialogue content should be available for published studies. One could even argue that dialogue content and source code should be added as supplementary materials to papers. This would serve multiple purposes, such

as helping to clarify to those reading the paper what it is that participants experienced, enabling more precise integration of results in new applications, and potential reuse of dialogue content in new systems. Most importantly, it would support the replication of studies.

Conclusion

This thesis investigated how health coaching conversations with conversational agents could be tailored to improve engagement. To that end, it introduced new tools for dialogue authoring and execution, investigated the tailoring of strategies underlying to a conversation, and developed and evaluated a model and implementation for tailoring conversations on a high level by selecting relevant topics during interaction.

As discussed in this chapter, there are still plenty of directions and considerations for future research when it comes to tailoring conversations with conversational agents, and the application of these agents as a stand-alone technology in healthcare and our daily life. With the ongoing developments in sensors, smart devices, intelligent interaction and digital health, conversational agents are gradually evolving to become intelligent applications that can provide us with the support we need, when we need it. In the end, this brings us closer to a future in which our conversational coaches can independently support us in achieving our healthy lifestyles.

&

Appendix A: Coaching strategy dialogues

This Appendix contains the text from the dialogues that were used in **Chapter 3** to present the coaching strategies to the users. The presenting coach and order of presentation were randomized for strategy one and two. The names of the coaches were inserted where it says '\$StrategyOneCoach' and '\$StrategyTwoCoach'. The strategy presentation started with a few introduction sentences, which were the following:

Coda: Coaches, now that you have finished your discussion: What's the plan?

Helen: Well, we have discussed and there are two coaching approaches that we think could be suitable.

Helen: We'd like to tell you about both. \$StrategyOneCoach and \$StrategyTwoCoach will both explain one.

User: Ok.

Helen: Helen: Good! Let's start with the first one... \$StrategyOneCoach?

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

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

Which was followed by the presentation of the other strategy for the participants' motivation group by the other coach.

And then there were some concluding remarks, namely:

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

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

A.1 Self Goal Setting

Coach: I suggest that we will support you in setting your own health goals.

Coach: For example, you can tell us how many steps you would like to take every day, or how many glasses of water you would like to drink.

(*Option 1*)

User: How many steps?

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

(*Option 2*)

User: How much water?

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

Coach: The most important thing is that you set a concrete goal for yourself. Then, together we will try to make sure that you stick to your goals!

Coach: We can help you choose a goal. After all, we are familiar with all the healthy activity lifestyle guidelines.

Coach: And we know a lot about healthy eating as well. But in the end, you have to set your own goals.

(*Optional, can be skipped.*)

User: But, why can't you set a goal for me?

Coach: If you set your own goals, chances are higher that you actually stick with them, compared to when we will try to commit you to something.

User: Okay, I think I understand?

Coach: Great! I hope this coaching approach will work well for you!



A.2 Health Education

Coach: I suggest we will provide you with tips and advice on healthy behaviour, and explain how this contributes to your health.

Coach: For example, why is it important to be physically active? And why is a healthy eating pattern good for you... .

(Option 1)

User: Why is physical activity important?

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

(Option 2)

User: Why is healthy eating important?

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

Coach: We will start by having a conversation about your general health knowledge – just to see what you already know about healthy eating and activity.

Coach: We will also talk about your reasons for wanting to live a healthier life. That way, we can better tailor our information to your wishes and needs.

Coach: And of course, there is always the possibility of asking questions!

(Optional, can be skipped.)

User: But, why is all of this important?

Coach: It may not always be easy to adopt a healthy lifestyle. If you know exactly what the potential benefits are, this will help you in achieving your goals.

User: Okay, I think I understand?

Coach: Great! I hope this coaching approach will work well for you!

A.3 Showing Progress

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

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

(Option 1)

User: How do you know how much I'm walking?

Coach: We have automatic step counters for that. If you agree to share this data, it should be no problem keeping track of your steps.

(Option 2)

User: How do you know what I'm eating?

Coach: We could keep track of that in a food diary, for example, but we'll figure that out when we get there.

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

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

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

(Optional, can be skipped.)

User: But, why is knowing this so important?

Coach: Insight! A healthy lifestyle starts with understanding your current behaviour. If we show you how you're doing, you can see how close you are to reaching your health goals.

User: Okay, I think I understand?

Coach: Great! I hope this coaching approach will work well for you!



&

A.4 Implementation Intentions

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

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

(Option 1)

User: Why three days?

Coach: It could be fewer, or less. That is up to you.

(Option 2)

User: Should it be the same every week?

Coach: No, you can select different days every week.

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

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

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

(Optional, can be skipped.)

User: But, why do I have to schedule all that?

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

User: Okay, I think I understand?

Coach: Great! I hope this coaching approach will work well for you!

Appendix B: Overview of actions extracted from literature (excluding BCTs)

Table B.1: An overview of the coaching actions that were defined based on the literature on relational agents and persuasive technology.

Coaching action	Description
Getting acquainted talk	Share initial information about yourself with the person to get acquainted with each other.
Reciprocal self-disclosure	Share information about your personal background with the user and ask them questions about theirs.
Expressing happiness to see the user	Express happiness to see the person again.
Social dialogue	Discuss a social topic of conversation with the person, such as e.g. the weather.
Explain interaction paradigm	Explain to the person how the system works and how to communicate.
Questions on feelings	Ask questions regarding the person's feelings with the aim of adjusting the coaching process to them.
Questions on context	Ask questions regarding the person's context with the aim of adjusting the coaching process to their specific situation.
Explain standards	Explain to the person what the advised behaviour is for their demographics.
Discuss 'healthy behaviour'	Discuss what healthy behaviour is with the person.
Compliment (behaviour)	Compliment the person on their behaviour.
Compliment (outcome)	Compliment the person on the outcomes of their behaviour.
Show collected data	Provide an overview of measured data.
Send a reminder	Notify the person of a previously planned activity.
Connect sensor	Discuss the connection of sensors and provide support in this process.
Explain use for sensors	Explain the use for sensors.
Explain use for self-monitoring	Explain the use for self-monitoring.
Provide assistance with sensor problems	Provide the assistance with sensor problems.
Explain widget	Explain where to find a widget and how to use it.



&

Table B.1: An overview of the coaching actions that were defined based on the literature on relational agents and persuasive technology (continued).

Coaching action	Description
Confirm measured values	Ask the person if the measured values are correct.
Explain coaching options	Explain the possible coaching that can be provided.
Ask for coaching preferences	Ask the person for their preferences when it comes to coaching.
Interpret collected data	Provide an interpretation of collected data in relation to the goals that were set.

Appendix C: An overview of the descriptions added to the topics for the card sort

Table C.1: An overview of the descriptions added to all defined topics for the card sort.

Topic	Description
Social	Non-coaching conversations aimed at building a relationship between coach and the person.
Introduction	Have a conversation with the person in which you introduce yourselves.
Share a background story	Provide some background information about yourself (e.g. about a hobby or recent activity).
Small talk	Have a conversation about a small talk topic, e.g. discuss the weather.
Welcome back	Welcome the person back and e.g. ask them how they have been.
Meta	Conversations about the system and the coaching process.
Explain how to use the system	Provide information on how to use (parts of) the system.
Assist with sensor connection	Explain how to connect sensors and provide (technical) assistance with the connection.
Discuss coaching approach	Explain options for coaching and discuss the person's preferred approach.
Coaching	Conversations related to behavior change for the benefit of the person using the system.
Goals & Planning	Discuss goals, and discuss plans and influential factors for reaching these goals.
Set a new goal	Discuss what the person's new goal should be.
Review existing goal	Discuss an existing goal and whether it should be adjusted.
Discuss actions to achieve goal	Discuss plans and how to commit to them.
Discuss facilitators and barriers	Ask about potential problems and help the person overcome them.
Monitoring	Discuss monitoring of the person's behavior.
Discuss sensors	Discuss monitoring of behavior through sensors.
Discuss self-monitoring	Discuss the person monitoring themselves.
Feedback	Provide an interpretation of the person's recorded behavior.



&

Table C.1: An overview of the descriptions added to all defined topics for the card sort (continued).

Reward & punishment	Discuss rewards or punishments as a consequence of the person's behavior.
Reward	Discuss reward as a consequence of the person's behavior.
Punishment	Discuss punishment as a consequence of the person's behavior.
Health education	Educate the person on healthy behavior and provide advice on how this might be achieved.
Inform 'what'	Discuss what healthy behavior is.
Inform 'why'	Discuss why certain behavior is (not) beneficial for the person.
Inform 'how'	Discuss things the person could do to improve or sustain their healthy behavior.
Do an exercise	Provide the person with an exercise they can do to help them in their behavior change.
Do mental exercise	Ask the person to perform an exercise that can help their mindset, such as thought experiments or imagining.
Do practical exercise	Ask the person to practice behavior or adapt their context to their (desired) new behavior.
Gather information	Ask the person questions so that the provided coaching can be personalized.
Reminders	Remind the person of something.

References

Chapter 1

- Abdullah, A. S., Gaehde, S. & Bickmore, T. W. (2018). A tablet based embodied conversational agent to promote smoking cessation among veterans: A feasibility study. *Journal of Epidemiology and Global Health*, 8(3-4), 225–230. <https://doi.org/10.2991/j.jegh.2018.08.104>
- Achterkamp, R., Cabrita, M., op den Akker, H., Hermens, H. J. & Vollenbroek-Hutten, M. M. R. (2013). Promoting a healthy lifestyle: Towards an improved personalized feedback approach. *HealthCom '13: Proceedings of the 15th IEEE International Conference on e-Health Networking, Applications and Services*, 725–727. <https://doi.org/10.1109/HealthCom.2013.6720772>
- Andersson, G., Carlbring, P., Berger, T., Almlöv, J. & Cuijpers, P. (2009). What makes internet therapy work? *Cognitive Behaviour Therapy*, 38(S1), 55–60. <https://doi.org/10.1080/16506070902916400>
- André, E. & Pelachaud, C. (2010). Interacting with embodied conversational agents. *Speech technology* (pp. 123–149). https://doi.org/10.1007/978-0-387-73819-2_8
- André, E. & Rist, T. (2001). Presenting through performing: On the use of multiple lifelike characters in knowledge-based presentation systems. *Knowledge-Based Systems*, 14(1-2), 3–13. [https://doi.org/10.1016/S0950-7051\(00\)00096-4](https://doi.org/10.1016/S0950-7051(00)00096-4)
- Benítez-Guijarro, A., Ruiz-Zafra, A., Callejas, Z., Medina-Medina, N., Benghazi, K. & Noguera, M. (2019). General architecture for development of virtual coaches for healthy habits monitoring and encouragement. *Sensors*, 19(1), 108. <https://doi.org/10.3390/s19010108>
- Bickmore, T. W. & Giorgino, T. (2006). Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, 39(5), 556–571. <https://doi.org/10.1016/j.jbi.2005.12.004>
- Bickmore, T. W., Gruber, A. & Picard, R. W. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 59(1), 21–30. <https://doi.org/10.1016/j.pec.2004.09.008>
- Bickmore, T. W., Kimani, E., Trinh, H., Pusateri, A., Paasche-Orlow, M. K. & Magnani, J. W. (2018). Managing chronic conditions with a smartphone-based conversational virtual agent. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 119–124. <https://doi.org/10.1145/3267851.3267908>
- Bickmore, T. W., Schulman, D. & Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6), 648–666. <https://doi.org/10.1080/08839514.2010.492259>
- Bickmore, T. W., Silliman, R. A., Nelson, K., Cheng, D. M., Winter, M., Henault, L. & Paasche-Orlow, M. K. (2013). A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society*, 61(10), 1676–1683. <https://doi.org/10.1111/jgs.12449>
- Bickmore, T. W., Trinh, H., Asadi, R. & Olafsson, S. (2018). Safety first: Conversational agents for health care. *Studies in Conversational UX Design*, 33–57. https://doi.org/10.1007/978-3-319-95579-7_3
- Bickmore, T. W., Utami, D., Matsuyama, R. & Paasche-Orlow, M. K. (2016). Improving access to online health information with conversational agents: A randomized controlled experiment. *Journal of Medical Internet Research*, 18(1), e5239. <https://doi.org/10.2196/jmir.5239>
- Bouton, M. E. (2014). Why behavior change is difficult to sustain. *Preventive Medicine*, 68, 29–36. <https://doi.org/10.1016/j.ypmed.2014.06.010>

- Brinkman, W.-P. (2016). Virtual health agents for behavior change: Research perspectives and directions. *GREATS '16: Proceedings of the Workshop on Graphical and Robotic Embodied Agents for Therapeutic Systems, held during the 16th International Conference on Intelligent Virtual Agents (IVA '16)*, 1–17.
- Buimer, H. P., Tabak, M., van Velsen, L., van der Geest, T. & Hermens, H. J. (2017). Exploring determinants of patient adherence to a portal-supported oncology rehabilitation program: Interview and data log analyses. *JMIR Rehabilitation and Assistive Technologies*, 4(2), e12. <https://doi.org/10.2196/rehab.6294>
- Callejas, Z. & Griol, D. (2021). Conversational agents for mental health and wellbeing. *Dialog systems: A perspective from language, logic and computation* (pp. 219–244). https://doi.org/10.1007/978-3-030-61438-6_11
- Castonguay, L. G., Constantino, M. J. & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy: Theory, Research, Practice, Training*, 43(3), 271–279. <https://doi.org/10.1037/0033-3204.43.3.271>
- Cole-Lewis, H., Ezeanochie, N. & Turgiss, J. (2019). Understanding health behavior technology engagement: Pathway to measuring digital behavior change interventions. *JMIR Formative Research*, 3(4), 1–10. <https://doi.org/10.2196/14052>
- Couper, M. P., Alexander, G. L., Zhang, N., Little, R. J. A., Maddy, N., Nowak, M. A., McClure, J. B., Calvi, J. J., Rolnick, S. J., Stopponi, M. A. & Johnson, C. C. (2010). Engagement and retention: Measuring breadth and depth of participant use of an online intervention. *Journal of Medical Internet Research*, 12(4), e52. <https://doi.org/10.2196/jmir.1430>
- de Vries, R. A. J., Truong, K. P., Kwint, S., Drossaert, C. H. C. & Evers, V. (2016). Crowd-designed motivation: Motivational messages for exercise adherence based on behavior change theory. *CHI '16: Proceedings of the 2016 ACM Conference on Human Factors in Computing Systems*, 297–308. <https://doi.org/10.1145/2858036.2858229>
- de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J. & Evers, V. (2017). A word of advice: How to tailor motivational text messages based on behavior change theory to personality and gender. *Personal and Ubiquitous Computing*, 21(4), 675–687. <https://doi.org/10.1007/s00779-017-1025-1>
- Fadhil, A., Wang, Y. & Reiterer, H. (2019). Assistive conversational agent for health coaching: A validation study. *Methods of Information in Medicine*, 58(1), 9–23. <https://doi.org/10.1055/s-0039-1688757>
- Fitzpatrick, K. K., Darcy, A. & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, 4(2), e19. <https://doi.org/10.2196/mental.7785>
- Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., Jack, B. W. & Bickmore, T. W. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient Education and Counseling*, 100(9), 1720–1729. <https://doi.org/10.1016/j.pec.2017.04.015>
- Glas, N. & Pelachaud, C. (2018). Topic management for an engaging conversational agent. *International Journal of Human-Computer Studies*, 120, 107–124. <https://doi.org/10.1016/j.ijhcs.2018.07.007>
- Grey, J. & Bryson, J. (2011). Procedural quests: A focus for agent interaction in role-playing-games. *AISB '11: Artificial Intelligence and the Simulation of Behaviour Symposium on AI and Games*, 3–10.
- Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J. & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54, 170–179. <https://doi.org/10.1016/j.chb.2015.07.045>

REFERENCES

- Hatcher, R. L. & Gillaspy, J. A. (2006). Development and validation of a revised short version of the working alliance inventory. *Psychotherapy Research*, 16(1), 12–25. <https://doi.org/10.1080/10503300500352500>
- Horvath, A. O. & Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *Journal of Counseling Psychology*, 36(2), 223–233. <https://doi.org/10.1037/0022-0167.36.2.223>
- Huber, M., Knottnerus, J. A., Green, L., van der Horst, H., Jadad, A. R., Kromhout, D., Leonard, B., Lorig, K., Loureiro, M. I., van der Meer, J. W. M., Schnabel, P., Smith, R., van Weel, C. & Smid, H. (2011). How should we define health? *BMJ*, 343(7817), 1–3. <https://doi.org/10.1136/bmj.d4163>
- Huber, M., van Vliet, M., Giezenberg, M., Winkens, B., Heerkens, Y., Dagnelie, P. C. & Knottnerus, J. A. (2016). Towards a 'patient-centred' operationalisation of the new dynamic concept of health: A mixed methods study. *BMJ Open*, 6(1), 1–11. <https://doi.org/10.1136/bmjopen-2015-010091>
- Kamphorst, B. A. (2017). E-coaching systems: What they are, and what they aren't. *Personal and Ubiquitous Computing*, 21(4), 625–632. <https://doi.org/10.1007/s00779-017-1020-6>
- Kantharaju, R. B., De Franco, D., Pease, A. & Pelachaud, C. (2018). Is two better than one? effects of multiple agents on user persuasion. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 255–262. <http://arxiv.org/abs/1904.05248>
- Kantharaju, R. B., Pease, A., Reidsma, D., Pelachaud, C., Snaith, M., Bruijnes, M., Klaassen, R., Beinema, T., Huizing, G., Simonetti, D., Heylen, D. & op den Akker, H. (2019). Integrating argumentation with social conversation between multiple virtual coaches. *IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 203–205.
- Kaptein, M., de Ruyter, B., Markopoulos, P. & Aarts, E. (2012). Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(2), 1–25. <https://doi.org/10.1145/2209310.2209313>
- Kim, Y. & Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education*, 26(1), 160–169. <https://doi.org/10.1007/s40593-015-0055-y>
- King, A. C., Campero, I., Sheats, J. L., Castro Sweet, C. M., Garcia, D., Chazaro, A., Blanco, G., Hauser, M., Fierros, F., Ahn, D. K., Diaz, J., Done, M., Fernandez, J. & Bickmore, T. W. (2017). Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: The COMPASS trial design, methods, and baseline characteristics. *Contemporary Clinical Trials*, 61, 115–125. <https://doi.org/10.1016/j.cct.2017.07.020>
- Kivelä, K., Elo, S., Kyngäs, H. & Kääriäinen, M. (2014). The effects of health coaching on adult patients with chronic diseases: A systematic review. *Patient Education and Counseling*, 97(2), 147–157. <https://doi.org/10.1016/j.pec.2014.07.026>
- Klaassen, R., Bul, K. C. M., op den Akker, R., van der Burg, G. J., Kato, P. M. & Di Bitonto, P. (2018). Design and evaluation of a pervasive coaching and gamification platform for young diabetes patients. *Sensors*, 18(2), 1–27. <https://doi.org/10.3390/s18020402>
- Klein, M. C. A., Mogles, N. & van Wissen, A. (2013). An intelligent coaching system for therapy adherence. *IEEE Pervasive Computing*, 12(3), 22–30. <https://doi.org/10.1109/MPRV.2013.41>
- Kramer, L. L., ter Stal, S., Mulder, B. C., de Vet, E. & van Velsen, L. (2020). Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *Journal of medical Internet Research*, 22(2), e14058. <https://doi.org/10.2196/14058>

- Krämer, N. C., Hoffmann, L. & Kopp, S. (2010). Know your users! empirical results for tailoring an agent's nonverbal behavior to different user groups. *IVA '10: Proceedings of the 10th ACM International Conference on Intelligent Virtual Agents*, 6356 LNCS, 468–474. https://doi.org/10.1007/978-3-642-15892-6_50
- Krebs, P., Prochaska, J. O. & Rossi, J. S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. *Preventive Medicine*, 51(3-4), 214–221. <https://doi.org/10.1016/j.ypmed.2010.06.004>
- Kuligowska, K. (2015). Commercial chatbot: Performance evaluation, usability metrics and quality standards of embodied conversational agents. *Professional Center for Business Research*, 2(2), 1–16.
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J. Y. C., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S. & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, 25(9), 1248–1258. <https://doi.org/10.1093/jamia/ocy072>
- Lisetti, C. L., Yasavur, U., Visser, U. & Rishe, N. (2011). Toward conducting motivational interviewing with an on-demand clinician avatar for tailored health behavior change interventions. *PervasiveHealth '11: Proceedings of the 5th International Conference on Pervasive Computing Technologies for Healthcare*, 246–249. <https://doi.org/10.4108/icst.pervasivehealth.2011.246078>
- Ma, T., Sharifi, H. & Chattopadhyay, D. (2019). Virtual humans in health-related interventions: A meta-analysis. *CHI '19: Proceedings of the 2019 ACM Conference on Human Factors in Computing Systems*, 1–6.
- Nakamura, J. & Csikszentmihalyi, M. (2002). The concept of flow. *Handbook of positive psychology* (pp. 89–105). <https://doi.org/10.1002/9780470172698.ch19>
- Nass, C., Steuer, J. & Tauber, E. R. (1994). Computers are social actors. *CHI '94: Proceedings of the 1994 ACM Conference on Human Factors in Computing Systems*, 72–78. <https://doi.org/10.1145/259963.260288>
- Nijland, N. (2011). *Grounding eHealth - towards a holistic framework for sustainable eHealth technologies* (Doctoral dissertation). University of Twente. <https://doi.org/10.3990/1.9789036531337>
- O'Brien, H. L. & Toms, E. G. (2013). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938–955. <https://doi.org/10.1002/asi>
- Office of the United Nations High Commissioner for Human Rights & World Health Organization. (2008). *The right to health* (tech. rep.). <https://www.ohchr.org/documents/publications/factsheet31.pdf>
- Olafsson, S., O'Leary, T. & Bickmore, T. W. (2019). Coerced change-talk with conversational agents promotes confidence in behavior change. *PervasiveHealth '19: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 31–40. <https://doi.org/10.1145/3329189.3329202>
- op den Akker, H., Jones, V. M. & Hermens, H. J. (2014). Tailoring real-time physical activity coaching systems: A literature survey and model. *User Modeling and User-Adapted Interaction*, 24(5), 351–392. <https://doi.org/10.1007/s11257-014-9146-y>
- op den Akker, H., op den Akker, R., Beinema, T., Banos, O., Heylen, D., Pease, A., Pelachaud, C., Traver Salcedo, V., Kyriazakos, S. & Hermens, H. J. (2018). Council of Coaches: A novel holistic behavior change coaching approach. *ICT4AWE '18: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 978–989. <https://doi.org/10.5220/0006787702190226>
- Perski, O., Blandford, A., West, R. & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, 7(2), 254–267. <https://doi.org/10.1007/s13142-016-0453-1>

REFERENCES

- Pickard, M. D. (2012). *Persuasive embodied agents: Using embodied agents to change people's behavior, beliefs, and assessments* (Doctoral dissertation). University of Arizona.
- Ryan, K., Dockray, S. & Linehan, C. (2019). A systematic review of tailored eHealth interventions for weight loss. *Digital Health*, 5, 1–29. <https://doi.org/10.1177/2055207619826685>
- Scholten, M. R., Kelders, S. M. & van Gemert-Pijnen, J. E. W. C. (2017). Self-guided web-based interventions: Scoping review on user needs and the potential of embodied conversational agents to address them. *Journal of Medical Internet Research*, 19(11), e383. <https://doi.org/10.2196/jmir.7351>
- Schulman, D. & Bickmore, T. W. (2009). Persuading users through counseling dialogue with a conversational agent. *PERSUASIVE '09: Proceedings of the 4th International Conference on Persuasive Technology*, 1–8. <https://doi.org/10.1145/1541948.1541983>
- Sebastian, J. & Richards, D. (2017). Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior*, 73, 479–488. <https://doi.org/10.1016/j.chb.2017.03.071>
- Smith, C., Crook, N., Dobnik, S., Charlton, D., Boye, J., Pulman, S., Santos de la Camara, R., Turunen, M., Benyon, D., Bradley, J., Gambäck, B., Hansen, P., Mival, O., Webb, N. & Cavazza, M. (2011). Interaction strategies for an affective conversational agent. *Presence*, 20(5), 395–411. https://doi.org/10.1162/PRES_a_00063
- Starr, J. (2008). *The coaching manual: The definitive guide to the process, principles, and skills of personal coaching*. Pearson Education.
- Tanaka, H., Negoro, H., Iwasaka, H. & Nakamura, S. (2017). Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLoS ONE*, 12(8), 1–15. <https://doi.org/10.1371/journal.pone.0182151>
- ter Stal, S., Broekhuis, M., van Velsen, L., Hermens, H. J. & Tabak, M. (2020). Embodied conversational agent appearance for health assessment of older adults: Explorative study. *JMIR Human Factors*, 7(3), e19987. <https://doi.org/10.2196/19987>
- Trinh, H., Shamekhi, A., Kimani, E. & Bickmore, T. W. (2018). Predicting user engagement in longitudinal interventions with virtual agents. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 9–16. <https://doi.org/10.1145/3267851.3267909>
- United Nations Department of Public Information. (2017). *2017 revision of world population prospects* (tech. rep.). World Health Organisation. http://www.un.org/en/webguidelines/pdf/DPI_SocialMedia_Guide.pdf
- Uribe, J. A., Duitama, J. F. & Gaviria Gómez, N. (2011). Personalized message emission in a mobile application for supporting therapeutic adherence. *HealthCom '11: Proceedings of the 13th IEEE International Conference on e-Health Networking, Applications and Services*, 15–20. <https://doi.org/10.1109/HEALTH.2011.6026734>
- van Velsen, L., Cabrita, M., op den Akker, H., Brandl, L., Isaac, J., Suárez, M., Gouveia, A., de Sousa, R. D., Rodrigues, A. M., Canhão, H., Evans, N., Blok, M., Alcobia, C. & Brodbeck, J. (2020). LEAVES (optimizing the mental health and resilience of older Adults that have lost their spouse via blended, online therapy): Proposal for an online service development and evaluation. *JMIR Research Protocols*, 9(9), e19344. <https://doi.org/10.2196/19344>
- van Wissen, A., Vinkers, C. & van Halteren, A. (2016). Developing a virtual coach for chronic patients: A user study on the impact of similarity, familiarity and realism. In A. Meschtscherjakow, B. De Ruyter, V. Fuchsberger, M. Murer & M. Tscheligi (Eds.), *Persuasive '16: Proceedings of the 11th international conference on persuasive technology* (pp. 263–275). Springer, Cham. https://doi.org/10.1007/B978-0-12-373932-2_00008-9
- Veletsianos, G. & Russell, G. S. (2014). Pedagogical agents. In J. M. Spector, M. D. Merrill, J. Elen & M. J. Bishop (Eds.), *Handbook of research on educational communications*

- and technology: Fourth edition* (pp. 1–12). <https://doi.org/10.1007/978-1-4614-3185-5>
- Wangberg, S. C., Bergmo, S., Trine & Johnson, J.-A. K. (2008). Adherence in internet-based interventions. *Patient Preference and Adherence*, 2, 57–66.
- Watson, A., Bickmore, T. W., Cange, A., Kulshreshtha, A. & Kvedar, J. (2012). An internet-based virtual coach to promote physical activity adherence in overweight adults: Randomized controlled trial. *Journal of Medical Internet Research*, 14(1), e1. <https://doi.org/10.2196/jmir.1629>
- World Health Organisation. (2021). Noncommunicable diseases fact sheet [Accessed: 17-08-2021].
- World Health Organization. (1946). *Constitution of the world health organization* (tech. rep.).
- Xu, A., Liu, Z., Guo, Y., Sinha, V. & Akkiraju, R. (2017). A new chatbot for customer service on social media. *CHI '17: Proceedings of the 2017 ACM Conference on Human Factors in Computing Systems*, 3506–3510. <https://doi.org/10.1145/3025453.3025496>
- Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., Merchant, G. C., Naughton, F. & Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine*, 51(5), 833–842. <https://doi.org/10.1016/j.amepre.2016.06.015>
- Zhou, S., Bickmore, T. W., Paasche-Orlow, M. K. & Jack, B. W. (2014). Agent-user concordance and satisfaction with a virtual hospital discharge nurse. *IVA '14: Proceedings of the 14th ACM International Conference on Intelligent Virtual Agents*, 8637 LNCS, 528–541. https://doi.org/10.1007/978-3-319-09767-1_63

Chapter 2

- Aleven, V., Sewall, J., McLaren, B. M. & Koedinger, K. R. (2006). Rapid authoring of intelligent tutors for real-world and experimental use. *ICALT '06: Proceedings of the 6th International Conference on Advanced Learning Technologies*, 847–851. <https://doi.org/10.1109/icalt.2006.1652575>
- Amazon. (2020). Amazon Lex [Accessed: 09-03-2020]. <https://aws.amazon.com/lex/>
- Articy Software GmbH & Co. KG. (2020). Articy [Accessed: 09-03-2020]. <https://www.articy.com/>
- Beinema, T., Davison, D., Reidsma, D., Banos, O., Bruijnes, M., Donval, B., Fides Valero, Á., Heylen, D., Hofst, D., Huizing, G., Kantharaju, R. B., Klaassen, R., Kolkmeier, J., Konsolakis, K., Pease, A., Pelachaud, C., Simonetti, D., Snaith, M., Traver, V., ... op den Akker, H. (2021). Agents United: An open platform for multi-agent conversational systems. *IVA '21: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*.
- Bex, F., Lawrence, J. & Reed, C. A. (2014). Generalising argument dialogue with the dialogue game execution platform. *COMMA '14: Proceedings of the 5th International Conference on Computational Models of Argument*, 141–152. <https://doi.org/10.3233/978-1-61499-436-7-141>
- BotSociety.io. (2020). BotSociety [Accessed: 08-05-2020]. <https://botsociety.io>
- Cai, Z., Graesser, A. & Hu, X. (2015). ASAT: AutoTutor Script Authoring Tool. In R. Sottilare, A. Graesser, X. Hu & K. Brawner (Eds.), *Design recommendations for intelligent tutoring systems, volume 3: Authoring tools & expert modelling techniques* (pp. 199–210).
- Code Whale Inc. (2020). POEditor [Accessed: 16-04-2020]. <https://poeditor.com/>
- Eastgate Systems, Inc. (2020). Storyspace 3 [Accessed: 11-03-2020]. <http://www.eastgate.com/storyspace/index.html>
- Fungus Games. (2020). Fungus [Accessed: 07-05-2020]. <https://fungusgames.com/>

REFERENCES

- Gaffney, C., Dagger, D. & Wade, V. (2008). A survey of soft skill simulation authoring tools. *HYPertext '08: Proceedings of the 19th ACM Conference on Hypertext and Hypermedia*, 181–185. <https://doi.org/10.1145/1379092.1379127>
- Gandhe, S., Whitman, N., Traum, D. R. & Artstein, R. (2005). An integrated authoring tool for tactical questioning dialogue systems. *Proceedings of the 6th ijcai workshop on knowledge and reasoning in practical dialogue systems* (pp. 10–17). papers: //28f4590b-a64b-4c0a-9ebc-51d6773cdd9a/Paper/p94
- Gebhard, P., Kipp, M., Klesen, M. & Rist, T. (2003). Authoring scenes for adaptive, interactive performances. *AAMAS '03: Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, 725–732. <https://doi.org/10.1145/860690.860692>
- Github Inc. (2020). Electron [Accessed: 28-04-2020]. <https://www.electronjs.org/>
- Google. (2020). Dialog Flow [Accessed: 11-03-2020]. <https://dialogflow.com/>
- Graham Nelson. (2019). Inform 7 [Accessed: 08-05-2020]. <http://inform7.com/>
- Green, D., Hargood, C. & Charles, F. (2018a). Contemporary issues in interactive storytelling authoring systems. *ICIDS '18: Proceedings of the 11th International Conference on Interactive Digital Storytelling*, 11318 LNCS, 501–513. https://doi.org/10.1007/978-3-030-04028-4_59
- Green, D., Hargood, C. & Charles, F. (2018b). Define “authoring tool”: A survey of interactive narrative authoring tools. *ICIDS '18: Proceedings of the 11th International Conference on Interactive Digital Storytelling*, 1–6.
- Green, D., Hargood, C. & Charles, F. (2019). Novella 2.0: A hypertextual architecture for interactive narrative in games. *HYPertext '19: Proceedings of the 30th ACM Conference on Hypertext and Social Media*, 77–86. <https://doi.org/10.1145/3342220.3343655>
- Grey, J. & Bryson, J. (2011). Procedural quests: A focus for agent interaction in role-playing-games. *AISB '11: Artificial Intelligence and the Simulation of Behaviour Symposium on AI and Games*, 3–10.
- Guimarães, M., Mascarenhas, S., Prada, R., Santos, P. A. & Dias, J. (2019). An accessible toolkit for the creation of socio-emotional agents. *AAMAS '19: Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems*, 2357–2359.
- IBM. (2020). Watson Assistant [Accessed: 11-03-2020]. <https://www.ibm.com/cloud/watson-assistant/>
- Inkle Studios. (2019). Inklewriter [Accessed: 08-05-2020]. <https://www.inklestudios.com/inklewriter/>
- Interactive Fiction Technology Foundation. (2020). Twine [Accessed: 11-03-2020]. <http://twinery.org/>
- Iurgel, I. A. (2008). Cyranus - an authoring tool for interactive edutainment applications. *Edutainment '06: Proceedings of the International Conference on Technologies for E-Learning and Digital Entertainment*, 3942 LNCS, 577–580.
- Johnson, W. L. & Valente, A. (2008). Collaborative authoring of serious games for language and culture. *Proceedings of the SimTecT 2008 Simulation Conference*.
- Jordan, P. W., Hall, B., Ringenberg, M., Cue, Y. & Rosé, C. (2007). Tools for authoring a dialogue agent that participates in learning studies. *Artificial Intelligence in Education*, 158(1-2), 43–50.
- Kim, Y. & Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education*, 26(1), 160–169. <https://doi.org/10.1007/s40593-015-0055-y>
- Kuligowska, K. (2015). Commercial chatbot: Performance evaluation, usability metrics and quality standards of embodied conversational agents. *Professional Center for Business Research*, 2(2), 1–16.
- Landbot.io. (2020). Landbot I.O. [Accessed: 08-05-2020]. <https://landbot.io/>

- Leuski, A. & Traum, D. R. (2011). NPCEditor: Creating virtual human dialogue using information retrieval techniques. *AI Magazine*, 32(2), 42–56. <https://doi.org/10.1609/aimag.v32i2.2347>
- Martens, C. & Iqbal, O. (2019). Villanelle: An authoring tool for autonomous characters in interactive fiction. *ICIDS '19: Proceedings of the 12th International Conference on Interactive Digital Storytelling*, 190–203. <https://doi.org/10.1007/978-3-030-33894-7>
- Medller, B. & Magerko, B. (2006). Scribe: A tool for authoring event driven interactive drama. *TIDSE '06: Proceedings of the International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, 4326 LNCS, 139–150.
- Mehm, F., Göbel, S. & Steinmetz, R. (2012). Authoring of serious adventure games in StoryTec. *E-Learning and Games for Training, Education, Health and Sports*, 7516 LNCS, 144–154. https://doi.org/10.1007/978-3-642-33466-5_16
- Microsoft. (2020). Azure Bot Service [Accessed: 11-03-2020]. <https://azure.microsoft.com/en-us/services/bot-service/>
- Mitrovic, A., Martin, B., Suraweera, P., Zakharov, K., Milik, N., Holland, J. & McGuigan, N. (2009). ASPIRE: An authoring system and deployment environment for constraint-based tutors. *International Journal of Artificial Intelligence in Education*, 19(2), 155–188.
- Niewiadomski, R., Bevacqua, E., Mancini, M. & Pelachaud, C. (2009). Greta: An interactive expressive ECA system. *AAMAS '09: Proceedings of the 8th International Joint Conference on Autonomous Agents and Multiagent Systems*, 1427–1428.
- Nijland, N. (2011). *Grounding eHealth - towards a holistic framework for sustainable eHealth technologies* (Doctoral dissertation). University of Twente. <https://doi.org/10.3990/1.9789036531337>
- op den Akker, H., op den Akker, R., Beinema, T., Banos, O., Heylen, D., Pease, A., Pelachaud, C., Traver Salcedo, V., Kyriazakos, S. & Hermens, H. J. (2018). Council of Coaches: A novel holistic behavior change coaching approach. *ICT4AWE '18: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 978–989. <https://doi.org/10.5220/0006787702190226>
- Ososky, S. J. & Sottilare, R. A. (2016). *A heuristic evaluation of the Generalized Intelligent Framework for Tutoring (GIFT) authoring tools* (tech. rep.).
- Radmatt. (2020). Yarn Editor [Accessed: 11-03-2020]. <https://radmatt.itch.io/dialogue-designer>
- Rasa Technologies GmbH. (2020). Rasa [Accessed: 11-03-2020]. <https://rasa.com/docs/rasa-x/>
- Roberts, M. J. (2020). TADS 3.1.3 [Accessed: 08-05-2020]. <http://inform7.com/>
- Roessingh Research and Development. (2020). The WOOL Dialogue Platform [Accessed: 06-05-2020]. www.woolplatform.eu
- Saarland University. (2020). DialogOS [Accessed: 26-01-2020]. <https://www.dialogos.app/en/index.html>
- SAP Conversational AI. (2020). Bot Builder [Accessed: 11-03-2020]. <https://cai.tools.sap/bot-builder>
- Sebastian, J. & Richards, D. (2017). Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior*, 73, 479–488. <https://doi.org/10.1016/j.chb.2017.03.071>
- SecretLab. (2020). Yarn Spinner [Accessed: 28-05-2020]. <https://yarnspinner.dev/>
- Skorupski, J. & Mateas, M. (2010). Novice-friendly authoring of plan-based interactive storyboards. *AIIDE '10: Proceedings of the 6th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 174–179.
- Snaith, M., Lawrence, J., Pease, A. & Reed, C. A. (2020). A modular platform for argument and dialogue. *COMMA '20: Proceedings of the 9th International Conference on Computational Models of Argument*, 473–474. <https://doi.org/10.3233/FAIA200540>

- Spierling, U. & Szilas, N. (2009). Authoring issues beyond tools. *ICIDS '09: Proceedings of the 2nd International Conference on Interactive Digital Storytelling*, 5915 LNCS, 50–61. https://doi.org/10.1007/978-3-642-10643-9_9
- Spierling, U., Weiß, S. A. & Müller, W. (2006). Towards accessible authoring tools for interactive storytelling. *TIDSE '06: Proceedings of the International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, 4326 LNCS, 169–180. https://doi.org/10.1007/11944577_17
- Stefnisson, I. S. & Thue, D. (2018). Mimisbrunnur: AI-assisted authoring for interactive storytelling. *AIIDE '18: Proceedings of the 14th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, (Aiide), 236–242.
- Tanaka, H., Negoro, H., Iwasaka, H. & Nakamura, S. (2017). Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLoS ONE*, 12(8), 1–15. <https://doi.org/10.1371/journal.pone.0182151>
- textadventures.co.uk. (2020a). Quest [Accessed: 11-03-2020]. <http://textadventures.co.uk/quest/>
- textadventures.co.uk. (2020b). Squiffy [Accessed: 11-03-2020]. <http://textadventures.co.uk/squiffy>
- Texture. (2019). Texture Writer [Accessed: 08-05-2020]. <https://texturewriter.com/>
- The Genarrator Community. (2019). Genarrator [Accessed: 08-05-2020]. <http://genarrator.org/>
- The Ren'Py Community. (2019). Ren'Py [Accessed: 08-05-2020]. <https://renpy.en.uptodown.com>
- The Yarn Team. (2020). Dialogue Designer [Accessed: 11-03-2020]. <https://radmatt.itch.io/dialogue-designer>
- USC Institute for Creative Technologies. (2021). Virtual human toolkit [Accessed: 26-01-2020]. <https://vhtoolkit.ict.usc.edu/>
- van Velsen, L., Cabrita, M., op den Akker, H., Brandl, L., Isaac, J., Suárez, M., Gouveia, A., de Sousa, R. D., Rodrigues, A. M., Canhão, H., Evans, N., Blok, M., Alcobia, C. & Brodbeck, J. (2020). LEAVES (optimizing the mental health and resiliencE of older Adults that haVe lost thEir spouse via blended, online therapy): Proposal for an online service development and evaluation. *JMIR Research Protocols*, 9(9), e19344. <https://doi.org/10.2196/19344>
- van Velsen, M. (2008). Towards real-time authoring of believable agents in interactive narrative. *IVA '08: Proceedings of the 8th ACM International Workshop on Intelligent Virtual Agents*, 5208 LNCS, 81–88. https://doi.org/10.1007/978-3-540-85483-8_8
- van Waterschoot, J., Bruijnes, M., Flokstra, J., Reidsma, D., Davison, D., Theune, M. & Heylen, D. (2018). Flipper 2.0: A pragmatic dialogue engine for embodied conversational agents. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 43–50.
- van Welbergen, H., Yaghoubzadeh, R. & Kopp, S. (2014). AsapRealizer 2.0: The next steps in fluent behavior realization for ECAs. *IVA '14: Proceedings of the 14th ACM International Conference on Intelligent Virtual Agents*, 8637 LNCS, 449–462. https://doi.org/10.1007/978-3-319-09767-1_56
- van Wissen, A., Vinkers, C. & van Halteren, A. (2016). Developing a virtual coach for chronic patients: A user study on the impact of similarity, familiarity and realism. In A. Meschtscherjakow, B. De Ruyter, V. Fuchsberger, M. Murer & M. Tscheligi (Eds.), *Persuasive '16: Proceedings of the 11th international conference on persuasive technology* (pp. 263–275). Springer, Cham. <https://doi.org/10.1007/B978-0-12-373932-2.00008-9>
- Veletsianos, G. & Russell, G. S. (2014). Pedagogical agents. In J. M. Spector, M. D. Merrill, J. Elen & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology: Fourth edition* (pp. 1–12). <https://doi.org/10.1007/978-1-4614-3185-5>



- Wang, X., Sohn, S. S. & Kapadia, M. (2019). Towards a conversational interface for authoring intelligent virtual characters. *IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 127–129. <https://doi.org/10.1145/3308532.3329431>
- Wangberg, S. C., Bergmo, S., Trine & Johnson, J.-A. K. (2008). Adherence in internet-based interventions. *Patient Preference and Adherence*, 2, 57–66.
- Watson, A., Bickmore, T. W., Cange, A., Kulshreshtha, A. & Kvedar, J. (2012). An internet-based virtual coach to promote physical activity adherence in overweight adults: Randomized controlled trial. *Journal of Medical Internet Research*, 14(1), e1. <https://doi.org/10.2196/jmir.1629>
- Wit.ai, Inc. (2020). Wit.ai [Accessed: 11-03-2020]. <https://wit.ai/>
- Xu, A., Liu, Z., Guo, Y., Sinha, V. & Akkiraju, R. (2017). A new chatbot for customer service on social media. *CHI '17: Proceedings of the 2017 ACM Conference on Human Factors in Computing Systems*, 3506–3510. <https://doi.org/10.1145/3025453.3025496>
- Zhou, S., Bickmore, T. W., Paasche-Orlow, M. K. & Jack, B. W. (2014). Agent-user concordance and satisfaction with a virtual hospital discharge nurse. *IVA '14: Proceedings of the 14th ACM International Conference on Intelligent Virtual Agents*, 8637 LNCS, 528–541. https://doi.org/10.1007/978-3-319-09767-1_63

Intermezzo 1

- Beinema, T., Broekhuis, M., op den Akker, H., van Velsen, L., ter Stal, S., Pease, A., De Franco, D. & Ramchandra Kotnis, S. (2019). Deliverable 2.5: Evaluation of second functional prototype and updated requirements. <https://cordis.europa.eu/project/id/769553/results>
- Beinema, T., Huizing, G., op den Akker, H., Snaith, M., Fides Valero, A., Traver, V., Pelachaud, C. & Konsolakis, K. (2018). Deliverable 3.1: Initial knowledge base design and coaching strategies. <https://cordis.europa.eu/project/id/769553/results>
- Beinema, T. & op den Akker, H. (2019). Deliverable 3.3: Definition of tailored coaching strategies. <https://cordis.europa.eu/project/id/769553/results>
- Beinema, T., op den Akker, H., Broekhuis, M., Huizing, G., van Velsen, L., ter Stal, S. & van Loon, J. (2018). Deliverable 2.4: Evaluation results of first functional prototype and updated requirements. <https://cordis.europa.eu/project/id/769553/results>
- Beinema, T., op den Akker, H. & Hermens, H. J. (2018). Creating an artificial coaching engine for multi-domain conversational coaches in eHealth applications. In E. André, T. W. Bickmore, S. Vrochidis & L. Wanner (Eds.), *ICA-HoGeCa '18: Proceedings of the AAMAS workshop on intelligent conversation agents in home and geriatric care applications co-located with the Federated AI Meeting* (pp. 35–39). CEUR.
- van Velsen, L., Broekhuis, M., Jansen-Kosterink, S. M. & op den Akker, H. (2019). Tailoring persuasive eHealth strategies for older adults on the basis of personal motivation: An online survey. *Journal of Medical Internet Research*, 21(9), e11759. <https://doi.org/10.2196/11759>

Chapter 3

- Abdullah, A. S., Gaehde, S. & Bickmore, T. W. (2018). A tablet based embodied conversational agent to promote smoking cessation among veterans: A feasibility study. *Journal of Epidemiology and Global Health*, 8(3-4), 225–230. <https://doi.org/10.2991/j.jegh.2018.08.104>

REFERENCES

- Abdulrahman, A. & Richards, D. (2020). Modelling working alliance using user-aware explainable embodied conversational agents for behavior change: Framework and empirical evaluation. *ICIS '19: Proceedings of the 40th International Conference on Information Systems*, 1–17.
- Achterkamp, R., Cabrita, M., op den Akker, H., Hermens, H. J. & Vollenbroek-Hutten, M. M. R. (2013). Promoting a healthy lifestyle: Towards an improved personalized feedback approach. *HealthCom '13: Proceedings of the 15th IEEE International Conference on e-Health Networking, Applications and Services*, 725–727. <https://doi.org/10.1109/HealthCom.2013.6720772>
- Acosta, J. C. & Ward, N. G. (2011). Achieving rapport with turn-by-turn, user-responsive emotional coloring. *Speech Communication*, 53(9–10), 1137–1148. <https://doi.org/10.1016/j.specom.2010.11.006>
- André, E. & Pelachaud, C. (2010). Interacting with embodied conversational agents. *Speech technology* (pp. 123–149). https://doi.org/10.1007/978-0-387-73819-2_8
- André, E. & Rist, T. (2001). Presenting through performing: On the use of multiple lifelike characters in knowledge-based presentation systems. *Knowledge-Based Systems*, 14(1–2), 3–13. [https://doi.org/10.1016/S0950-7051\(00\)00096-4](https://doi.org/10.1016/S0950-7051(00)00096-4)
- Asbjørnsen, R. A., Smedsrød, M. L., Nes, L. S., Wentzel, J., Varsi, C., Hjelmesæth, J. & van Gemert-Pijnen, J. E. W. C. (2019). Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: Scoping review. *Journal of Medical Internet Research*, 21(6), e14265. <https://doi.org/10.2196/14265>
- Beinema, T., op den Akker, H. & Hermens, H. J. (2018). Creating an artificial coaching engine for multi-domain conversational coaches in eHealth applications. In E. André, T. W. Bickmore, S. Vrochidis & L. Wanner (Eds.), *ICA-HoGeCa '18: Proceedings of the AAMAS workshop on intelligent conversation agents in home and geriatric care applications co-located with the Federated AI Meeting* (pp. 35–39). CEUR.
- Bickmore, T. W., Gruber, A. & Picard, R. W. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 59(1), 21–30. <https://doi.org/10.1016/j.pec.2004.09.008>
- Bickmore, T. W. & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2), 293–327.
- Bickmore, T. W., Schulman, D. & Yin, L. (2009). Engagement vs. deceit: Virtual humans with human autobiographies. *IVA '09: Proceedings of the 9th ACM International Workshop on Intelligent Virtual Agents*, 5773 LNCS, 6–19. https://doi.org/10.1007/978-3-642-04380-2_4
- Bickmore, T. W., Schulman, D. & Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6), 648–666. <https://doi.org/10.1080/08839514.2010.492259>
- Bickmore, T. W., Trinh, H., Asadi, R. & Olafsson, S. (2018). Safety first: Conversational agents for health care. *Studies in Conversational UX Design*, 33–57. https://doi.org/10.1007/978-3-319-95579-7_3
- Bickmore, T. W., Utami, D., Matsuyama, R. & Paasche-Orlow, M. K. (2016). Improving access to online health information with conversational agents: A randomized controlled experiment. *Journal of Medical Internet Research*, 18(1), e5239. <https://doi.org/10.2196/jmir.5239>
- Bouton, M. E. (2014). Why behavior change is difficult to sustain. *Preventive Medicine*, 68, 29–36. <https://doi.org/10.1016/j.ypmed.2014.06.010>
- Castonguay, L. G., Constantino, M. J. & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy: Theory, Research, Practice, Training*, 43(3), 271–279. <https://doi.org/10.1037/0033-3204.43.3.271>

- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5), 752–766. <https://doi.org/10.1037/0022-3514.39.5.752>
- Chew, L. D., Bradley, K. A. & Boyko, E. J. (2004). Brief questions to identify patients with inadequate health literacy. *Family Medicine*, 36(8), 588–94.
- Das, K. S. J., Beinema, T., op den Akker, H. & Hermens, H. J. (2019). Generation of multi-party dialogues among embodied conversational agents to promote active living and healthy diet for subjects suffering from type 2 diabetes. *ICT4AWE '19: Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 297–304. <https://doi.org/10.5220/0007750602970304>
- de Vries, R. A. J., Truong, K. P., Kwint, S., Drossaert, C. H. C. & Evers, V. (2016). Crowd-designed motivation: Motivational messages for exercise adherence based on behavior change theory. *CHI '16: Proceedings of the 2016 ACM Conference on Human Factors in Computing Systems*, 297–308. <https://doi.org/10.1145/2858036.2858229>
- de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J. & Evers, V. (2017). A word of advice: How to tailor motivational text messages based on behavior change theory to personality and gender. *Personal and Ubiquitous Computing*, 21(4), 675–687. <https://doi.org/10.1007/s00779-017-1025-1>
- Dijkstra, A. (2008). The psychology of tailoring-ingredients in computer-tailored persuasion. *Social and Personality Psychology Compass*, 2(2), 765–784.
- Fadhil, A., Wang, Y. & Reiterer, H. (2019). Assistive conversational agent for health coaching: A validation study. *Methods of Information in Medicine*, 58(1), 9–23. <https://doi.org/10.1055/s-0039-1688757>
- Fisher, J. D. & Fisher, W. A. (1992). Changing aids-risk behavior. *Psychological Bulletin*, 111(3), 455–474. <https://doi.org/10.1037/0033-2909.111.3.455>
- Fogg, B. J. (2009). A behavior model for persuasive design. *PERSUASIVE '09: Proceedings of the 4th International Conference on Persuasive Technology*, 1–7. <https://doi.org/10.1145/1541948.1541999>
- Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., Jack, B. W. & Bickmore, T. W. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient Education and Counseling*, 100(9), 1720–1729. <https://doi.org/10.1016/j.pec.2017.04.015>
- Gourlan, M., Trouilloud, D. & Boiché, J. (2016). Motivational profiles for physical activity practice in adults with type 2 diabetes: A self-determination theory perspective. *Behavioral Medicine*, 42(4), 227–237. <https://doi.org/10.1080/08964289.2014.1001810>
- Hardiker, N. R. & Grant, M. J. (2011). Factors that influence public engagement with eHealth: A literature review. *International Journal of Medical Informatics*, 80(1), 1–12. <https://doi.org/10.1016/j.ijmedinf.2010.10.017>
- Hatcher, R. L. & Gillaspy, J. A. (2006). Development and validation of a revised short version of the working alliance inventory. *Psychotherapy Research*, 16(1), 12–25. <https://doi.org/10.1080/10503300500352500>
- Hawkins, R. P., Kreuter, M., Resnicow, K., Fishbein, M. & Dijkstra, A. (2008). Understanding tailoring in communicating about health. *Health Education Research*, 23(3), 454–466. <https://doi.org/10.1093/her/cyn004>
- Horvath, A. O. & Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *Journal of Counseling Psychology*, 36(2), 223–233. <https://doi.org/10.1037/0022-0167.36.2.223>
- Huber, M., van Vliet, M., Giezenberg, M., Winkens, B., Heerkens, Y., Dagnelie, P. C. & Knottnerus, J. A. (2016). Towards a 'patient-centred' operationalisation of the new dynamic concept of health: A mixed methods study. *BMJ Open*, 6(1), 1–11. <https://doi.org/10.1136/bmjopen-2015-010091>

REFERENCES

- Hurmuz, M. Z. M., Jansen-Kosterink, S. M., op den Akker, H. & Hermens, H. J. (2020). User experience and potential health effects of a conversational agent-based electronic health intervention: Protocol for an observational cohort study. *JMIR Research Protocols*, 9(4), e16641. <https://doi.org/10.2196/16641>
- Kamphorst, B. A. (2017). E-coaching systems: What they are, and what they aren't. *Personal and Ubiquitous Computing*, 21(4), 625–632. <https://doi.org/10.1007/s00779-017-1020-6>
- Kanthalraju, R. B., De Franco, D., Pease, A. & Pelachaud, C. (2018). Is two better than one? effects of multiple agents on user persuasion. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 255–262. <http://arxiv.org/abs/1904.05248>
- Kanthalraju, R. B., Pease, A., Reidsma, D., Pelachaud, C., Snaith, M., Bruijnes, M., Klaassen, R., Beinema, T., Huizing, G., Simonetti, D., Heylen, D. & op den Akker, H. (2019). Integrating argumentation with social conversation between multiple virtual coaches. *IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 203–205.
- Kaptein, M., de Ruyter, B., Markopoulos, P. & Aarts, E. (2012). Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(2), 1–25. <https://doi.org/10.1145/2209310.2209313>
- Kim, S., Lee, J. & Gweon, G. (2019). Comparing data from chatbot and web surveys effects of platform and conversational style on survey response quality. *CHI '19: Proceedings of the 2019 ACM Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3290605.3300316>
- King, A. C., Campero, I., Sheats, J. L., Castro Sweet, C. M., Garcia, D., Chazaro, A., Blanco, G., Hauser, M., Fierros, F., Ahn, D. K., Diaz, J., Done, M., Fernandez, J. & Bickmore, T. W. (2017). Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: The COMPASS trial design, methods, and baseline characteristics. *Contemporary Clinical Trials*, 61, 115–125. <https://doi.org/10.1016/j.cct.2017.07.020>
- Klaassen, R., Bul, K. C. M., op den Akker, R., van der Burg, G. J., Kato, P. M. & Di Bitonto, P. (2018). Design and evaluation of a pervasive coaching and gamification platform for young diabetes patients. *Sensors*, 18(2), 1–27. <https://doi.org/10.3390/s18020402>
- Kramer, L. L., ter Stal, S., Mulder, B. C., de Vet, E. & van Velsen, L. (2020). Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *Journal of medical Internet Research*, 22(2), e14058. <https://doi.org/10.2196/14058>
- Krämer, N. C., Hoffmann, L. & Kopp, S. (2010). Know your users! empirical results for tailoring an agent's nonverbal behavior to different user groups. *IVA '10: Proceedings of the 10th ACM International Conference on Intelligent Virtual Agents*, 6356 LNCS, 468–474. https://doi.org/10.1007/978-3-642-15892-6_50
- Krebs, P., Prochaska, J. O. & Rossi, J. S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. *Preventive Medicine*, 51(3-4), 214–221. <https://doi.org/10.1016/j.ypmed.2010.06.004>
- Ma, T., Sharifi, H. & Chattopadhyay, D. (2019). Virtual humans in health-related interventions: A meta-analysis. *CHI '19: Proceedings of the 2019 ACM Conference on Human Factors in Computing Systems*, 1–6.
- Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J. & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, 46(1), 81–95. <https://doi.org/10.1007/s12160-013-9486-6>
- Miltenberger, R. G. (2008). *Behaviour modification: Principles and practices*.

- Nass, C., Steuer, J. & Tauber, E. R. (1994). Computers are social actors. *CHI '94: Proceedings of the 1994 ACM Conference on Human Factors in Computing Systems*, 72–78. <https://doi.org/10.1145/259963.260288>
- Nijland, N. (2011). *Grounding eHealth – towards a holistic framework for sustainable eHealth technologies* (Doctoral dissertation). University of Twente. <https://doi.org/10.3990/1.9789036531337>
- Norman, P., Boer, H. & Seydel, E. R. (2005). Protection motivation theory. *Predicting health behaviour: Research and practice with social cognition models* (pp. 81–126). Maidenhead: Open University Press.
- Oinas-Kukkonen, H. & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, 24(1), 485–500.
- Olafsson, S., O'Leary, T. & Bickmore, T. W. (2019). Coerced change-talk with conversational agents promotes confidence in behavior change. *PervasiveHealth '19: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 31–40. <https://doi.org/10.1145/3329189.3329202>
- op den Akker, H., Jones, V. M. & Hermens, H. J. (2014). Tailoring real-time physical activity coaching systems: A literature survey and model. *User Modeling and User-Adapted Interaction*, 24(5), 351–392. <https://doi.org/10.1007/s11257-014-9146-y>
- op den Akker, H., op den Akker, R., Beinema, T., Banos, O., Heylen, D., Pease, A., Pelachaud, C., Traver Salcedo, V., Kyriazakos, S. & Hermens, H. J. (2018). Council of Coaches: A novel holistic behavior change coaching approach. *ICT4AWE '18: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 978–989. <https://doi.org/10.5220/0006787702190226>
- Paramythios, A., Weibelzahl, S. & Masthoff, J. (2010). Layered evaluation of interactive adaptive systems: Framework and formative methods. *User Modeling and User-Adapted Interaction*, 20(5), 383–453. <https://doi.org/10.1007/s11257-010-9082-4>
- Payne, J., Szymkowiak, A., Robertson, P. & Johnson, G. (2013). Gendering the machine: Preferred virtual assistant gender and realism in self-service. *IVA '13: Proceedings of the 13th ACM International Workshop on Intelligent Virtual Agents*, 8108 LNCS, 106–115. https://doi.org/10.1007/978-3-642-40415-3_9
- Pelletier, L. G., Rocchi, M. A., Vallerand, R. J., Deci, E. L. & Ryan, R. M. (2013). Validation of the revised sport motivation scale (SMS-II). *Psychology of Sport and Exercise*, 14(3), 329–341. <https://doi.org/10.1016/j.psychsport.2012.12.002>
- Perski, O., Blandford, A., West, R. & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, 7(2), 254–267. <https://doi.org/10.1007/s13142-016-0453-1>
- Pezzullo, L. G., Wiggins, J. B., Franksky, M. H., Min, W., Boyer, K. E., Mott, B. W., Wiebe, E. N. & Lester, J. C. (2017). "thanks alisha, keep in touch": Gender effects and engagement with virtual learning companions. *AIED '17: Proceedings of the International Conference on Artificial Intelligence in Education (AIED 2017)*, 10331 LNCS, 299–310. https://doi.org/10.1007/978-3-319-61425-0_25
- Pickard, M. D. (2012). *Persuasive embodied agents: Using embodied agents to change people's behavior, beliefs, and assessments* (Doctoral dissertation). University of Arizona.
- Prochaska, J. O. & Velicer, W. F. (1997). The transtheoretical model of health behaviour change. *American Journal of Health Promotion*, 12(1), 38–48. <https://doi.org/10.4278/0890-1171-12.1.38>
- Rafaeli, S. (1988). Interactivity: From new media to communication. *Sage annual review of communication research: Advancing communication science* (pp. 110–134).
- Roessingh Research and Development. (2020). The WOOL Dialogue Platform [Accessed: 06-05-2020]. www.woolplatform.eu

REFERENCES

- Rogers, R. W. (1989). Cognitive and psychological processes in fear appeals and attitude change: A revised theory of protection motivation. In J. Cacioppo (Ed.), *Social psychophysiology: A sourcebook* (pp. 153–176). Guildford Press.
- Ruijten, P. A., Ham, J. & Midden, C. J. (2014). Investigating the influence of social exclusion on persuasion by a virtual agent. *PERSUASIVE '14: Proceedings of the 9th International Conference on Persuasive Technology*, 8462 LNCS, 191–200. https://doi.org/10.1007/978-3-319-07127-5_17
- Ruttkay, Z., Dormann, C. & Noot, H. (2004). Embodied conversational agents on a common ground: A framework for design and evaluation. *From brows to trust: Evaluating embodied conversational agents* (pp. 27–66). Kluwer.
- Ryan, K., Dockray, S. & Linehan, C. (2019). A systematic review of tailored eHealth interventions for weight loss. *Digital Health*, 5, 1–29. <https://doi.org/10.1177/2055207619826685>
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Scholten, M. R., Kelders, S. M. & van Gemert-Pijnen, J. E. W. C. (2017). Self-guided web-based interventions: Scoping review on user needs and the potential of embodied conversational agents to address them. *Journal of Medical Internet Research*, 19(11), e383. <https://doi.org/10.2196/jmir.7351>
- Schulman, D. & Bickmore, T. W. (2009). Persuading users through counseling dialogue with a conversational agent. *PERSUASIVE '09: Proceedings of the 4th International Conference on Persuasive Technology*, 1–8. <https://doi.org/10.1145/1541948.1541983>
- Tedesco, D. P. & Tullis, T. S. (2006). A comparison of methods for eliciting post-task subjective ratings in usability testing. *Usability Professionals Association (UPA)*, 1–9.
- ter Stal, S., Tabak, M., op den Akker, H., Beinema, T. & Hermens, H. J. (2019). Who do you prefer? the effect of age, gender and role on users' first impressions of embodied conversational agents in eHealth. *International Journal of Human-Computer Interaction*, 36(9), 881–892. <https://doi.org/10.1080/10447318.2019.1699744>
- van Velsen, L., Broekhuis, M., Jansen-Kosterink, S. M. & op den Akker, H. (2019). Tailoring persuasive eHealth strategies for older adults on the basis of personal motivation: An online survey. *Journal of Medical Internet Research*, 21(9), e11759. <https://doi.org/10.2196/11759>
- van Wissen, A., Vinkers, C. & van Halteren, A. (2016). Developing a virtual coach for chronic patients: A user study on the impact of similarity, familiarity and realism. In A. Meschtscherjakow, B. De Ruyter, V. Fuchsberger, M. Murer & M. Tscheligi (Eds.), *Persuasive '16: Proceedings of the 11th international conference on persuasive technology* (pp. 263–275). Springer, Cham. <https://doi.org/10.1016/B978-0-12-373932-2.00008-9>
- Wangberg, S. C., Bergmo, S., Trine & Johnson, J.-A. K. (2008). Adherence in internet-based interventions. *Patient Preference and Adherence*, 2, 57–66.
- Watson, A., Bickmore, T. W., Cange, A., Kulshreshtha, A. & Kvedar, J. (2012). An internet-based virtual coach to promote physical activity adherence in overweight adults: Randomized controlled trial. *Journal of Medical Internet Research*, 14(1), e1. <https://doi.org/10.2196/jmir.1629>
- World Health Organization. (1946). *Constitution of the world health organization* (tech. rep.).
- World Health Organization. (2015). *World report on ageing and health* (tech. rep.). World Health Organization.
- World Health Organization. (2018). Noncommunicable diseases fact sheet [Accessed: March 2020].



- Xu, A., Liu, Z., Guo, Y., Sinha, V. & Akkiraju, R. (2017). A new chatbot for customer service on social media. *CHI '17: Proceedings of the 2017 ACM Conference on Human Factors in Computing Systems*, 3506–3510. <https://doi.org/10.1145/3025453.3025496>
- Zhou, S., Bickmore, T. W., Paasche-Orlow, M. K. & Jack, B. W. (2014). Agent-user concordance and satisfaction with a virtual hospital discharge nurse. *IVA '14: Proceedings of the 14th ACM International Conference on Intelligent Virtual Agents*, 8637 LNCS, 528–541. https://doi.org/10.1007/978-3-319-09767-1_63

Intermezzo 2

- Beinema, T., op den Akker, H., Jansen-Kosterink, S. M., ter Stal, S. & van den Boer, J. (2019). Deliverable 3.4: Final coaching actions and content. <https://cordis.europa.eu/project/id/769553/results>
- Das, K. S. J., Beinema, T., op den Akker, H. & Hermens, H. J. (2019). Generation of multi-party dialogues among embodied conversational agents to promote active living and healthy diet for subjects suffering from type 2 diabetes. *ICT4AWE '19: Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 297–304. <https://doi.org/10.5220/0007750602970304>
- Snaith, M., Conway, N., Beinema, T., De Franco, D., Pease, A., Kantharaju, R. B., Janier, M., Huizing, G., Pelachaud, C. & op den Akker, H. (2021). A multimodal corpus of simulated consultations between a patient and multiple healthcare professionals. *Language Resources and Evaluation*, 1–16. <https://doi.org/10.1007/s10579-020-09526-0>
- Snaith, M., De Franco, D., Beinema, T., op den Akker, H. & Pease, A. (2018). A dialogue game for multi-party goal-setting in health coaching. *COMMA '18: Proceedings of the 7th International Conference on Computational Models of Argument*, 337–344. <https://doi.org/10.3233/978-1-61499-906-5-337>
- van der Kamp, M., Beinema, T., Broekhuis, M., op den Akker, H., De Franco, D., Pease, A., Ramchandra Kotnis, S., Ellemose Vad, F. & Welan, E. (2019). Deliverable 2.6: Evaluation of the third functional prototype and updated requirements. <https://cordis.europa.eu/project/id/769553/results>

Chapter 4

- Abdullah, A. S., Gaehde, S. & Bickmore, T. W. (2018). A tablet based embodied conversational agent to promote smoking cessation among veterans: A feasibility study. *Journal of Epidemiology and Global Health*, 8(3-4), 225–230. <https://doi.org/10.2991/j.jegh.2018.08.104>
- Achterkamp, R., Cabrita, M., op den Akker, H., Hermens, H. J. & Vollenbroek-Hutten, M. M. R. (2013). Promoting a healthy lifestyle: Towards an improved personalized feedback approach. *HealthCom '13: Proceedings of the 15th IEEE International Conference on e-Health Networking, Applications and Services*, 725–727. <https://doi.org/10.1109/HealthCom.2013.6720772>
- Andersson, G., Carlbring, P., Berger, T., Almlöv, J. & Cuijpers, P. (2009). What makes internet therapy work? *Cognitive Behaviour Therapy*, 38(S1), 55–60. <https://doi.org/10.1080/16506070902916400>
- Beinema, T., Davison, D., Reidsma, D., Banos, O., Bruijnes, M., Donval, B., Fides Valero, Å., Heylen, D., Hofs, D., Huizing, G., Kantharaju, R. B., Klaassen, R., Kolkmeier, J., Konsolakis, K., Pease, A., Pelachaud, C., Simonetti, D., Snaith, M., Traver, V., ... op den Akker, H. (2021). Agents United: An open platform for multi-

REFERENCES

- agent conversational systems. *IVA '21: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*.
- Beinema, T., op den Akker, H., van Velsen, L. & Hermens, H. J. (2021). Tailoring coaching strategies to users' motivation in a multi-agent health coaching application. *Computers in Human Behavior*, 121, 106787. <https://doi.org/10.1016/j.chb.2021.106787>
- Benítez-Guijarro, A., Ruiz-Zafra, A., Callejas, Z., Medina-Medina, N., Benghazi, K. & Noguera, M. (2019). General architecture for development of virtual coaches for healthy habits monitoring and encouragement. *Sensors*, 19(1), 108. <https://doi.org/10.3390/s19010108>
- Beun, R. J., Brinkman, W.-P., Fitrianie, S., Griffioen-Both, F., Horsch, C., Lancee, J. & Spruit, S. (2016). Improving adherence in automated e-coaching. a case from insomnia therapy. *PERSUASIVE '16: Proceedings of the 11th International Conference on Persuasive Technology*, 9638 LNCS, 276–287. <https://doi.org/10.1007/978-3-319-31510-2>
- Beun, R. J., Fitrianie, S., Griffioen-Both, F., Spruit, S., Horsch, C., Lancee, J. & Brinkman, W.-P. (2017). Talk and tools: The best of both worlds in mobile user interfaces for e-coaching. *Personal and Ubiquitous Computing*, 21(4), 661–674. <https://doi.org/10.1007/s00779-017-1021-5>
- Beun, R. J., Griffioen-Both, F., Ahn, R., Fitrianie, S. & Lancee, J. (2014). Modeling interaction in automated e-coaching: A case from insomnia therapy. *COGNITIVE '14: Proceedings of the 6th International Conference on Advanced Cognitive Technologies and Applications*, 1–4.
- Bickmore, T. W. (2010). Relational agents for chronic disease self management. *Health Informatics: A Patient-Centered Approach to Diabetes*, 1–19.
- Bickmore, T. W. & Giorgino, T. (2006). Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, 39(5), 556–571. <https://doi.org/10.1016/j.jbi.2005.12.004>
- Bickmore, T. W., Gruber, A. & Picard, R. W. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 59(1), 21–30. <https://doi.org/10.1016/j.pec.2004.09.008>
- Bickmore, T. W., Mauer, D., Crespo, F. & Brown, T. (2008). Negotiating task interruptions with virtual agents for health behavior change. *AAMAS '08: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, 1217–1220.
- Bickmore, T. W. & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2), 293–327.
- Bickmore, T. W., Schulman, D. & Sidner, C. L. (2011). A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. *Journal of Biomedical Informatics*, 44(2), 183–197. <https://doi.org/10.1016/j.jbi.2010.12.006>
- Bickmore, T. W., Schulman, D. & Sidner, C. L. (2013). Automated interventions for multiple health behaviors using conversational agents. *Patient Education and Counseling*, 92(2), 142–148. <https://doi.org/10.1016/j.pec.2013.05.011>
- Bickmore, T. W., Schulman, D. & Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6), 648–666. <https://doi.org/10.1080/08839514.2010.492259>
- Bickmore, T. W., Silliman, R. A., Nelson, K., Cheng, D. M., Winter, M., Henault, L. & Paasche-Orlow, M. K. (2013). A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society*, 61(10), 1676–1683. <https://doi.org/10.1111/jgs.12449>
- Bickmore, T. W., Trinh, H., Asadi, R. & Olafsson, S. (2018). Safety first: Conversational agents for health care. *Studies in Conversational UX Design*, 33–57. https://doi.org/10.1007/978-3-319-95579-7_3

- Brinkman, W.-P. (2016). Virtual health agents for behavior change: Research perspectives and directions. *GREATS '16: Proceedings of the Workshop on Graphical and Robotic Embodied Agents for Therapeutic Systems, held during the 16th International Conference on Intelligent Virtual Agents (IVA '16)*, 1–17.
- Buimer, H. P., Tabak, M., van Velsen, L., van der Geest, T. & Hermens, H. J. (2017). Exploring determinants of patient adherence to a portal-supported oncology rehabilitation program: Interview and data log analyses. *JMIR Rehabilitation and Assistive Technologies*, 4(2), e12. <https://doi.org/10.2196/rehab.6294>
- Callejas, Z., Griol, D., McTear, M. F. & López-Cózar, R. (2014). A virtual coach for active ageing based on sentient computing and m-Health. *IWAAL '14: Proceedings of the International Workshop on Ambient Assisted Living*, 8868 LNCS, 59–66. <https://doi.org/10.1111/exsy.12454>
- Castonguay, L. G., Constantino, M. J. & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy: Theory, Research, Practice, Training*, 43(3), 271–279. <https://doi.org/10.1037/0033-3204.43.3.271>
- Chalaguine, L. A., Hunter, A., Potts, H. & Hamilton, F. (2019). Impact of argument type and concerns in argumentation with a chatbot. *ICTAI '19: Proceedings of the 31st International Conference on Tools with Artificial Intelligence*, 1557–1562. <https://doi.org/10.1109/ictai.2019.00224>
- de Kok, I., Hough, J., Frank, C., Schlangen, D. & Kopp, S. (2014). Dialogue structure of coaching sessions. *DialWatt '14: Proceedings of the 18th SemDial Workshop on the Semantics and Pragmatics of Dialogue*, 167–169. <http://pub.uni-bielefeld.de/download/2685764/2706495>
- DeSmet, A., De Bourdeaudhuij, I., Chastin, S., Crombez, G., Maddison, R. & Cardon, G. (2019). Adults' preferences for behavior change techniques and engagement features in a mobile app to promote 24-hour movement behaviors: Cross-sectional survey study. *JMIR mHealth and uHealth*, 7(12), e15707. <https://doi.org/10.2196/15707>
- de Vries, R. A. J., Truong, K. P., Kwint, S., Drossaert, C. H. C. & Evers, V. (2016). Crowd-designed motivation: Motivational messages for exercise adherence based on behavior change theory. *CHI '16: Proceedings of the 2016 ACM Conference on Human Factors in Computing Systems*, 297–308. <https://doi.org/10.1145/2858036.2858229>
- de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J. & Evers, V. (2017). A word of advice: How to tailor motivational text messages based on behavior change theory to personality and gender. *Personal and Ubiquitous Computing*, 21(4), 675–687. <https://doi.org/10.1007/s00779-017-1025-1>
- Fadhil, A., Wang, Y. & Reiterer, H. (2019). Assistive conversational agent for health coaching: A validation study. *Methods of Information in Medicine*, 58(1), 9–23. <https://doi.org/10.1055/s-0039-1688757>
- Fan, H. & Poole, M. S. (2006). What is personalization? perspectives on the design and implementation of personalization in information systems. *Journal of Organizational Computing and Electronic Commerce*, 16(3-4), 179–202. https://doi.org/10.1207/s15327744joce1603{\&}4_2
- Fitrianie, S., Griffioen-Both, F., Spruit, S., Lancee, J. & Beun, R. J. (2015). Automated dialogue generation for behavior intervention on mobile devices. *Procedia Computer Science*, 63, 236–243. <https://doi.org/10.1016/j.procs.2015.08.339>
- Fogg, B. J. (2002). Persuasive technology: Using computers to change what we think and do. *Ubiquity*, 2002(December), 2.
- Følstad, A. & Brandtzaeg, P. B. (2017). Chatbots and the new world of hci. *Interactions*, 24(4), 38–42. <https://doi.org/10.1145/3085558>
- Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., Jack, B. W. & Bickmore, T. W. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility

- randomized control trial. *Patient Education and Counseling*, 100(9), 1720–1729. <https://doi.org/10.1016/j.pec.2017.04.015>
- Glas, N. & Pelachaud, C. (2018). Topic management for an engaging conversational agent. *International Journal of Human-Computer Studies*, 120, 107–124. <https://doi.org/10.1016/j.ijhcs.2018.07.007>
- Gross, C., Schachner, T., Hasl, A., Kohlbrenner, D., Clarenbach, C. F., Wangenheim, F. V. & Kowatsch, T. (2021). Personalization of conversational agent-patient interaction styles for chronic disease management: Two consecutive cross-sectional questionnaire studies. *Journal of Medical Internet Research*, 23(5), e26643. <https://doi.org/10.2196/26643>
- Gupta, I., Di Eugenio, B., Ziebart, B., Liu, B., Gerber, B., Sharp, L., Davis, R. & Baiju, A. (2018). Towards building a virtual assistant health coach. *ICHI '18: Proceedings of the 2018 IEEE International Conference on Healthcare Informatics*, 419–421. <https://doi.org/10.1109/ICHI.2018.00081>
- Horvath, A. O. & Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *Journal of Counseling Psychology*, 36(2), 223–233. <https://doi.org/10.1037/0022-0167.36.2.223>
- Hurmuz, M. Z. M., Jansen-Kosterink, S. M., op den Akker, H. & Hermens, H. J. (2020). User experience and potential health effects of a conversational agent-based electronic health intervention: Protocol for an observational cohort study. *JMIR Research Protocols*, 9(4), e16641. <https://doi.org/10.2196/16641>
- Janz, N. K. & Becker, M. H. (1984). The Health Belief Model: A decade later. *Health Education Quarterly*, 11(1), 1–47. <https://doi.org/10.1177/109019818401100101>
- Jofré, N., Rodríguez, G., Alvarado, Y., Fernández, J. & Guerrero, R. (2017). Natural user interfaces: A physical activity trainer. In A. E. De Giusti (Ed.), *Cacic '17: Argentine congress of computer science* (pp. 122–131). Springer International Publishing.
- Kaptein, M., de Ruyter, B., Markopoulos, P. & Aarts, E. (2012). Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(2), 1–25. <https://doi.org/10.1145/2209310.2209313>
- King, A. C., Campero, I., Sheats, J. L., Castro Sweet, C. M., Garcia, D., Chazaro, A., Blanco, G., Hauser, M., Fierros, F., Ahn, D. K., Diaz, J., Done, M., Fernandez, J. & Bickmore, T. W. (2017). Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: The COMPASS trial design, methods, and baseline characteristics. *Contemporary Clinical Trials*, 61, 115–125. <https://doi.org/10.1016/j.cct.2017.07.020>
- Klein, M. C. A., Mogles, N. & van Wissen, A. (2013). An intelligent coaching system for therapy adherence. *IEEE Pervasive Computing*, 12(3), 22–30. <https://doi.org/10.1109/MPRV.2013.41>
- Kowatsch, T., Nißen, M., Rüegger, D., Stieger, M., Flückiger, C., Allemand, M. & von Wangenheim, F. (2018). The impact of interpersonal closeness cues in text-based healthcare chatbots on attachment bond and the desire to continue interacting: An experimental design. *ECIS '18: Proceedings of the 26th European Conference on Information Systems*.
- Kramer, L. L., ter Stal, S., Mulder, B. C., de Vet, E. & van Velsen, L. (2020). Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *Journal of medical Internet Research*, 22(2), e14058. <https://doi.org/10.2196/14058>
- Krämer, N. C., Hoffmann, L. & Kopp, S. (2010). Know your users! empirical results for tailoring an agent's nonverbal behavior to different user groups. *IVA '10: Proceedings of the 10th ACM International Conference on Intelligent Virtual Agents*, 6356 LNCS, 468–474. https://doi.org/10.1007/978-3-642-15892-6_50



- Krebs, P., Prochaska, J. O. & Rossi, J. S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. *Preventive Medicine*, 51(3-4), 214–221. <https://doi.org/10.1016/j.ypmed.2010.06.004>
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J. Y. C., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S. & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, 25(9), 1248–1258. <https://doi.org/10.1093/jamia/ocy072>
- Locke, E. A. & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717. <https://doi.org/10.1037/0003-066X.57.9.705>
- McTear, M. F., Callejas, Z. & Griol, D. (2016). Evaluating the conversational interface. *The conversational interface* (pp. 379–402). https://doi.org/10.1007/978-3-32967-3_17
- Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J. & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, 46(1), 81–95. <https://doi.org/10.1007/s12160-013-9486-6>
- Miltenberger, R. G. (2008). *Behaviour modification: Principles and practices*.
- Montenegro, C., López Zorrilla, A., Mikel Olaso, J., Santana, R., Justo, R., Lozano, J. A. & Torres, M. I. (2019). A dialogue-act taxonomy for a virtual coach designed to improve the life of elderly. *Multimodal Technologies and Interaction*, 3(3), 52. <https://doi.org/10.3390/mti3030052>
- Nass, C., Steuer, J. & Tauber, E. R. (1994). Computers are social actors. *CHI '94: Proceedings of the 1994 ACM Conference on Human Factors in Computing Systems*, 72–78. <https://doi.org/10.1145/259963.260288>
- Nijland, N. (2011). *Grounding eHealth – towards a holistic framework for sustainable eHealth technologies* (Doctoral dissertation). University of Twente. <https://doi.org/10.3990/1.9789036531337>
- Norman, P., Boer, H. & Seydel, E. R. (2005). Protection motivation theory. *Predicting health behaviour: Research and practice with social cognition models* (pp. 81–126). Maidenhead: Open University Press.
- Oinas-Kukkonen, H. & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, 24(1), 485–500.
- Olafsson, S., O'Leary, T. & Bickmore, T. W. (2019). Coerced change-talk with conversational agents promotes confidence in behavior change. *PervasiveHealth '19: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 31–40. <https://doi.org/10.1145/3329189.3329202>
- op den Akker, H., Jones, V. M. & Hermens, H. J. (2014). Tailoring real-time physical activity coaching systems: A literature survey and model. *User Modeling and User-Adapted Interaction*, 24(5), 351–392. <https://doi.org/10.1007/s11257-014-9146-y>
- op den Akker, H., op den Akker, R., Beinema, T., Banos, O., Heylen, D., Pease, A., Pelachaud, C., Traver Salcedo, V., Kyriazakos, S. & Hermens, H. J. (2018). Council of Coaches: A novel holistic behavior change coaching approach. *ICT4AWE '18: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 978–989. <https://doi.org/10.5220/0006787702190226>
- op den Akker, R., Klaassen, R. & Nijholt, A. (2016). Virtual coaches for healthy lifestyle. *Toward robotic socially believable behaving systems - volume ii* (pp. 121–149). <https://doi.org/10.1007/978-3-319-31053-4>
- Perski, O., Blandford, A., West, R. & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from

REFERENCES

- critical interpretive synthesis. *Translational Behavioral Medicine*, 7(2), 254–267. <https://doi.org/10.1007/s13142-016-0453-1>
- Riou, M. (2015). A methodology for the identification of topic transitions in interaction. *Discours*, (16), 3–28. <https://doi.org/10.4000/discours.8997>
- Roessingh Research and Development. (2020). The WOOL Dialogue Platform [Accessed: 06-05-2020]. www.woolplatform.eu
- Ruttkay, Z. & van Welbergen, H. (2008). Elbows higher! performing, observing and correcting exercises by a virtual trainer. *IVA '08: Proceedings of the 8th ACM International Workshop on Intelligent Virtual Agents*, 5208 LNCS, 409–416. https://doi.org/10.1007/978-3-540-85483-8_41
- Ryan, K., Dockray, S. & Linehan, C. (2019). A systematic review of tailored eHealth interventions for weight loss. *Digital Health*, 5, 1–29. <https://doi.org/10.1177/2055207619826685>
- Schulman, D. & Bickmore, T. W. (2009). Persuading users through counseling dialogue with a conversational agent. *PERSUASIVE '09: Proceedings of the 4th International Conference on Persuasive Technology*, 1–8. <https://doi.org/10.1145/1541948.1541983>
- Schwarzer, R., Lippke, S. & Luszczynska, A. (2011). Mechanisms of health behavior change in persons with chronic illness or disability: The health action process approach (HAPA). *Rehabilitation Psychology*, 56(3), 161–170. <https://doi.org/10.1037/a0024509>
- Sebastian, J. & Richards, D. (2017). Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior*, 73, 479–488. <https://doi.org/10.1016/j.chb.2017.03.071>
- Smith, C., Cavazza, M., Charlton, D., Zhang, L., Turunen, M. & Hakulinen, J. (2008). Integrating planning and dialogue in a lifestyle agent. *IVA '08: Proceedings of the 8th ACM International Workshop on Intelligent Virtual Agents*, 5208 LNCS, 146–153. https://doi.org/10.1007/978-3-540-85483-8_15
- Smith, C., Charlton, D., Zhang, L., Cavazza, M., Hakulinen, J. & Turunen, M. (2008). An embodied conversational agent as a lifestyle advisor. *AAMAS '08: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, 1663–1664.
- Smith, C., Crook, N., Dobnik, S., Charlton, D., Boye, J., Pulman, S., Santos de la Camara, R., Turunen, M., Benyon, D., Bradley, J., Gambäck, B., Hansen, P., Mival, O., Webb, N. & Cavazza, M. (2011). Interaction strategies for an affective conversational agent. *Presence*, 20(5), 395–411. https://doi.org/10.1162/PRES_a_00063
- Snaith, M., De Franco, D., Beinema, T., op den Akker, H. & Pease, A. (2018). A dialogue game for multi-party goal-setting in health coaching. *COMMA '18: Proceedings of the 7th International Conference on Computational Models of Argument*, 337–344. <https://doi.org/10.3233/978-1-61499-906-5-337>
- Snaith, M., Lawrence, J., Pease, A. & Reed, C. A. (2020). A modular platform for argument and dialogue. *COMMA '20: Proceedings of the 9th International Conference on Computational Models of Argument*, 473–474. <https://doi.org/10.3233/FAIA200540>
- Starr, J. (2008). *The coaching manual: The definitive guide to the process, principles, and skills of personal coaching*. Pearson Education.
- ter Stal, S., Broekhuis, M., van Velsen, L., Hermens, H. J. & Tabak, M. (2020). Embodied conversational agent appearance for health assessment of older adults: Explorative study. *JMIR Human Factors*, 7(3), e19987. <https://doi.org/10.2196/19987>
- Uribe, J. A., Duitama, J. F. & Gaviria Gómez, N. (2011). Personalized message emission in a mobile application for supporting therapeutic adherence. *HealthCom '11: Proceedings of the 13th IEEE International Conference on e-Health Networking, Applications and Services*, 15–20. <https://doi.org/10.1109/HEALTH.2011.6026734>
- van Velsen, L., Broekhuis, M., Jansen-Kosterink, S. M. & op den Akker, H. (2019). Tailoring persuasive eHealth strategies for older adults on the basis of personal motivation:



- An online survey. *Journal of Medical Internet Research*, 21(9), e11759. <https://doi.org/10.2196/11759>
- van Velsen, L., Cabrita, M., op den Akker, H., Brandl, L., Isaac, J., Suárez, M., Gouveia, A., de Sousa, R. D., Rodrigues, A. M., Canhão, H., Evans, N., Blok, M., Alcobia, C. & Brodbeck, J. (2020). LEAVES (optimizing the mental health and resilience of older Adults that have lost their spouse via blended, online therapy): Proposal for an online service development and evaluation. *JMIR Research Protocols*, 9(9), e19344. <https://doi.org/10.2196/19344>
- Wangberg, S. C., Bergmo, S., Trine & Johnson, J.-A. K. (2008). Adherence in internet-based interventions. *Patient Preference and Adherence*, 2, 57–66.
- Watson, A., Bickmore, T. W., Cange, A., Kulshreshtha, A. & Kvedar, J. (2012). An internet-based virtual coach to promote physical activity adherence in overweight adults: Randomized controlled trial. *Journal of Medical Internet Research*, 14(1), e1. <https://doi.org/10.2196/jmir.1629>
- Yasavur, U., Lisetti, C. L. & Rishe, N. (2013). Modeling brief alcohol intervention dialogue with MPDs for delivery by ECAs. *IVA '13: Proceedings of the 13th ACM International Workshop on Intelligent Virtual Agents*, 8108 LNCS, 92–105. https://doi.org/10.1007/978-3-642-40415-3_8
- Zhang, Z. & Bickmore, T. W. (2018). Medical shared decision making with a virtual agent. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 113–118. <https://doi.org/10.1145/3267851.3267883>

Intermezzo 3

Beinema, T., Davison, D., Reidsma, D., Banos, O., Bruijnes, M., Donval, B., Fides Valero, Á., Heylen, D., Hofs, D., Huizing, G., Kantharaju, R. B., Klaassen, R., Kolkmeier, J., Konsolakis, K., Pease, A., Pelachaud, C., Simonetti, D., Snaith, M., Traver, V., ... op den Akker, H. (2021). Agents United: An open platform for multi-agent conversational systems. *IVA '21: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*.

Chapter 5

- Abdullah, A. S., Gaehde, S. & Bickmore, T. W. (2018). A tablet based embodied conversational agent to promote smoking cessation among veterans: A feasibility study. *Journal of Epidemiology and Global Health*, 8(3-4), 225–230. <https://doi.org/10.2991/j.jegh.2018.08.104>
- Agarwal, R. & Prasad, J. (1998). A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research*, 9(2), 204–215. <https://doi.org/10.1287/isre.9.2.204>
- Andersson, G., Carlbring, P., Berger, T., Almlöv, J. & Cuijpers, P. (2009). What makes internet therapy work? *Cognitive Behaviour Therapy*, 38(S1), 55–60. <https://doi.org/10.1080/16506070902916400>
- André, E. & Pelachaud, C. (2010). Interacting with embodied conversational agents. *Speech technology* (pp. 123–149). https://doi.org/10.1007/978-0-387-73819-2_8
- André, E. & Rist, T. (2001). Presenting through performing: On the use of multiple lifelike characters in knowledge-based presentation systems. *Knowledge-Based Systems*, 14(1-2), 3–13. [https://doi.org/10.1016/S0950-7051\(00\)00096-4](https://doi.org/10.1016/S0950-7051(00)00096-4)

REFERENCES

- Beinema, T., op den Akker, H. & Hermens, H. J. (2018). Creating an artificial coaching engine for multi-domain conversational coaches in eHealth applications. In E. André, T. W. Bickmore, S. Vrochidis & L. Wanner (Eds.), *ICA-HoGeCa '18: Proceedings of the AAMAS workshop on intelligent conversation agents in home and geriatric care applications co-located with the Federated AI Meeting* (pp. 35–39). CEUR.
- Beinema, T., op den Akker, H., van Velsen, L. & Hermens, H. J. (2021). Tailoring coaching strategies to users' motivation in a multi-agent health coaching application. *Computers in Human Behavior*, 121, 106787. <https://doi.org/10.1016/j.chb.2021.106787>
- Benítez-Guijarro, A., Ruiz-Zafra, A., Callejas, Z., Medina-Medina, N., Benghazi, K. & Noguera, M. (2019). General architecture for development of virtual coaches for healthy habits monitoring and encouragement. *Sensors*, 19(1), 108. <https://doi.org/10.3390/s19010108>
- Bickmore, T. W. & Giorgino, T. (2006). Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, 39(5), 556–571. <https://doi.org/10.1016/j.jbi.2005.12.004>
- Bickmore, T. W., Mauer, D., Crespo, F. & Brown, T. (2007). Persuasion, task interruption and health regimen adherence. *PERSUASIVE '07: Proceedings of the 2nd International Conference on Persuasive Technology*, 4744 LNCS, 1–11. https://doi.org/10.1007/978-3-540-77006-0_1
- Bickmore, T. W., Schulman, D. & Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6), 648–666. <https://doi.org/10.1080/08839514.2010.492259>
- Bickmore, T. W., Trinh, H., Asadi, R. & Olafsson, S. (2018). Safety first: Conversational agents for health care. *Studies in Conversational UX Design*, 33–57. https://doi.org/10.1007/978-3-319-95579-7_3
- Bickmore, T. W., Utami, D., Matsuyama, R. & Paasche-Orlow, M. K. (2016). Improving access to online health information with conversational agents: A randomized controlled experiment. *Journal of Medical Internet Research*, 18(1), e5239. <https://doi.org/10.2196/jmir.5239>
- Bouton, M. E. (2014). Why behavior change is difficult to sustain. *Preventive Medicine*, 68, 29–36. <https://doi.org/10.1016/j.ypmed.2014.06.010>
- Brinkman, W.-P. (2016). Virtual health agents for behavior change: Research perspectives and directions. *GREATS '16: Proceedings of the Workshop on Graphical and Robotic Embodied Agents for Therapeutic Systems, held during the 16th International Conference on Intelligent Virtual Agents (IVA '16)*, 1–17.
- Buimer, H. P., Tabak, M., van Velsen, L., van der Geest, T. & Hermens, H. J. (2017). Exploring determinants of patient adherence to a portal-supported oncology rehabilitation program: Interview and data log analyses. *JMIR Rehabilitation and Assistive Technologies*, 4(2), e12. <https://doi.org/10.2196/rehab.6294>
- Chew, L. D., Bradley, K. A. & Boyko, E. J. (2004). Brief questions to identify patients with inadequate health literacy. *Family Medicine*, 36(8), 588–94.
- Cole-Lewis, H., Ezeanochie, N. & Turgiss, J. (2019). Understanding health behavior technology engagement: Pathway to measuring digital behavior change interventions. *JMIR Formative Research*, 3(4), 1–10. <https://doi.org/10.2196/14052>
- Couper, M. P., Alexander, G. L., Zhang, N., Little, R. J. A., Maddy, N., Nowak, M. A., McClure, J. B., Calvi, J. J., Rolnick, S. J., Stopponi, M. A. & Johnson, C. C. (2010). Engagement and retention: Measuring breadth and depth of participant use of an online intervention. *Journal of Medical Internet Research*, 12(4), e52. <https://doi.org/10.2196/jmir.1430>
- Crutzen, R., Cyr, D. & de Vries, N. K. (2011). Bringing loyalty to e-Health: Theory validation using three internet-delivered interventions. *Journal of Medical Internet Research*, 13(3), e73. <https://doi.org/10.2196/jmir.1837>

- Das, K. S. J., Beinema, T., op den Akker, H. & Hermens, H. J. (2019). Generation of multi-party dialogues among embodied conversational agents to promote active living and healthy diet for subjects suffering from type 2 diabetes. *ICT4AWE '19: Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 297–304. <https://doi.org/10.5220/0007750602970304>
- de Vries, R. A. J., Truong, K. P., Kwint, S., Drossaert, C. H. C. & Evers, V. (2016). Crowd-designed motivation: Motivational messages for exercise adherence based on behavior change theory. *CHI '16: Proceedings of the 2016 ACM Conference on Human Factors in Computing Systems*, 297–308. <https://doi.org/10.1145/2858036.2858229>
- Ekeland, A. G., Bowes, A. & Flottorp, S. (2010). Effectiveness of telemedicine: A systematic review of reviews. *International Journal of Medical Informatics*, 79(11), 736–771. <https://doi.org/10.1016/j.ijmedinf.2010.08.006>
- Ekeland, A. G., Bowes, A. & Flottorp, S. (2012). Methodologies for assessing telemedicine: A systematic review of reviews. *International Journal of Medical Informatics*, 81(1), 1–11. <https://doi.org/10.1016/j.ijmedinf.2011.10.009>
- Fadhil, A., Wang, Y. & Reiterer, H. (2019). Assistive conversational agent for health coaching: A validation study. *Methods of Information in Medicine*, 58(1), 9–23. <https://doi.org/10.1055/s-0039-1688757>
- Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., Jack, B. W. & Bickmore, T. W. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient Education and Counseling*, 100(9), 1720–1729. <https://doi.org/10.1016/j.pec.2017.04.015>
- Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J. & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54, 170–179. <https://doi.org/10.1016/j.chb.2015.07.045>
- Hardiker, N. R. & Grant, M. J. (2011). Factors that influence public engagement with eHealth: A literature review. *International Journal of Medical Informatics*, 80(1), 1–12. <https://doi.org/10.1016/j.ijmedinf.2010.10.017>
- Hayashi, Y. & Ogawa, H. (2012). Facilitating creative interpretations on collaboration with multiple conversational agents. *APCHI '12: Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction*, 443–449.
- Huber, M., van Vliet, M., Giezenberg, M., Winkens, B., Heerkens, Y., Dagnelie, P. C. & Knottnerus, J. A. (2016). Towards a 'patient-centred' operationalisation of the new dynamic concept of health: A mixed methods study. *BMJ Open*, 6(1), 1–11. <https://doi.org/10.1136/bmjopen-2015-010091>
- Hurmuz, M. Z. M., Jansen-Kosterink, S. M., op den Akker, H. & Hermens, H. J. (2020). User experience and potential health effects of a conversational agent-based electronic health intervention: Protocol for an observational cohort study. *JMIR Research Protocols*, 9(4), e16641. <https://doi.org/10.2196/16641>
- Joseph-Williams, N., Elwyn, G. & Edwards, A. (2014). Knowledge is not power for patients: A systematic review and thematic synthesis of patient-reported barriers and facilitators to shared decision making. *Patient Education and Counseling*, 94(3), 291–309. <https://doi.org/10.1016/j.pec.2013.10.031>
- Kairy, D., Lehoux, P., Vincent, C. & Visintin, M. (2009). A systematic review of clinical outcomes, clinical process, healthcare utilization and costs associated with telerehabilitation. *Disability and Rehabilitation*, 31(6), 427–447. <https://doi.org/10.1080/09638280802062553>
- Kamphorst, B. A. (2017). E-coaching systems: What they are, and what they aren't. *Personal and Ubiquitous Computing*, 21(4), 625–632. <https://doi.org/10.1007/s00779-017-1020-6>

REFERENCES

- Kanthalraju, R. B., De Franco, D., Pease, A. & Pelachaud, C. (2018). Is two better than one? effects of multiple agents on user persuasion. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 255–262. <http://arxiv.org/abs/1904.05248>
- Kanthalraju, R. B., Pease, A., Reidsma, D., Pelachaud, C., Snaith, M., Bruijnes, M., Klaassen, R., Beinema, T., Huizing, G., Simonetti, D., Heylen, D. & op den Akker, H. (2019). Integrating argumentation with social conversation between multiple virtual coaches. *IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 203–205.
- King, A. C., Campero, I., Sheats, J. L., Castro Sweet, C. M., Garcia, D., Chazaro, A., Blanco, G., Hauser, M., Fierros, F., Ahn, D. K., Diaz, J., Done, M., Fernandez, J. & Bickmore, T. W. (2017). Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: The COMPASS trial design, methods, and baseline characteristics. *Contemporary Clinical Trials*, 61, 115–125. <https://doi.org/10.1016/j.cct.2017.07.020>
- Klaassen, R., Bul, K. C. M., op den Akker, R., van der Burg, G. J., Kato, P. M. & Di Bitonto, P. (2018). Design and evaluation of a pervasive coaching and gamification platform for young diabetes patients. *Sensors*, 18(2), 1–27. <https://doi.org/10.3390/s18020402>
- Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A. & Murphy, S. A. (2015). Micro-randomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*, 34, 1220–1228. <https://doi.org/10.1016/j.physbeh.2017.03.040>
- Kohl, L. F. M., Crutzen, R. & de Vries, N. K. (2013). Online prevention aimed at lifestyle behaviors: A systematic review of reviews. *Journal of Medical Internet Research*, 15(7), e146. <https://doi.org/10.2196/jmir.2665>
- Kramer, L. L., Mulder, B. C., van Velsen, L. & de Vet, E. (2021). Use and effect of web-based embodied conversational agents for improving eating behavior and decreasing loneliness among community-dwelling older adults: Protocol for a randomized controlled trial. *JMIR Research Protocols*, 10(1), e22186. <https://doi.org/10.2196/22186>
- Kramer, L. L., ter Stal, S., Mulder, B. C., de Vet, E. & van Velsen, L. (2020). Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *Journal of medical Internet Research*, 22(2), e14058. <https://doi.org/10.2196/14058>
- Krämer, N. C., Hoffmann, L. & Kopp, S. (2010). Know your users! empirical results for tailoring an agent's nonverbal behavior to different user groups. *IVA '10: Proceedings of the 10th ACM International Conference on Intelligent Virtual Agents*, 6356 LNCS, 468–474. https://doi.org/10.1007/978-3-642-15892-6_50
- Krebs, P., Prochaska, J. O. & Rossi, J. S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. *Preventive Medicine*, 51(3-4), 214–221. <https://doi.org/10.1016/j.ypmed.2010.06.004>
- LaPlante, C. & Peng, W. (2011). A systematic review of e-health interventions for physical activity: An analysis of study design, intervention characteristics, and outcomes. *Telemedicine and e-Health*, 17(7), 509–523. <https://doi.org/10.1089/tmj.2011.0013>
- Lee, M. K., Kiesler, S. & Forlizzi, J. (2010). Receptionist or information kiosk: How do people talk with a robot? *CSCW '10: Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 31–40. <https://doi.org/10.1145/1718918.1718927>
- Ma, T., Sharifi, H. & Chattopadhyay, D. (2019). Virtual humans in health-related interventions: A meta-analysis. *CHI '19: Proceedings of the 2019 ACM Conference on Human Factors in Computing Systems*, 1–6.
- Nakamura, J. & Csikszentmihalyi, M. (2002). The concept of flow. *Handbook of positive psychology* (pp. 89–105). <https://doi.org/10.1002/9780470172698.ch19>



- Nijland, N. (2011). *Grounding eHealth – towards a holistic framework for sustainable eHealth technologies* (Doctoral dissertation). University of Twente. <https://doi.org/10.3990/1.9789036531337>
- O'Brien, H. L. & Toms, E. G. (2013). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938–955. <https://doi.org/10.1002/asi>
- Olafsson, S., O'Leary, T. & Bickmore, T. W. (2019). Coerced change-talk with conversational agents promotes confidence in behavior change. *PervasiveHealth '19: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 31–40. <https://doi.org/10.1145/3329189.3329202>
- op den Akker, H., Jones, V. M. & Hermens, H. J. (2014). Tailoring real-time physical activity coaching systems: A literature survey and model. *User Modeling and User-Adapted Interaction*, 24(5), 351–392. <https://doi.org/10.1007/s11257-014-9146-y>
- op den Akker, H., op den Akker, R., Beinema, T., Banos, O., Heylen, D., Pease, A., Pelachaud, C., Traver Salcedo, V., Kyriazakos, S. & Hermens, H. J. (2018). Council of Coaches: A novel holistic behavior change coaching approach. *ICT4AWE '18: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 978–989. <https://doi.org/10.5220/0006787702190226>
- Payne, J., Szymkowiak, A., Robertson, P. & Johnson, G. (2013). Gendering the machine: Preferred virtual assistant gender and realism in self-service. *IVA '13: Proceedings of the 13th ACM International Workshop on Intelligent Virtual Agents*, 8108 LNCS, 106–115. https://doi.org/10.1007/978-3-642-40415-3_9
- Perski, O., Blandford, A., West, R. & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, 7(2), 254–267. <https://doi.org/10.1007/s13142-016-0453-1>
- Pezzullo, L. G., Wiggins, J. B., Frankosky, M. H., Min, W., Boyer, K. E., Mott, B. W., Wiebe, E. N. & Lester, J. C. (2017). "thanks alisha, keep in touch": Gender effects and engagement with virtual learning companions. *AIED '17: Proceedings of the International Conference on Artificial Intelligence in Education (AIED 2017)*, 10331 LNCS, 299–310. https://doi.org/10.1007/978-3-319-61425-0_25
- Ring, L., Barry, B., Totzke, K. & Bickmore, T. W. (2013). Addressing loneliness and isolation in older adults: Proactive affective agents provide better support. *ACII '13: Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction*, 61–66. <https://doi.org/10.1109/ACII.2013.17>
- Roessingh Research and Development. (2020). The WOOL Dialogue Platform [Accessed: 06-05-2020]. www.woolplatform.eu
- Ruttkay, Z., Dormann, C. & Noot, H. (2004). Embodied conversational agents on a common ground: A framework for design and evaluation. *From brows to trust: Evaluating embodied conversational agents* (pp. 27–66). Kluwer.
- Ryan, K., Dockray, S. & Linehan, C. (2019). A systematic review of tailored eHealth interventions for weight loss. *Digital Health*, 5, 1–29. <https://doi.org/10.1177/2055207619826685>
- Sebastian, J. & Richards, D. (2017). Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior*, 73, 479–488. <https://doi.org/10.1016/j.chb.2017.03.071>
- Trinh, H., Shamekh, A., Kimani, E. & Bickmore, T. W. (2018). Predicting user engagement in longitudinal interventions with virtual agents. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 9–16. <https://doi.org/10.1145/3267851.3267909>
- van Velsen, L., Broekhuis, M., Jansen-Kosterink, S. M. & op den Akker, H. (2019). Tailoring persuasive eHealth strategies for older adults on the basis of personal motivation:

- An online survey. *Journal of Medical Internet Research*, 21(9), e11759. <https://doi.org/10.2196/11759>
- van Velsen, L., Cabrita, M., op den Akker, H., Brandl, L., Isaac, J., Suárez, M., Gouveia, A., de Sousa, R. D., Rodrigues, A. M., Canhão, H., Evans, N., Blok, M., Alcobia, C. & Brodbeck, J. (2020). LEAVES (optimizing the mental health and resilience of older Adults that have lost their spouse via blended, online therapy): Proposal for an online service development and evaluation. *JMIR Research Protocols*, 9(9), e19344. <https://doi.org/10.2196/19344>
- Wangberg, S. C., Bergmo, S., Trine & Johnson, J.-A. K. (2008). Adherence in internet-based interventions. *Patient Preference and Adherence*, 2, 57–66.
- Watson, A., Bickmore, T. W., Cange, A., Kulshreshtha, A. & Kvedar, J. (2012). An internet-based virtual coach to promote physical activity adherence in overweight adults: Randomized controlled trial. *Journal of Medical Internet Research*, 14(1), e1. <https://doi.org/10.2196/jmir.1629>
- World Health Organization. (1946). *Constitution of the world health organization* (tech. rep.).
- Xiao, J., Stasko, J. & Catrambone, R. (2002). Embodied conversational agents as a UI paradigm: A framework for evaluation. *Proceedings of the AAMAS02 Workshop on Embodied Conversational Agents - Let's Specify and Evaluate Them!*
- Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., Merchant, G. C., Naughton, F. & Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine*, 51(5), 833–842. <https://doi.org/10.1016/j.amepre.2016.06.015>
- Zhang, Z. & Bickmore, T. W. (2018). Medical shared decision making with a virtual agent. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 113–118. <https://doi.org/10.1145/3267851.3267883>

Chapter 6

- Abdulrahman, A. & Richards, D. (2020). Modelling working alliance using user-aware explainable embodied conversational agents for behavior change: Framework and empirical evaluation. *ICIS '19: Proceedings of the 40th International Conference on Information Systems*, 1–17.
- Abdulrahman, A., Richards, D. & Bilgin, A. A. (2021). Reason explanation for encouraging behaviour change intention. *AAMAS '21: Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems*, 68–77.
- Achterkamp, R., Cabrita, M., op den Akker, H., Hermens, H. J. & Vollenbroek-Hutten, M. M. R. (2013). Promoting a healthy lifestyle: Towards an improved personalized feedback approach. *HealthCom '13: Proceedings of the 15th IEEE International Conference on e-Health Networking, Applications and Services*, 725–727. <https://doi.org/10.1109/HealthCom.2013.6720772>
- Acosta, J. C. & Ward, N. G. (2011). Achieving rapport with turn-by-turn, user-responsive emotional coloring. *Speech Communication*, 53(9–10), 1137–1148. <https://doi.org/10.1016/j.specom.2010.11.006>
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50(2), 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L)
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R. & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>



- Beinema, T., Davison, D., Reidsma, D., Banos, O., Bruijnes, M., Donval, B., Fides Valero, Á., Heylen, D., Hofs, D., Huizing, G., Kantharaju, R. B., Klaassen, R., Kolkmeier, J., Konsolakis, K., Pease, A., Pelachaud, C., Simonetti, D., Snaith, M., Traver, V., ... op den Akker, H. (2021). Agents United: An open platform for multi-agent conversational systems. *IVA '21: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*.
- Bickmore, T. W. & Giorgino, T. (2006). Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, 39(5), 556–571. <https://doi.org/10.1016/j.jbi.2005.12.004>
- Bickmore, T. W., Gruber, A. & Picard, R. W. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 59(1), 21–30. <https://doi.org/10.1016/j.pec.2004.09.008>
- Bickmore, T. W., Pfeifer, L. M., Byron, D., Forsythe, S., Henault, L. E., Jack, B. W., Silliman, R. A. & Paasche-Orlow, M. K. (2010). Usability of conversational agents by patients with inadequate health literacy: Evidence from two clinical trials. *Journal of Health Communication*, 15(S2), 197–210. <https://doi.org/10.1080/10810730.2010.499991>
- Bickmore, T. W., Trinh, H., Asadi, R. & Olafsson, S. (2018). Safety first: Conversational agents for health care. *Studies in Conversational UX Design*, 33–57. https://doi.org/10.1007/978-3-319-95579-7_3
- Bickmore, T. W., Trinh, H., Olafsson, S., O'Leary, T. K., Asadi, R., Rickles, N. M. & Cruz, R. (2018). Patient and consumer safety risks when using conversational assistants for medical information: An observational study of Siri, Alexa, and Google Assistant. *Journal of Medical Internet Research*, 20(9), 1–13. <https://doi.org/10.2196/11510>
- Callejas, Z. & Griol, D. (2021). Conversational agents for mental health and wellbeing. *Dialog systems: A perspective from language, logic and computation* (pp. 219–244). https://doi.org/10.1007/978-3-030-61438-6_11
- Carey, R. N., Connell, L. E., Johnston, M., Rothman, A. J., de Bruin, M., Kelly, M. P. & Michie, S. (2018). Behavior change techniques and their mechanisms of action: A synthesis of links described in published intervention literature. *Annals of Behavioral Medicine*, 53(8), 693–707. <https://doi.org/10.1093/abm/kay078>
- Cole-Lewis, H., Ezeanochie, N. & Turgiss, J. (2019). Understanding health behavior technology engagement: Pathway to measuring digital behavior change interventions. *JMIR Formative Research*, 3(4), 1–10. <https://doi.org/10.2196/14052>
- Creed, C., Beale, R. & Cowan, B. (2015). The impact of an embodied agent's emotional expressions over multiple interactions. *Interacting with Computers*, 27(2), 172–188. <https://doi.org/10.1093/iwc/iwt064>
- Devillers, L. (2020). Human-robot interactions and affecting computing: The ethical implications. *Dagstuhl Reports*, 10(1), 19–24.
- de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J. & Evers, V. (2017). A word of advice: How to tailor motivational text messages based on behavior change theory to personality and gender. *Personal and Ubiquitous Computing*, 21(4), 675–687. <https://doi.org/10.1007/s00779-017-1025-1>
- Diederich, S., Brendel, A., Morana, S. & Kolbe, L. (2021). On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*.
- Dingler, T., Kwasnicka, D., Wei, J., Gong, E. & Oldenburg, B. (2021). The use and promise of conversational agents in digital health. *IMIA Yearbook of Medical Informatics 2021*, 32–40.
- El Kamali, M., Angelini, L., Caon, M., Carrino, F., Rocke, C., Guye, S., Rizzo, G., Mastropietro, A., Sykora, M., Elayan, S., Kniestedt, I., Ziylan, C., Lettieri, E., Khaled, O. A. & Mugellini, E. (2020). Virtual coaches for older adults' wellbeing: A systematic review. *IEEE Access*, 8, 101884–101902. <https://doi.org/10.1109/ACCESS.2020.2996404>

REFERENCES

- European Commission. (2021). *Proposal for a regulation of the european parliament and of the council laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts* (tech. rep.). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- Fan, H. & Poole, M. S. (2006). What is personalization? perspectives on the design and implementation of personalization in information systems. *Journal of Organizational Computing and Electronic Commerce*, 16(3-4), 179–202. https://doi.org/10.1207/s15327744jocce1603_4_2
- Feine, J., Gnewuch, U., Morana, S. & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human Computer Studies*, 132, 138–161. <https://doi.org/10.1016/j.ijhcs.2019.07.009>
- Fitrianie, S., Bruijnes, M., Li, F. & Brinkman, W.-P. (2021). Questionnaire items for evaluating artificial social agents - expert generated, content validated and reliability analysed. *IVA '21: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, 84–86. <https://doi.org/10.1145/3472306.3478341>
- Fitrianie, S., Bruijnes, M., Richards, D., Abdulrahman, A. & Brinkman, W.-P. (2019). What are we measuring anyway? a literature survey of questionnaires used in studies reported in the intelligent virtual agent conferences. *IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 159–161.
- Fitrianie, S., Bruijnes, M., Richards, D., Bönsch, A. & Brinkman, W.-P. (2020). The 19 unifying questionnaire constructs of artificial social agents: An IVA community analysis. *IVA '20: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, (7). <https://doi.org/10.1145/3383652.3423873>
- Gambino, A., Fox, J. & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, 1, 71–86. <https://doi.org/10.30658/hmc.1.5>
- Godolphin, W. (2009). Shared decision making. *Healthcare Quarterly*, 12, 64–67. <https://doi.org/10.1055/s-0039-1677417>
- Gratch, J., Ning, W., Gerten, J., Fast, E. & Duffy, R. (2007). Creating rapport with virtual agents. *IVA '07: Proceedings of the 7th ACM International Workshop on Intelligent Virtual Agents*, 4722 LNCS, 125–138. https://doi.org/10.1007/978-3-540-74997-4_12
- Griffin, A. C., Xing, Z., Khairat, S., Wang, Y., Bailey, S., Arguello, J. & Chung, A. E. (2020). Conversational agents for chronic disease self-management: A systematic review. *AMIA Annual Symposium Proceedings*, 504–513.
- Hardeman, W., Houghton, J., Lane, K., Jones, A. & Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAs) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1). <https://doi.org/10.1186/s12966-019-0792-7>
- Iacobucci, G. (2020). Row over Babylon's chatbot shows lack of regulation. *BMJ*, 368(m815), 1–2. <https://doi.org/10.1136/bmj.m815>
- Joseph-Williams, N., Elwyn, G. & Edwards, A. (2014). Knowledge is not power for patients: A systematic review and thematic synthesis of patient-reported barriers and facilitators to shared decision making. *Patient Education and Counseling*, 94(3), 291–309. <https://doi.org/10.1016/j.pec.2013.10.031>
- Kamphorst, B. A. (2017). E-coaching systems: What they are, and what they aren't. *Personal and Ubiquitous Computing*, 21(4), 625–632. <https://doi.org/10.1007/s00779-017-1020-6>
- Kamphorst, B. A. (2020). *Autonomy-respectful e-coaching systems: Fending off complacency* (Doctoral dissertation). Utrecht University.
- Kelders, S. M., Kip, H. & Greeff, J. (2020). Psychometric evaluation of the twente engagement with Health technologies scale (TWEETS): Evaluation study. *Journal of Medical Internet Research*, 22(10), e17757. <https://doi.org/10.2196/17757>

- Kirakowski, J., O'Donnell, P. & Yiu, A. (2007). The perception of artificial intelligence as "human" by computer users. *HCI '07: Proceedings of the International Conference on Human-Computer Interaction*, 4552 LNCS, 376–384. https://doi.org/10.1007/978-3-540-73110-8_40
- Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A. & Murphy, S. A. (2015). Micro-randomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology*, 34, 1220–1228. <https://doi.org/10.1016/j.physbeh.2017.03.040>
- Klein, M. C. A., Mogles, N. & van Wissen, A. (2013). An intelligent coaching system for therapy adherence. *IEEE Pervasive Computing*, 12(3), 22–30. <https://doi.org/10.1109/MPRV.2013.41>
- Kocaballi, A. B., Berkovsky, S., Quiroz, J. C., Laranjo, L., Tong, H. L., Rezazadegan, D., Briatore, A. & Coiera, E. (2019). The personalization of conversational agents in health care: Systematic review. *Journal of Medical Internet Research*, 21(11), e15360. <https://doi.org/10.2196/15360>
- Komatsu, T., Kurosawa, R. & Yamada, S. (2012). How does the difference between users' expectations and perceptions about a robotic agent affect their behavior? *International Journal of Social Robotics*, 4(2), 109–116. <https://doi.org/10.1007/s12369-011-0122-y>
- Kramer, L. L., ter Stal, S., Mulder, B. C., de Vet, E. & van Velsen, L. (2020). Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *Journal of medical Internet Research*, 22(2), e14058. <https://doi.org/10.2196/14058>
- Krebs, P., Prochaska, J. O. & Rossi, J. S. (2010). A meta-analysis of computer-tailored interventions for health behavior change. *Preventive Medicine*, 51(3-4), 214–221. <https://doi.org/10.1016/j.ypmed.2010.06.004>
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J. Y. C., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S. & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, 25(9), 1248–1258. <https://doi.org/10.1093/jamia/ocy072>
- Lee, M., Noortman, R., Zaga, C., Starke, A., Huisman, G. & Andersen, K. (2021). Conversational futures: Emancipating conversational interactions for futures worth wanting. *CHI '21: Proceedings of the 2021 ACM Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3411764.3445244>
- Lefebvre, R. C., Tada, Y., Hilfiker, S. W. & Baur, C. (2010). The assessment of user engagement with eHealth content: The eHealth engagement scale. *Journal of Computer-Mediated Communication*, 15(4), 666–681. <https://doi.org/10.1111/j.1083-6101.2009.01514.x>
- Lentferink, A. J., Oldenhuis, H. K. E., de Groot, M., Polstra, L., Velthuijsen, H. & van Gemert-Pijnen, J. E. W. C. (2017). Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: A scoping review. *Journal of Medical Internet Research*, 19(8), e277. <https://doi.org/10.2196/jmir.7288>
- Locke, E. A. & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717. <https://doi.org/10.1037/0003-066X.57.9.705>
- Loveys, K., Sebaratnam, G., Sagar, M. & Broadbent, E. (2020). The effect of design features on relationship quality with embodied conversational agents: A systematic review. *International Journal of Social Robotics*, 12, 1293–1312. <https://doi.org/10.1007/s12369-020-00680-7>
- Mankins, J. C. (2009). Technology readiness and risk assessments: A new approach. *Acta Astronautica*, 65(9–10), 1208–1215. <https://doi.org/10.1016/j.actaastro.2009.03.059>

REFERENCES

- Mensio, M., Rizzo, G. & Morisio, M. (2018). The rise of emotion-aware conversational agents: Threats in digital emotions. *WWW '18: Proceedings of The Web Conference 2018*, 1541–1544. <https://doi.org/10.1145/3184558.3191607>
- Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J. & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, 46(1), 81–95. <https://doi.org/10.1007/s12160-013-9486-6>
- Michie, S., van Stralen, M. M. & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1), 42. <https://doi.org/10.1186/1748-5908-6-42>
- Milne-Ives, M., de Cock, C., Lim, E., Shehadeh, M. H., de Pennington, N., Mole, G., Normando, E. & Meinert, E. (2020). The effectiveness of artificial intelligence conversational agents in health care: Systematic review. *Journal of Medical Internet Research*, 22(10), e20346. <https://doi.org/10.2196/20346>
- Miner, A. S., Milstein, A., Schueler, S., Hegde, R., Mangurian, C. & Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine*, 176(5), 619–625. <https://doi.org/10.1001/jamainternmed.2016.0400>
- Montenegro, J. L. Z., da Costa, C. A. & da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Systems with Applications*, 129, 56–67. <https://doi.org/10.1016/j.eswa.2019.03.054>
- Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A. & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52(6), 446–462. <https://doi.org/10.1007/s12160-016-9830-8>
- Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Nass, C., Steuer, J. & Tauber, E. R. (1994). Computers are social actors. *CHI '94: Proceedings of the 1994 ACM Conference on Human Factors in Computing Systems*, 72–78. <https://doi.org/10.1145/259963.260288>
- Nijland, N. (2011). *Grounding eHealth - towards a holistic framework for sustainable eHealth technologies* (Doctoral dissertation). University of Twente. <https://doi.org/10.3990/1.9789036531337>
- Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C. & Peters, C. (2020). Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, 7, 1–21. <https://doi.org/10.3389/frobt.2020.00092>
- Oinas-Kukkonen, H. & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, 24(1), 485–500.
- op den Akker, H., Jones, V. M. & Hermens, H. J. (2014). Tailoring real-time physical activity coaching systems: A literature survey and model. *User Modeling and User-Adapted Interaction*, 24(5), 351–392. <https://doi.org/10.1007/s11257-014-9146-y>
- Paramythios, A., Weibelzahl, S. & Masthoff, J. (2010). Layered evaluation of interactive adaptive systems: Framework and formative methods. *User Modeling and User-Adapted Interaction*, 20(5), 383–453. <https://doi.org/10.1007/s11257-010-9082-4>
- Perski, O., Blandford, A., Garnett, C., Crane, D. H., West, R. & Michie, S. (2018). A self-report measure of engagement with digital behavior change interventions (DBCIs): Development and psychometric evaluation of the "DBCI engagement scale". *Translational Behavioral Medicine*, 10(1), 267–277. <https://doi.org/10.1093/tbm/ibz039>
- Perski, O., Blandford, A., West, R. & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: A systematic review using principles from



- critical interpretive synthesis. *Translational Behavioral Medicine*, 7(2), 254–267. <https://doi.org/10.1007/s13142-016-0453-1>
- Prochaska, J. O. & Velicer, W. F. (1997). The transtheoretical model of health behaviour change. *American Journal of Health Promotion*, 12(1), 38–48. <https://doi.org/10.4278/0890-1171-12.1.38>
- Rheu, M., Shin, J. Y., Peng, W. & Huh-Yoo, J. (2020). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human-Computer Interaction*, 37(1), 81–96. <https://doi.org/10.1080/10447318.2020.1807710>
- Richards, D. & Bransky, K. (2014). ForgetMeNot: What and how users expect intelligent virtual agents to recall and forget personal conversational content. *International Journal of Human-Computer Studies*, 72(5), 460–476. <https://doi.org/10.1016/j.ijhcs.2014.01.005>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Schembре, S. M., Liao, Y., Robertson, M. C., Dunton, G. F., Kerr, J., Haffey, M. E., Burnett, T., Basen-Engquist, K. & Hicklen, R. S. (2018). Just-in-time feedback in diet and physical activity interventions: Systematic review and practical design framework. *Journal of Medical Internet Research*, 20(3), 1–14. <https://doi.org/10.2196/jmir.8701>
- Scholten, M. R., Kelders, S. M. & van Gemert-Pijnen, J. E. W. C. (2017). Self-guided web-based interventions: Scoping review on user needs and the potential of embodied conversational agents to address them. *Journal of Medical Internet Research*, 19(11), e383. <https://doi.org/10.2196/jmir.7351>
- Schwarzer, R., Lippke, S. & Luszczynska, A. (2011). Mechanisms of health behavior change in persons with chronic illness or disability: The health action process approach (HAPA). *Rehabilitation Psychology*, 56(3), 161–170. <https://doi.org/10.1037/a0024509>
- Sheh, R. K. (2017). Different XAI for different HRI. *AAAI Fall Symposium Series: Technical Report*, 114–117.
- Short, C. E., DeSmet, A., Woods, C., Williams, S. L., Maher, C., Middelweerd, A., Müller, A. M., Wark, P. A., Vandelanotte, C., Poppe, L., Hingle, M. D. & Crutzen, R. (2018). Measuring engagement in eHealth and mHealth behavior change interventions: Viewpoint of methodologies. *Journal of Medical Internet Research*, 20(11), e9397. <https://doi.org/10.2196/jmir.9397>
- Snaith, M., Conway, N., Beinema, T., De Franco, D., Pease, A., Kantharaju, R. B., Janier, M., Huizing, G., Pelachaud, C. & op den Akker, H. (2021). A multimodal corpus of simulated consultations between a patient and multiple healthcare professionals. *Language Resources and Evaluation*, 1–16. <https://doi.org/10.1007/s10579-020-09526-0>
- Snaith, M., De Franco, D., Beinema, T., op den Akker, H. & Pease, A. (2018). A dialogue game for multi-party goal-setting in health coaching. *COMMA '18: Proceedings of the 7th International Conference on Computational Models of Argument*, 337–344. <https://doi.org/10.3233/978-1-61499-906-5-337>
- Snaith, M., Lawrence, J., Pease, A. & Reed, C. A. (2020). A modular platform for argument and dialogue. *COMMA '20: Proceedings of the 9th International Conference on Computational Models of Argument*, 473–474. <https://doi.org/10.3233/FAIA200540>
- Snaith, M., Nielsen, R. Ø., Kotnis, S. R. & Pease, A. (2021). Ethical challenges in argumentation and dialogue in a healthcare context. *Argument and Computation*, 12(2), 249–264. <https://doi.org/10.3233/AAC-200908>
- Stigglebout, A. M., van der Weijden, T., de Wit, M. P., Frosch, D., Légaré, F., Montori, V. M., Trevena, L. & Elwyn, G. (2012). Shared decision making: Really putting patients at the centre of healthcare. *BMJ*, 344, e256. <https://doi.org/10.1136/bmj.e256>

REFERENCES

- Tedesco, D. P. & Tullis, T. S. (2006). A comparison of methods for eliciting post-task subjective ratings in usability testing. *Usability Professionals Association (UPA)*, 1–9.
- ter Stal, S., Tabak, M., op den Akker, H., Beinema, T. & Hermens, H. J. (2019). Who do you prefer? the effect of age, gender and role on users' first impressions of embodied conversational agents in eHealth. *International Journal of Human-Computer Interaction*, 36(9), 881–892. <https://doi.org/10.1080/10447318.2019.1699744>
- The European Parliament and Council. (2017). *Regulation on medical devices* (tech. rep.). <http://data.europa.eu/eli/reg/2017/745/2020-04-24>
- The European Parliament and Council. (2018). *General Data Protection Regulation* (tech. rep.). <https://gdpr-info.eu/>
- Trinh, H., Shamekhi, A., Kimani, E. & Bickmore, T. W. (2018). Predicting user engagement in longitudinal interventions with virtual agents. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 9–16. <https://doi.org/10.1145/3267851.3267909>
- Tudor Car, L., Dhinagaran, D. A., Kyaw, B. M., Kowatsch, T., Joty, S., Theng, Y. L. & Atun, R. (2020). Conversational agents in health care: Scoping review and conceptual analysis. *Journal of Medical Internet Research*, 22(8), e17158. <https://doi.org/10.2196/17158>
- van Velsen, L., Broekhuis, M., Jansen-Kosterink, S. M. & op den Akker, H. (2019). Tailoring persuasive eHealth strategies for older adults on the basis of personal motivation: An online survey. *Journal of Medical Internet Research*, 21(9), e11759. <https://doi.org/10.2196/11759>
- World Health Organization. (2011a). *Global recommendations on physical activity for health; 18-64 years old* (tech. rep.). <https://doi.org/10.1080/11026480410034349>
- World Health Organization. (2011b). *Global recommendations on physical activity for health; 65 years and above* (tech. rep.). World Health Organization. <https://www.who.int/dietphysicalactivity/publications/recommendations65yearsold/en>
- Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., Merchant, G. C., Naughton, F. & Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine*, 51(5), 833–842. <https://doi.org/10.1016/j.amepre.2016.06.015>
- Zhang, Z. & Bickmore, T. W. (2018). Medical shared decision making with a virtual agent. *IVA '18: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents*, 113–118. <https://doi.org/10.1145/3267851.3267883>



Summary

While a healthy lifestyle is important for our well-being and can prevent illnesses, changing our behaviour to adopt a healthier lifestyle can be difficult. Digital health (eHealth) applications can support people in this process. However, it has been found that the use of these applications decreases rapidly when the novelty effect wears off. Two causes for this are the lack of human involvement and the lack of personalised content. Therefore, conversational agents are added to eHealth applications in the role of health coaches to provide a social incentive, and adjustment of their dialogue content to users is investigated to increase personal relevance (tailoring). Such tailoring has previously been found to be effective in eHealth applications, but has not extensively been investigated for interactive two-way communication over a longer period of time, such as in coaching conversations. The objective of this thesis research was therefore to address the following question: **“How can we tailor users’ coaching conversations with conversational agents to improve engagement?”**

To provide users with tailored coaching conversations, dialogue content is needed that can be adapted. In addition, it is also important that such dialogue content is in line with health coaching guidelines and that the advice the user receives is safe. Thus, it is beneficial if dialogues can also be written and tested by (non-technical) domain experts. **Chapter 2**, therefore, describes the development of a set of tools that support this process: the WOOL Dialogue Platform. These tools allow domain experts to author and test dialogues, while they also include dialogue execution libraries to support agent developers. The chapter describes the developed tools, dialogue language, and use cases from projects in which the WOOL Dialogue Platform is used.

Following the development of dialogue tools, **Chapter 3** describes a study that investigates if the long-term strategies that are underlying to a coaching dialogue can be tailored. In the study, 108 participants interacted with a group of conversational agents, during which they answered a questionnaire on motivation to live healthy, were classified into motivation groups and were presented by the agents with two strategies: a positively and a negatively tailored strategy for their motivation group. Participants were asked to rate these strategies and to choose their favourite. In addition, they were asked to rate the agents’ likeability at two moments during the experiment. Results showed that it is indeed possible to tailor coaching strategies to users’ motivation to live healthy in the context of coaching dialogues, and they confirm that users can be both intrinsically motivated and externally regulated. Furthermore, the agents’ likeability did not seem to affect users’ ratings for the strategies, but rather the discussed dialogue content seemed to affect agents’ likeability scores.

Where **Chapter 3** investigated tailoring of the underlying strategies, the last

two chapters in this thesis focus on dynamically tailoring coaching dialogues during interaction. To that end, **Chapter 4** discriminates five levels in the process of tailoring conversations, namely the *domain*, *topic*, *action*, *dialogue act* and *utterance* level. The chapter then focuses on the definition of a blueprint topic model to support automatic tailoring on the topic level. The model was constructed on the basis of literature on behavioural psychology, persuasive technology and relational agents. It includes 30 hierarchically structured topics that cover 115 actions that a conversational agent could perform as a coach. The blueprint topic model was validated in a closed card-sort study with 11 experts. Results showed that most of the defined topics were confirmed to cover their core actions. Cards that were sorted incorrectly were often missing an immediacy indicator in their description or were based on actions for which it is more complex to imagine a corresponding dialogue.

Following the blueprint topic model, in **Chapter 5** a proof-of-concept implementation of a component that performs tailored selection of conversational topics is evaluated. The chapter describes the setup and results of a micro-randomized trial that compares two conditions. In the first condition, the conversational agent suggested a specific topic to discuss (coach initiative), which is an application of the topic selection component. In the second condition, the user was shown a menu dialogue from which they could select what they wanted to discuss (user initiative). The condition was randomly selected each time a participant interacted with the conversational agent. In the study, 82 participants interacted with the application over 4-8 weeks, which resulted in 1,011 interactions. Results showed that the coach suggesting a specific (tailored) topic was perceived as a natural interaction and that acceptance of these suggestions by participants was high (71.1% overall, 75.8% for coaching topics). Participants did care about the type of topic that was suggested, as supported by the lower acceptance rate for background story suggestions in the second round. Furthermore, there were indications that a coach's initiative might need to be tailored to users' active involvement in their behaviour change process.

Finally, **Chapter 6** discusses the main findings and conclusions of this thesis and does so following three themes, namely: 1) content and system design, 2) tailoring content and conversations, and 3) evaluation of tailored content. In short, it concludes that content should be carefully constructed and reported, that tailoring approaches should be combined, and that evaluation of tailored applications remains complicated due to the many options for tailoring and factors of influence in long-term daily life situations. The second part of the chapter then discusses considerations for future work on tailoring and conversational agents in eHealth and does so for five themes: safety, verification and validation; autonomy and trust; social interaction; input and output modalities; and open science.



Samenvatting

Een gezonde levensstijl is belangrijk voor ons welzijn en kan ziektes voorkomen. Het kan echter moeilijk zijn om ons gedrag te veranderen. Digitale gezondheidsapplicaties (eHealth-applicaties) kunnen mensen die een gezonde(re) levensstijl willen ontwikkelen ondersteunen bij aanpassing van hun gedrag, maar de praktijk laat zien dat het gebruik van dergelijke applicaties al na de eerste keer snel afneemt. Twee oorzaken hiervoor zijn het gebrek aan menselijke betrokkenheid en het ontbreken van gepersonaliseerde inhoud. Om het gebruik van applicaties te verbeteren worden er conversational agents als gezondheidscoaches aan toegevoegd en wordt het aanpassen van hun dialooginhoud aan gebruikers onderzocht om de persoonlijke relevantie te vergroten (tailoring). Dergelijke tailoring is eerder effectief gebleken in eHealth-applicaties, maar is niet uitgebreid onderzocht voor interactieve tweerichtingscommunicatie over een langere periode, zoals in coachinggesprekken. Het onderzoek in dit proefschrift was er daarom op gericht om de volgende vraag te beantwoorden: **“Hoe kunnen we coachinggesprekken met conversational agents aanpassen aan gebruikers om hun betrokkenheid bij de applicatie te vergroten?”**

Om coachinggesprekken op maat te bieden is aanpasbare dialooginhoud nodig. Daarnaast is het ook belangrijk dat de dialooginhoud richtlijnen voor gezondheidscoaching volgt en dat het advies aan de gebruiker verantwoord is. Het heeft van ook belangrijke voordelen als dialogen ook door (niet-technische) domeinexperts kunnen worden geschreven en getest. **Hoofdstuk 2** beschrijft de ontwikkeling van software om dit proces ondersteunen: het WOOL Dialogue Platform. Met de WOOL-editor kunnen domeinexperts dialogen schrijven en testen. WOOL ondersteunt ook agent-ontwikkelaars doordat het ook software voor het uitvoeren van dialogen omvat. Het hoofdstuk beschrijft de dialoogtaal, de ontwikkelde software, en voorbeelden uit projecten waarin het dialoog platform wordt ingezet.

Hoofdstuk 3 beschrijft een studie die onderzocht of de langetermijnstrategieën die ten grondslag liggen aan een coachingdialoog aangepast kunnen worden. In het onderzoek hadden 108 deelnemers een gesprek met een groep conversational agents. Ze beantwoordden vragen over hun motivatie om gezond te willen leven, waarna ze werden ingedeeld in motivatiegroepen. Vervolgens kregen ze op basis daarvan twee strategieën gepresenteerd door de agents: een positieve en een negatieve strategie voor hun motivatiegroep. De deelnemers werd gevraagd om deze strategieën te beoordelen en hun favoriet te kiezen. Daarnaast werd hen op twee momenten gevraagd om de sympathie van de agents te beoordelen. De resultaten toonden aan dat het mogelijk is om coachingstrategieën af te stemmen op motivatie om gezond te leven en ze bevestigen dat gebruikers ook twee typen motivatie kunnen hebben. Daarbij leek de sympathiescore van de agents geen invloed te hebben op de gebruikersbeoordelingen voor de strategieën, terwijl de

besproken dialooginhoud juist wel de sympathiescores leek te beïnvloeden.

Waar **Hoofdstuk 3** het afstemmen van de onderliggende strategieën onderzocht, richten de laatste twee hoofdstukken van dit proefschrift zich op het dynamisch afstemmen van coachingdialogen tijdens interactie. **Hoofdstuk 4** begint met vijf niveaus waarin de conversatie kan worden aangepast: *domein, onderwerp, actie, dialoogactie en uiting*. Het hoofdstuk richt zich vervolgens op de definitie van een blauwdruk-onderwerpmmodel om automatische tailoring op onderwerpniveau te ondersteunen. Het model is ontwikkeld op basis van literatuur over gedragspsychologie, persuasieve technologie en relational agents. Het bevat 30 hiërarchisch gestructureerde onderwerpen die 115 acties omvatten die een agent als coach zou kunnen uitvoeren. Het blauwdruk-onderwerpmmodel is vervolgens gevalideerd in een gesloten kaartsorteerstudie met 11 experts. De resultaten toonden aan dat het merendeel van de gedefinieerde onderwerpen hun kernacties goed afdekt.

Na het blauwdruk-onderwerpmmodel wordt in **Hoofdstuk 5** een proof-of-concept-implementatie geëvalueerd van een software-component die gebruikersspecifieke gespreksonderwerpen selecteert. Het hoofdstuk beschrijft een micro-randomized trial die twee condities vergelijkt. In de eerste conditie suggereert de conversational agent een specifiek onderwerp om te bespreken (coachinitiatief), wat een toepassing is van de onderwerpselectiecomponent. In de tweede conditie kiest de gebruiker zelf een onderwerp (gebruikersinitiatief). De conditie werd willekeurig gekozen elke keer als een deelnemer met de agent interacteerde. In twee rondes van 4-8 weken gebruikten 82 deelnemers de applicatie, wat resulteerde in 1.011 interacties. De resultaten toonden aan dat het coachinitiatief werd gezien als een natuurlijke interactie en dat de acceptatie van zulke suggesties door de deelnemers hoog was (71,1% in het algemeen, 75,8% voor coachingonderwerpen). Verder waren er aanwijzingen dat het gesuggerde onderwerp uitmaakt en dat het initiatief van een coach wellicht moet worden afgestemd op de actieve betrokkenheid van gebruikers bij hun gedragsveranderingsproces.

Hoofdstuk 6 vat de belangrijkste bevindingen en conclusies van dit proefschrift samen aan de hand van drie thema's, namelijk: 1) inhoud en systeemontwerp; 2) het aanpassen van inhoud en gesprekken aan gebruikers; en 3) evaluatie van aangepaste interacties. In het kort komt het erop neer dat de inhoud zorgvuldig moet worden geconstrueerd en gerapporteerd, verschillende aanpakken voor het aanpassen van conversaties moeten worden gecombineerd en de evaluatie van applicaties die zich aanpassen aan gebruikers gecompliceerd blijft vanwege de talrijke opties voor aanpassing en de vele factoren die van invloed zijn in dagelijkse situaties en op de lange termijn. Het tweede deel van het hoofdstuk bespreekt vervolgens overwegingen voor toekomstig onderzoek op het gebied van tailoring en conversational agents in eHealth-applicaties en doet dit voor vijf thema's: veiligheid en verifieerbaarheid, autonomie en vertrouwen, sociale interactie, input- en outputmodaliteiten en open wetenschap.



Dankwoord

Zo! Na drie jaar bij RRD en een jaar bij de UT is mijn proefschrift nu dan toch zo goed als af. Vier jaar waarin ik kennis heb mogen maken met de Twentse taal en cultuur, verschillende steden en conferenties heb mogen bezoeken, en ook zeker heel veel enthousiaste, leuke en lieve mensen heb mogen leren kennen. Ik wil daarom iedereen ook enorm bedanken!

Ten eerste, mijn promotor en co-promotor, Hermie en Harm. Hermie, bedankt voor de alle adviezen, gesprekken en het aan de slag mogen bij het PeHT programma. Hoewel de PhD-overleggen soms een omweg konden nemen via andere gespreksonderwerpen – zo kan de overeenkomst tussen het schrijven van text-based games en coaching dialogen toch ineens voor een overleg van ruim 1.5 uur zorgen – gaven ze ook altijd nieuwe inzichten voor het werk in dit proefschrift.

Harm, bedankt voor je enthousiasme, optimisme, en het vertrouwen om aan jouw Council of Coaches te mogen werken. Ik kan me nog goed herinneren dat Silke tijdens één van onze eerste overleggen hard moest lachen om mijn verwarde blik toen er 'heanig an' uit mijn laptop kwam. Hoewel dat soms nog steeds wat oefening nodig heeft, was dit proefschrift niet geweest wat het nu is zonder jouw bemoedigende woorden en goede adviezen als ik weer eens vast zat.

I also want to thank all the members of the graduation committee that were willing to accept the task of reviewing my thesis, providing suggestions, and being present at my PhD defence.

Silke en Roos, ik ben vereerd dat jullie mijn paranifmen willen zijn. Silke, onze tijd en kantoorruimte bij RRD overlatten precies. Of het nu ging om de directheid van e-mails aanpassen, eierkoeken en chocola delen tijdens het werken, verbazing omdat er Spaans uit jouw laptop kwam, geknoei tijdens de lunch of 'gewoon' een agent applicatie bouwen – wat ben ik blij dat dat samen kon! Bedankt!

Roos, het enthousiasme waarmee jij nieuwe dingen gewoon probeert en activiteiten op gang brengt is echt bewonderenswaardig! Het zal je misschien niet verbazen dat ik nog steeds het moment moet tegenkomen waarop ik hardlopen echt leuk ga vinden, maar met jou en Flower kwam het toch behoorlijk dichtbij. Dank voor alle gezelligheid!

I'm also thankful for all of the people that I had the pleasure to work with during the Council of Coaches project. I've lost count of the times that I have heard people say that it was a great project with great people and good collaboration. From walking around, paella for dinner, and tapas workshops in Valencia, whiskey tastings on a boat in Dundee, blob posters that tried to get us up 10 flights of stairs (you can also read them backwards on the way down!), technical integration meetings at the UT, being demonstrated the importance of intonation in uses of

the word ‘doei’, boat tours in Paris, being welcomed to Dundee on a work visit and learning about ‘half four at the circle’, to ‘quickly having dinner and going to bed early’ at AAMAS and instead ending up at dinner until midnight with a large group of people in one of the oldest restaurants in the world (with cloud berries for dessert). I’m really happy that I got to experience it with all of you!

Dan collega’s, met ten eerste natuurlijk de oud-collega’s van RRD, die altijd klaar stonden om een vraag te beantwoorden, advies te geven of koffiepauze te houden waarbij het altijd maar weer de vraag was welke onderwerpen er voorbij zouden komen. Lex, die als ‘oom met goed advies’ toch ook bij een flink deel van de hoofdstukken in dit proefschrift betrokken is geweest. Door onze werkreis naar Coimbra voor SmartWork kan ik ‘trein was vol’ en vliegen zonder ticket toch maar weer mooi van de reisbingokaart afstrepken. Dennis, bedankt voor al je geduld als ik weer eens in je kantoor stond met een ‘het is stuk’ (waarbij het dan altijd nog maar weer de vraag was of het niet toch een typefout was). Marijke, bedankt voor de fijne samenwerking, koppen thee, gezelligheid en de terugkeerde verbazing rondom technologie die ineens werk weigerde. Marian en Stephanie voor alle hulp bij de COUCH evaluaties. Junioren in het algemeen, bedankt! Fijne collega’s hebben is één, maar buzzwordbingo, weekendjes weg, rolstoelbasketbal, spelletjesavonden en borrels met heksenkaas, zijn toch echt een hele grote bonus.

Dan natuurlijk ook nieuwe collega’s. Wisselen van baan tijdens een lockdown terwijl in eerste instantie de mensen waarmee je in je thuiskantoor overlegt hetzelfde bleven was een vreemde ervaring. Ik weet nog steeds niet exact wat er allemaal onder PeHT valt, maar wat kreeg ik er leuke collega’s bij! Ook mensen van BSS en in het bijzonder ook de eHealth people, dank dat ik erbij mocht!

Naast collega’s zijn er natuurlijk ook familie en vrienden die ik wil bedanken. Nijmegenaren, we hebben elkaar, zeker met Corona veel te weinig gezien, maar dat het klopt als we elkaar zien zegt genoeg. Ook inmiddels oud-bewoners van Barslecht en andere Enschedeërs, zonder jullie was de stap om naar Enschede te verhuizen een stuk groter geweest. En natuurlijk moet het ook in mijn dankwoord toch ook even over paarden gaan (voor de oplettende lezer waren we dat punt eigenlijk al voorbij). Of het nu om SOs, bier drinken na de les, of filmavonden ging, wat ben ik blij met alle lieve mensen bij Hippo met wie ik de afgelopen jaren het paardrijden, mijn donderdagavonden en alles daaromheen heb mogen delen.

Familie lijkt er soms vanzelfsprekend, maar wat is het fijn om mensen te hebben die er altijd voor je zijn. Mam, pap, Esther, Thomas en oma, merci!

Last, maar zeker not least, Bob. Niet alleen om het geweldige geduld dat jij wel met Illustrator hebt, maar voor alles <3



About the author

Tessa Beinema was born in Malden, the Netherlands, in June 1991. In 2010, she started the Artificial Intelligence bachelor program at the Faculty of Social Sciences, Radboud University, Nijmegen, the Netherlands. After her bachelors, Tessa started her masters in Artificial Intelligence with a specialisation in Web and Language Interaction.

Throughout her education she was interested in the interaction between intelligent systems and their users, and how such systems could be adjusted to their users. This interest was also reflected in her master thesis research at Dedicon, which focused on the development of a web search interface for the visually impaired.

After receiving her Masters degree in 2017, Tessa moved to Enschede to start as a junior researcher and PhD student at Roessingh Research and Development and the University of Twente. She worked in the European Council of Coaches project (Horizon 2020), where she led the work package on 'shared knowledge base and coaching strategies'. Her research presented in this thesis is situated in that project, but she also contributed to the several other projects, and the open source WOOL and Agents United platforms.

Once the Council of Coaches project had finished, in September 2020, Tessa joined the Personalised eHealth Technology programme of the University of Twente as a researcher, while also finishing her PhD. She coordinates the Agents United Alliance, is involved in the development of a telemonitoring platform and is a coordinator for the VSNU Digital Society Health and Well-being programme line.

Publications

International journal papers

Beinema, T., op den Akker, H., Hurmuz, M., Jansen-Kosterink, S., Hermens, H. Automatic topic selection for long-term interaction with embodied conversational agents in health coaching: A micro-randomized trial. *Manuscript submitted for publication.*

Beinema, T., op den Akker, H., Hofs, D., van Schooten, B. The WOOL Dialogue Platform: Enabling interdisciplinary user-friendly development of dialogue for conversational agents. *Manuscript submitted for publication.*

Beinema, T., op den Akker, H. Hermens, H., van Velsen, L. What to discuss? – A blueprint topic model for health coaching dialogues with conversational agents. *Manuscript submitted for publication.*

Hurmuz, M., Jansen-Kosterink, S., **Beinema, T.**, De Franco, D., op den Akker, H., Hermens, H. (2021) User experience, use, and potential health effect of a conversational agent-based electronic health intervention: An observational cohort Study. *Manuscript submitted for publication*, available as a preprint with *The Lancet*.

Beinema, T., op den Akker, H., van Velsen, L., Hermens, H. (2021). Tailoring coaching strategies to users' motivation in a multi-agent health coaching application. In *Computers in Human Behavior*, 121, 106787.

Snaith, M., Conway, N., **Beinema, T.**, De Franco, D., Pease, A., Kantharaju, R., Janier, M., Huizing, G., Pelachaud, C., op den Akker, H. (2021). A multimodal corpus of mock patient interviews. In *Language Resources and Evaluation*, 55, 1077–1092.

Ter Stal, S., Tabak, M., op den Akker, H., **Beinema, T.**, Hermens, H. (2019). Who do you prefer? The effect of age, gender and role on users' first impressions of embodied conversational agents in eHealth. In *International Journal of Human-Computer Interaction*, 36(9), 881–892.



International conference papers

Beinema, T.*., Davison, D.*., Reidsma, D.*., Banos, O., Bruijnes, M., Donval, B., Fides Valero, A., Heylen, D., Hofs, D., Huizing, G., Kantharaju, R.B., Klaassen, R., Kolkmeier, J., Konsolakis, K., Pease, A., Pelachaud, C., Simonetti, D., Snaith, M., Traver, V., van Loon, J., Visser, J., Weusthof, M., Yunus, F., Hermens, H.+, and op den Akker, H.+, (2021). Agents United: An open platform for multi-agent conversational systems. Published in *IVA '21: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*. Online, Japan.

* Ex aequo first authors. + Ex aequo last authors.

Katharaju, R. B., Pease, A., Reidsma, D., Pelachaud, C., Snaith, M., Bruijnes, M., Klaassen, R., **Beinema, T.**, Huizing, G., Simonetti, D., Heylen, D., op den Akker, H. (2019). Integrating argumentation with social conversation between multiple virtual coaches. Published in *IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*.

Das, K.S.J., **Beinema, T.**, Op den Akker, H., Hermens, H. (2019). Generation of multi-party dialogues among embodied conversational agents to promote active living and healthy diet for subjects suffering from type 2 diabetes. Published in *ICT4AWE '19: Proceedings of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 297–304. Kreta, Greece.

Snaith, M., De Franco, D., **Beinema, T.**, Op den Akker, H. and Pease, A. (2018, September 14). A dialogue game for multi-party goal-setting in health coaching. Published in *COMMA '18: Proceedings of the 7th International Conference on Computational Models of Argument*, 337–344. Warsaw, Poland.

Beinema, T., Op den Akker, H., Hermens, H. (2018). Creating an artificial coaching engine for multi-domain conversational coaches in eHealth applications. In *ICA-HoGeCa '18: Proceedings of the AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications co-located with the Federated AI Meeting (FAIM 2018)*. Stockholm, Sweden. André, E., Bickmore, T., Vrochidis, S., Wanner, L. (Eds.). CEUR-WS.org, online.

Op den Akker, H., op den Akker, R., **Beinema, T.**, Banos, O., Heylen, D., Bedsted, B., Pease, A., Pelachaud, C., Traver Salcedo, V., Kyriazakos, S., and Hermens, H. (2018). Council of Coaches: A novel holistic behaviour change coaching approach. In *ICT4AWE '18: Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health*. Funchal, Madeira, Portugal.

Conference contributions

Beinema, T., op den Akker, H., van Velsen, L., Hermens, H. (2021) What to discuss? Automatic topic selection for embodied conversational health coaches. Updated abstract and oral presentation at the *Supporting Health by Technology Conference (HealthByTech 2021)*. Online, the Netherlands.

Beinema, T., op den Akker, H., van Velsen, L., Hermens, H. (2020) What to discuss? Automatic topic selection for embodied conversational health coaches. Abstract in proceedings of the cancelled *Supporting Health by Technology Symposium (HealthByTech 2020)*. Enschede, the Netherlands.

Beinema, T., op den Akker, H., Hermens, H. (2018). Creating an artificial coaching engine for multi-domain conversational coaches in eHealth applications. Poster presented at the *AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications (ICA-HoGeCa2018)* held in conjunction with the Federated AI Meeting (FAIM). Stockholm, Sweden.

Op den Akker, H., **Beinema, T.**, ter Stal, S., Broekhuis, M. (2018). Designing virtual agents for persuasive health coaching. Workshop at *HealthByTech 2018: 8th Supporting Health By Technology Symposium*. Enschede, the Netherlands

Beinema, T. (2018). Creating an artificial coaching engine for multi-domain conversational coaches in eHealth applications. Presentation given at the *Roessingh Research and Development Symposium: Enabling Technology for Active Life and Better Health*. Enschede, the Netherlands.

Beinema, T., Dotsch, R., Grootjen, F.A. (2014) The detection of facial expressions for action coordination. Extended abstract and poster presented at *BNAIC 2014: The 26th Benelux Conference on Artificial Intelligence*. Nijmegen, the Netherlands.

Progress range

The following publications have been published in the Progress range by Roessingh Research and Development, Enschede, the Netherlands. Copies can be ordered, when available, via info@rrd.nl.

1. Pot, J.W.G.A, Boer H., van Harten, W.H., Hermens, H.J., Seydel, E.R. Comprehensive Need-Assessment. Ontwikkeling van een meetinstrument voor zorgbehoeften en kwaliteitsbeoordeling door patiënten. September 1994, ISBN 90-25452-01-22
2. Van Leerdam, N.G.A, Hermens, H.J. Revalidatietechnologie in Euregio. July 1995, ISBN 90-75452-02-0
3. Duda, L., van Noort, L.O., Röseler, S., Greitemann, B.O.L, van Harten, W.H., Klazinga, N.S. Rehabilitation in Germany and the Netherlands, A comparison of two rehabilitation systems. August 1995, ISBN 90-75452-03-9
4. Hermens, H.J., Nene, A.V., Zilvold, G. Electrophysiological Kinesiology, Proceedings of the 11th congress of the International Society of Electrophysiology and Kinesiology in Enschede, the Netherlands 1996. October 1996, ISBN 90-75452-04-7
5. Van Harten, W.H. Bouwen aan een kwaliteitssysteem in de revalidatiezorg. Een poging tot constructieve technology assessment van een kwaliteitssysteem in een gezondheidszorginstelling. December 1997, ISBN 90-75452-07-1
6. Baardman, G., IJzerman, M.J. Design and evaluation of a hybrid orthosis for people with paraplegia. November 1997, ISBN 90-75452-8-X
7. Hutten, M.M.R. Lumbar Dynamometry: A useful method for assessment of patients with chronic low back pain? November 1999, ISBN 90-75452-13-6
8. Van der Salm, A., van Harten W.H., Maathuis, C.G.B. Ketenkwaliteit Cerebrale Parese Zorg. Een beschrijving van de cerebrale parese zorg en mogelijke verbeteringen hierin. April 2001, ISBN 90-75453-19-5
9. Nederhand, M.J. Muscle activation patterns in post traumatic neck pain. March 2003, ISBN 90-75452-27-6
10. Jannink, M.J.A. Usability of custom-made orthopaedic shoes in patients with degenerative disorders of the foot. September 2004, ISBN 90-75452-28-4
11. Blokhorst, M.G.B.G. State-dependent factors and attention in whiplash associated disorder. January 2005, ISBN 90-365-2111-4
12. Buurke, J.H. Walking after stroke. Co-ordination patterns & functional recovery. February 2005, ISBN 90-365-140-8
13. Van der Salm, A. Spasticity reduction using electrical stimulation in the lower limb of spinal cord injury patients. October 2005, ISBN 90-365-2253-6
14. Snoek, G.J. Patient preferences for reconstructive interventions of the upper limb in tetraplegia. December 2005, ISBN 90-365-2255-2
15. De Kroon, J.R. Therapeutic electrical stimulation of the upper extremity in stroke. December 2005, ISBN 90-365-2269-2
16. Van Dijk, H. Motor skill learning, age and augmented feedback. March 2006, ISBN 90-365-202-9
17. Mes, C.A.J. Improving non-optimal results in chronic pain treatment. January 2007, ISBN 90-365-2435-0

-
18. Voerman, G.E. Musculoskeletal neck-shoulder pain: a new ambulant myofeedback intervention approach. March 2007, ISBN 90-365-2460-1
 19. Kallenberg, L.A.C. Multi-channel array EMG in chronic neck-shoulder pain. March 2007, ISBN 90-365-2459-8
 20. Huis in 't Veld, M.H.A. Work-releated neck-shoulder pain. The role of cognitive-behavioural factors and remotely supervised treatment. December 2007, ISBN 978-90-365-2584-8
 21. Fleuren, J.F.M. Assessment of spasticity. From EMG to patients' perception. October 2009, ISBN 978-90-365-2869-6
 22. Reenalda, J. Dynamic sitting to prevent pressure ulcers in spinal cord injured. October 2009, ISBN 978-90-365-2884-9
 23. Prange, G.B. Rehabilitation robotics. Stimulating restoration of arm function after stroke. October 2009, ISBN 978-90-365-2901-3
 24. Vos-van der Hulst, M. Prognostic factors and underlying mechanisms in chronic low back pain. February 2010, ISBN 978-90-365-2881-8
 25. Kottink-Hutten, A.I.R. Assessment of a two-channel implantable peroneal nerve stimulator post-stroke. February 2010, ISBN 978-90-365-2959-4
 26. Van Weering, M.G.H. Towards a new treatment for chronic low back pain patients. Using activity monitoring and personalized feedback. May 2011, ISBN 978-90-365-3305-8
 27. Gulmans, J. Crossing boundaries. Improving communication in cerebral palsy care. February 2012, ISBN 978-90-365-3305-8
 28. Molier, B.I. Influence of augmented feedback on learning upper extremity tasks after stroke. March 2012, ISBN 978-90-365-3293-9
 29. Dubbeldam, R. Towards a better understanding of foot and ankle kinematics in rheumatoid arthritis. The effects of walking speed and structural impairments. October 2012, ISBN 978-90-365-3407-9
 30. Evering, R.M.H. Ambulatory feedback at daily physical activity patterns. A treatment for the chronic fatigue syndrome in the home environment? April 2013, ISBN 978-90-365-3512-0
 31. Malhorta, S. Does spasticity interfere with functional recovery after stroke? A novel approach to understand, measure and treat spasticity after acute stroke. November 2013, ISBN 978-90-365-2567-0
 32. Tabak, M. Telemedicine for patients with COPD. New treatment approaches to improve daily activity behaviour. February 2014, ISBN 978-94-6108-590-0
 33. Trompetter, H.R. ACT with pain. Measurement, efficacy and mechanisms of Acceptance and Commitment Therapy. September 2014, ISBN 978-90-365-3708-7
 34. Op den Akker, H. Smart tailoring of real-time physical activity coaching systems. October 2014, ISBN 978-90-365-3762-9
 35. Jansen-Kosterink, S.M. The added value of telemedicine for physical rehabilitation. December 2014, ISBN 978-90-823196-0-6
 36. Velstra, I.M. Advanced insights in upper limb function of individuals with cervical spinal cord injury. December 2015, ISBN 978-90-365-3929-6
 37. Kloosterman, M.G.M. Keep on rolling. Functional evaluation of power-assisted wheelchair use. June 2016, ISBN 978-90-365-4299-9
 38. Prinsen, E.C. Adapting to change. Influence of a microprocessor-controlled prosthetic knee on gait adaptations. December 2016, ISBN 978-90-365-4206-7



39. Wolvers, M.D.J. A coach in your pocket. On chronic cancer-related fatigue and physical behavior. March 2017, ISBN 978-90-365-4299-9
40. Cabrita, M. Active and pleasant ageing supported by technology. November 2017, ISBN 978-90-365-4407-8
41. Haarman, J.A.M. TIBAR. Therapist Inspired Balance Assisting Robot. November 2017, ISBN 978-90-365-4407-8
42. Nijenhuis, S.M. Roll up your sleeves! Technology-supported arm and hand training at home after stroke. April 2018, ISBN 978-90-365-4510-5
43. Cranen, K. Acceptance of telerehabilitation in chronic pain: the patients' perspective. June 2018, ISBN 978-90-365-4555-6
44. Boerema, S.T. Sensing human activity to improve sedentary lifestyle. September 2018, ISBN 978-90-365-4604-1
45. Radder, B. The wearable hand robot. Supporting impaired hand function in activities of daily living and rehabilitation. November 2018, ISBN 978-90-365-4658-4
46. Krabben T. A reaching hand. Towards an active therapeutic device for the upper extremity following stroke. December 2018, ISBN 978-90-365-4660-7
47. Timmerman, J.G. Cancer rehabilitation at home. The potential of telehealthcare to support functional recovery of lung cancer survivors. January 2019, ISBN 978-90-4701-7
48. Nikamp-Simons, C.D.M. The sooner the better?! Providing ankle-foot orthoses in the rehabilitation after stroke. May 2019, ISBN 978-90-365-4747-5
49. Achterkamp, R. Towards a balanced and active lifestyle. June 2019, ISBN 978-94-6323-656-0
50. Engbers, C. Keep Cycling. How technology can support safe and comfortable cycling for older adults. September 2019, ISBN 978-90-365-4848-9
51. Ommeren, A.L. Offering a helping hand. Getting a grip on needs and preferences of stroke patients regarding soft-robotic technology supporting hand function. October 2019, ISBN 978-90-365-4835-9
52. Ter Stal, S. Look who's talking. Appearance of embodied conversational agents in eHealth. March 2021, ISBN 978-90-365-5126-7
53. Beinema, T.C. Tailoring coaching conversations with virtual health coaches. December 2021, ISBN 978-90-365-5260-8



Roessingh
Research and Development

Progress in rehabilitation science

53



ISBN 978-90-365-5260-8