Understanding the interplay between technology and social ties in later life:

How social ties promote use of technology and how technology can promote social relationships

Svetlana Nikitina

Advisor partner University: Prof. Fabio Casati

Advisor home University: Prof. Olga Berestenva

Department of Information Engineering and Computer Science
University of Trento

This dissertation is submitted for the degree of
Doctor of Philosophy

October 2019

Acknowledgements

I would like to express my gratitude to the many people who supported me during my PhD. I would like to thank my supervisor prof. Fabio Casati and my colleagues for the guidance and support through the PhD path, that was challenging, intensive and sometimes not an easy journey. For the things created, failed and learned together.

I would like to thank my family and my fiancé for endless support, care and inspiration you surround me, that gifted me confidence and freedom for brave steps forward and my new beginnings. My friends, thank you for advising and supporting me during this path.

I'm thankful to Russian and Italian Ministry of Education, The European Union's Horizon 2020 Research and Innovation programme for for supporting my PhD in University of Trento.

Abstract

Meaningful social connections are an important part of our lives, especially as we age, and are associated with life satisfaction and psychological well-being. At the same time making friends and creating connections is known to be challenging in older age.

In this thesis, we focus on studying how technology can help to collect information about older adults that can be useful for facilitating friendship formation and social interactions among users.

We start by describing early work that shows the opportunities of technology in improving well-being of older adults. The conducted studies and review work highlights the potential of social interactions in motivating older adults for technology use and exercising.

We then study factors affecting people's social connectedness and friendships. The study highlights that common life points are related to higher levels of connectedness and frequency of interactions. We then move the focus on studying friendship formation in later life, and specifically on how technology can help to facilitate friendship formation.

From observations in the nursing homes we see that *reminiscence* is often used to collect information about a person's history and values, we look at this practice as a way to identify information potentially useful to recommend friendships, especially in nursing homes context.

We conduct Interviews and observations with nursing homes stakeholders and gerontology doctors to define requirements and opportunities of reminiscence conversational agent suitable to their current practices. We then conduct a study to explore how the concept of the bot and features are perceived by elderly, NH staff and doctors.

Finally, we present the work carried out to define and validate the concept of a reminiscence-based conversational agent aimed at: i) conducting storytelling conversations that are engaging and natural and ii) being effective in collecting information about the user (e.g values, interests, places) that later can be used for recommending potential friends.

Keywords: Older adults, Reminiscence, Loneliness, Social isolation, Social interactions, Conversational agents

Table of contents

List of figures

Chapter 1

Introduction

Meaningful social connections are an important part of our lives, especially as we age, and are associated with life satisfaction and psychological well-being [61, 94].

Social isolation is the lack of social connections (network size) and communication interactions. It is an objective quantitative measure of the number of contacts of a person [72]. *Loneliness* is instead the (subjective) dissatisfaction with the quality of social contact, and requires subjective measurement (e.g., UCLA Loneliness Scale) [72].

Both social isolation and loneliness have a strong negative effect on physical and cognitive health as well as on well-being in late adulthood, and results in much higher mortality rates (29% and 26%, respectively), which are figures as high as the one for death caused by well-established risk-factors such an alcohol [72]. Extensive research has found that social isolation leads to elevated blood pressure, dementia, depression, and cognitive decline [21, 61, 71, 184], with recent studies pointing to negative effect of social isolation and loneliness on physical health, but a stronger effect of loneliness in mental health [41].

In addition to having severe adverse effects, isolation and loneliness are also widespread and increasing. Surveys from USA [48], China [189] and Europe (European social survey, self-rated loneliness scale) [190] report on the presence of social isolation and loneliness in older adults on these different geographical and cultural regions with measures varying depending on the country and the scale used, with more marked presence on the oldest old [190]. For instance, the results of nationwide survey in Finland (2002) with a sample of 3858 community-dwelling older adults (75+) [161], show the presence of social isolation in 46% of the older adults, and 37% suffering from loneliness. The presence of loneliness and social isolation also became higher in the last years. Loneliness has been increasing (from 20% to 35%) in the US in just a decade (2000 - 2010) [48] for people 45+, and similarly (from 15% to 29%) in China in 8 years for older adults 60+ (national surveys of older people done in 1992 and 2000) [189].

Changes in social roles and limitations to participation in social activities put older adults at risk of loneliness and social isolation. Research shows that factors contributing to social isolation can be loss (in its many forms, e.g., loss of a spouse, loss of social capital, loss of respect), poor physical health, mental illness, low morale, being a caregiver, geographic location, communication and transport difficulties [47, 63, 55]. For these, and many other reasons, older adults find it difficult to make new friends or keep contact with friends, family and especially younger family members.

The use of technology might help to reduce loneliness and promote social interactions. Previous reviews [39, 36, 113] summarize years of research and show not only that interventions can be effective but also that computer and internet technologies can be effective in reducing loneliness among older adults. Technology-supported interventions aimed at promoting social interactions in older age explore cases with the use of such technological components as: digital frames [83, 174], social networks [130], photo albums [134], storytelling solutions [16] and telementoring systems [18]. Some of the tested technologies are tailor made by the researchers [83, 130, 16, 18] and some are off-the-shelf products (e.g Livescribe pen, Vodafone photo frame) [174, 134].

However social interactions are not only a risk factor (when it's absent), they can be a persuasive factor for older adults to use technology and to overcome an initial barrier or fear. Recent studies applied and explored social persuasion to motivate older adults to participate in exercising activities. In the study of Mubin et al. [116] with collaborative game authors conclude that social interaction was an important positive factor. In the Age invader study [90] with VR environment for playing interactive games remotely researchers also report that social interaction was one of the most important key for game enjoyment.

To explore the opportunities of social interactions as persuasion method for technology adoption and use, our research group has studied the effect of social interactions on the participation to fitness activities in an online virtual gym. The technology support for the study was provided by *Gymcentral*, a tablet and Web app developed by the team that allows trainees of different functional abilities to follow online group exercises from home, watching the exercises videos under the supervision of a remote coach. The application included the virtual environment representing a fitness club, where older adults could meet, train and interact sharing the same virtual space. The development of the application was motivated by the evidence that regular physical activity has strong effect to health and well-being in older age [3, 150, 97] and the goal to provide the possibility for physical training for older adults who due to mobility limitations can not leave their homes and train to support their health.

The study was conducted in Trento, Italy, with 65+ older adults (including adults 80 and over) Participants were recruited from local retirement organization and randomly assigned to 2 groups (control and social). The *social* group had access to social persuasion features in the app (virtual co-presence, messaging) and users were able to see and chat with each other while in the control group the users were able to communicate only with the remote coach. Both groups were performing physical exercises via Gymcental application following an 8-week exercise program.

The results of the study showed that presence of social interactions among participants in a group positively affected to the motivation for training and use the application. The level of attendance registered by the app in the social group was higher than in the control group, co-presence of users caused longer duration of stay in the classroom, more than 50% of received online invitations caused the participants to start training immediately.

Inspired by the results of the study and given the limitation that all the study participants were from the same retirement organization, we decided to explore whether the effect we have observed in the social condition group persists when people do not know each other before using the tool. We have formulated the research question as following:

RQ1. How do online group exercising and baseline measures influence the adherence of older adults to a training program?

We have replicated two follow-up pilot studies with groups of older adults on studying remote physical exercising via Gymcentral application in Tomsk, Russia that adhered to the same protocol and conditions, except for the group cohesion (the factor of knowing each other) setting:

i) Tomsk 1 study - participants with high group cohesion, recruited from two organizations, and with the majority performing shared activities (computer courses and hobbies classes).

ii) Tomsk 2 study - participants with low group cohesion, recruited from various organizations, with weak or no ties with each other.

Both pilot studies were follow-ups to a previous pilot performed in Trento, Italy, and so they follow the same study design.

One of the important findings from the studies was that participants who knew each other before the experiments (high group cohestion) and had some common interests (e.g participated together in the same retirement/leisure organizations) were more socially active during intervention and tend to communicate more via the app with other participants compared to people who didn't know each other before the experiment [121].

Other important result was that in the Tomsk studies online group-exercising (social condition) did not result in higher adherence when compared with individual training (with persuasion features). We didn't observe the same affect as in the previous study conducted with italian older adults in Trento where social group condition had higher adherence to the training.

We attribute this effect to: i) weaker cohesion among participants in Tomsk2, which might have reduced the effect of normative influence and peer support, ii) persuasion features in the individual training condition that raised the adherence by 10% compared with our previous study [51]. This results adds to the evidence that group-exercising in low cohesion groups results in an adherence comparable to that of individual training with contact (with a coach). It also partially supports the evidence that group exercising in high-cohesion groups results in higher adherence than individual training with contact. The Gymcentral application and the studies conducted are presented in the detail in the Chapter 2.

In the light of this findings we have started to research on the factors affecting to the people's motivations behind social connections and friendships. Such knowledge can help to identify how to group and match people meaningfully in online application that involve social interactions.

Specifically, we investigate the following research questions:

RQ2. Are common life points related to the level of online and face-to-face interactions, and how?

This research question can help us to understand if and how we can predict the frequency of social interactions based on the similarity of people.

RQ3. Is connectedness related to common life points and frequency of online and face-to-face interactions, and how?

Recent research on the friendship formation topic suggests that one of the core factors affecting to friendship are: i) proximity ii) common life points (values, interests, activities). People who have common values and shares same interests are more tend to communicate and become friends.

To answer our research questions, our research group have conducted a study exploring the relationship between connectedness, social interactions and common life points of user's on Facebook [142]. We observed that the more common life points (shared beliefs, activities, and interests) friends share the more frequent their online social interactions are, and that shared interests and activities are determinant to this effect. We have also seen that higher levels of interaction and common life points are related to higher levels of connectedness. This suggests that one potential direction to creating bonds is generating opportunities for similar people to have meaningful interactions.

After exploring factors the are correlated with people's connectedness and friendship, we focus on studying friendship formation of people in later life, and specifically on how technology can help to collect information about older adults that can be useful for facilitating friendship formation.

Indeed, making friends in older age is known to be challenging. Exploratory studies in nursing homes in Italy from our research group (Ibarra et al. 2017) [78] with elderly residents, their relatives and NH staff showed that:

- it is hard for residents to establish friendship relationships in the nursing home;
- there is an issue to introduce a new member in a NH community;
- NH staff conducts activities to stimulate communication, such as storytelling, reminiscence, games;

To tackle the challenge, we start from the concept and activity of *reminiscence*. Reminiscence is the process of collecting and recalling past memories through pictures, stories and other mementos [181]. The practice of reminiscence, and more generally, reminiscence therapy and life review, have well documented benefits on emotional and mental wellbeing [15, 151], interpersonal relationships and interactions [5, 77], and preserving personal identity [42, 57], making it a very desirable practice, especially for older adults.

Because reminiscence is often used (also in nursing homes) to collect informaton about a person's history and values, we look at this practice as a way to identify information potentially useful to recommend friendships, especially in nursing homes context. Therefore, based on the findings from the studies mentioned above, we propose the concept of a *reminiscence chatbot*, a smart conversational agent that can drive personal and social reminiscence/storytelling sessions with older adults in a way that is engaging and fun, while effectively collecting and organizing memories and stories. The tool has the purpose of: i) engaging older adults in storytelling conversations that are stimulating and fun, ii) collecting and organizing memories and profile information (learning), iii) proposing potential friends based on the possible common life points. The concept of the reminiscence conversational agent is explained in the detail in the Chapter 5.

The objectives of our work are therefore to define the requirements and opportunities for the reminiscence chatbot that leads to the successful human-chatbot interaction. We consider a reminiscence chatbot to be successful if i) it is engaging and natural, but also ii) it is effective in collecting information about the user (e.g values, interests, places) that later can be used for recommending potential friends. Furthermore, since reminiscence is a delicate activity which can be emotionally intense if not run properly, the chatbot needs to follow accepted practices for ensuring an emotionally safe reminiscence session.

We pose the following research question:

RQ4. Is a chatbot acceptable (both to our target users and according to doctors) as an instrument for storytelling (reminiscing) and collecting information about the person?

To answer RQ4 we run a study to explore and evaluate: i) the feasibility and acceptance of a reminiscence- based storytelling conversational agent with different stakeholders: elderly, nursing home staff and gerontology department doctors, ii) perception of the reminiscence tool activities/functionalities and expected benefits, iii) preferred context of use of the tool by the stakeholders.

1.1 Methods and thesis structure

The work presented here is in large part based on research publications conducted during the years of the doctoral studies. For clarity, we will include the citations to these publications. Below, we present the methods followed during our research work, indicated as these are structured in the book.

Chapter 2. Social Interactions. Active aging

A summary of the studies run using Gymcentral in Tomsk, Russia.

The content of this chapter integrates the studies published as:
Nikitina, S., Didino, D., Baez, M., & Casati, F. (2018). Feasibility of Virtual Tablet-Based Group Exercise Among Older Adults in Siberia: Findings From Two Pilot Trials. *JMIR mHealth and uHealth*, 6(2), e40.

My contribution in this work is: conducting the literature review, recruiting participants, planning and performing the experiments, contributing to the data analysis, preparing figures and tables, contributing to writing the paper.

A review on technologies for home-based training to help older adults remain physically active – in particular, those who cannot leave their homes regularly or easily.

The content of this chapter has been published as: Khaghani-Far, I., Nikitina, S., Baez, M., Taran, E. A., & Casati, F. (2016). Fitness applications for home-based training. *IEEE Pervasive Computing*, 15(4), 56-65.

My contribution in this work is: conducting the literature review and market apps review, contributing to data analysis, contributing to writing the paper.

Chapter 4. What Makes People Bond?: Social Interactions and Common Life Points

A section on friendship formation and social bonds.

An online survey run on Facebook using profile and interactions information to understand the relationships between connectedness, social interactions and common life points.

The content of this chapter has been published as:

Sanchiz, E., Ibarra, F., Nikitina, S., Báez, M., and Casati, F. (2016). What makes people bond?: A study on social interactions and common life points on Facebook. In *2016 International Conference on Collaboration Technologies and Systems (CTS)*, pages 26–30 [142].

My contribution in this work is: conducting the literature review, contributing to writing the paper.

Chapter 5. Concept of conversational agent for reminiscence and re-connection.

Conversational agent design and model

Here: The design and the concept of the reminiscence chatbot. Conceptual model, data model, interface mockup.

We explore the needs and challenges of nursing home stakeholders, with particular attention to current practices from relatives, nursing home staff and the doctors who practice reminiscence sessions.

The other part of content of this chapter has been published as: Nikitina, Svetlana, Sara Callaioli, and Marcos Baez. "Smart conversational agents for reminiscence." arXiv preprint arXiv:1804.06550 (2018).

Designing and running crowdsourcing experiments in order to define the conversational patterns for reminiscence sessions.

The part of content of this chapter has been published as: Nikitina, Svetlana, Florian Daniel, Marcos Baez, and Fabio Casati. "Crowdsourcing for Reminiscence Chatbot Design." arXiv preprint arXiv:1805.12346 (2018).

My contribution in these two papers is: conducting literature review, designing the chatbot conceptual model, contributing to the experiments design and conducting experiments, contributing to writing the paper.

Chapter 7. Understanding the requirements and behaviors for reminiscence chatbots

Building on the concept developed for our reminiscence-based conversational agent, we conduct validation studies with nursing home stakeholders (residents, staff members, doctors), considering opportunities and challenges for the use of the tool.

My contribution in this work is: planning and designing the study, preparing the storyboards, conducting the interviews, analyzing the data.

Chapter 8. Conclusions and future work

We close by summarizing the contributions of this work and connecting our results to the research questions of this thesis (RQs).

Chapter 2

Social Interactions. Active aging

2.1 Introduction

2.1.1 Background

Regular physical activity is a key factor to a successful ageing, contributing to positive outcomes in health and well-being in later life [156, 3, 150, 97]. It can improve physical function [97], slow the progression of degenerative diseases [150], reduce risk of falls [156] but also improve cognitive performance, mood and quality of life of older adults [3, 97]. A physically inactive lifestyle, on the contrary, can increase the risk of developing chronic diseases, one of the leading causes of death and disability in older adults [12, 186].

Engaging in regular physical activity can be challenging. Older adults might suffer from reduced mobility, low self-efficacy, lack the proper infrastructures in their communities or simply find it difficult to leave home and participate in physical activities on a regular basis [132, 46]. For these and many other reasons, physical inactivity is still prevalent in older adults [69], leading to the undesired effects on health and well-being.

Intervention programs to promote physical activity have shown to be effective in increasing and maintaining physical activity [49]. In particular, group-based interventions have shown promising results in long term settings, with higher adherence compared to individual home-based interventions. Studies have also reported a preference by older adults for group exercising [43], and discussed the potential of the social context to motivate social interactions and increase social well-being [51].

However, despite the body of literature on the topic, little attention has been paid on populations living under difficult environmental conditions and undergoing complex social changes, such as the Siberian community. Seasonal fluctuation has been found to determine the level of physical and social activities of older adults [129] leading to less opportunities to go out and interact, especially in high latitudes where winter can result in a decline of physical functions of older adults, such as a ankle strength [19]. Recent history has also shaped the lives of older adults in Russia. The breakup of Soviet Union in the early 90s, and the difficult years that followed, negatively affected the social and economic well-being of Russian population: The life expectancy of men is 14 years lower than in the European Union [149], and loneliness levels are among the highest in Europe [188]. The social, political, and economical uncertainty also deeply affected quality of life, with a decrease in life satisfaction and happiness [2].

The above observations point to the need for solutions that can help older adults living under the above conditions to keep physically and socially active. Technology-supported interventions have been shown in the past to be successful in this goal [117].

2.1.2 Related work

Recent research has demonstrated an effectiveness of technology-supported exercise interventions for older adults in terms of physical fitness [117]. However, although there is an ongoing discussion on whether group exercising or home-based individual exercising is more effective in increasing adherence of individuals to training programs (e.g., [173, 8]) and despite calls for analysis focusing on understanding group-based exercising in terms of cohesiveness (frequency of contact and group dynamics) [26], no intervention has compared the effectiveness of individual and (different types of) group settings in a technology-supported intervention.

Research has also shown a preference by older adults in group training [43, 51]. However, implementing group exercising can be challenging, especially in a heterogeneous elderly population, with individual differences leading to motivational issues and problems in tailoring the training [43].

Fitness applications for home-based training have been widely explored in technology-supported interventions (see [88] for a review), however we are not aware of interventions supporting online group-exercising for individuals of different levels of fitness. Consequently, there is very limited research on the effects of level of fitness, social support and subjective well-being in online group settings. The exception comes from a recent study on an internet-based group training intervention [50] relying on a general-purpose teleconference software to deliver real time exercises to older adults in rural areas. Although targeting homogeneous groups, focused on physical fitness outcomes, and limited to a small sample of 10 older adults, the study highlights some interesting challenges in deploying this type of technology.

In our previous study [51, 10], we made some steps to test the feasibility of a tool for online group-exercising, namely Gymcentral, that allows individual of different levels of fitness to follow exercises with the remote company of others. We conducted a 8-week pilot study exploring the effects of online group exercise training in Trento, Italy, with 37 adults 65 years old and above, who followed the Otago exercise program [65] aiming at strength and balance improvement in older age. The specific focus of the study was on technology acceptance, attitude and preference towards group training, and effects on physical and social well-being.

Participants were randomly assigned to either an intervention or control group. The intervention group was performing online group exercises, with access to personalised training, persuasion strategies and the ability to interact socially via app messages with other trainees and the coach. The control group was assigned to an individual training condition, limited to personalised training and communications with the coach only via phone calls. Besides indicating high usability of the application, results point to i) acceptance and preference of older adults towards online group training, ii) to higher adherence to the online group exercising compared to individual training with no persuasion strategies, iii) to the potential of the group setting to reduce the negative effect that a low starting level of fitness has on adherence, iv) and to positive effects on physical and well-being outcomes for both intervention and control groups.

Still, despite the prior work and the extensive existing literature, open questions remain:

1. How does the online group exercising translate to other cultural and environmental settings?

2. How effective is online training with groups of different levels of cohesion?

3. How does online group exercising compare with individual training featuring persuasion strategies?

2.1.3 Objectives

In this chapter we report on two pilot studies of an online exercise intervention with older adults living in Tomsk, Siberian Federal District (Russia). The aim of the intervention was to enable older adults of different levels of fitness to follow a personalized exercise program from home, with the (virtual) company of training companions and under the supervision of a remote coach. This was done with the support of a tablet app offering group exercising in a virtual gym while leveraging on the social context of the group exercising to enable social interactions and feedback.

The main objective of the pilot was to study the feasibility of online group exercising under different cohesion settings among Siberian older adults. We focused on the technology acceptance, on the adherence to the training (especially in light of pre-measures of social support, as well as on the enjoyment of physical activity and leg muscle strength). As a secondary objective, we also explored the effects of the technology-supported intervention on subjective well-being and loneliness.

In addition to testing a novel technology in a home-based physical intervention and pursuing the above objectives, this work contributes to the body of literature on the topic with the first reported intervention in a Siberian region. The technology-based intervention reported in this paper has been already studied in an Italian context [51, 9], under different but complementary settings, which we take as an opportunity to compare results and discuss cultural differences.

2.2 Methods

2.2.1 Gymcentral Training Apps

The technology support was provided by Gymcentral, a tablet and Web application that allows trainees of different functional abilities to follow online group-exercises from home, under the supervision of a remote coach [10]. Gymcentral serves the needs of trainees and coach via the *Trainee* and *Coach* apps (see Figure 2.1).

The design of the trainee app is based on a virtual gym environment that provides the following main features:

- Tailored training program. It delivers video exercises that are tailored to the abilities and progress of individual trainees. Trainees may receive exercises of different intensity level or not receive some exercises depending on their condition and the Coach assessment.

- Online group exercising. It allows trainees to participate in online group exercise sessions in a virtual classroom. Trainees can see the video of the coach and also the presence of other trainees via avatars. However, differences in functional abilities or the intensity level of the exercises remain hidden.

- Persuasion strategies. It provides individual persuasion features such as positive and negative reinforcement and self-monitoring (implemented using a growing garden metaphor), as well as social persuasion features such as social learning, social support, social facilitation, and normative influence.

- Remote monitoring and feedback. Participation to training sessions and completeness of exercises are recorded by the app and made available to the training coach. The coach can act on this data to provide feedback (using the communication features) and increase or tailor the intensity of the training program.

- Communication features. It enables trainees to share public messages with all the other trainees in a bulletin board or to exchange private messages with individual trainees (or the training coach) using an internal messaging feature.

The monitoring and feedback is supported by the coach app, a companion Web app for the training expert.

Details about the features of the Gymcentral application are discussed in detail in the study by Baez et al [10], and the underlying conceptual model by Far et al [52].

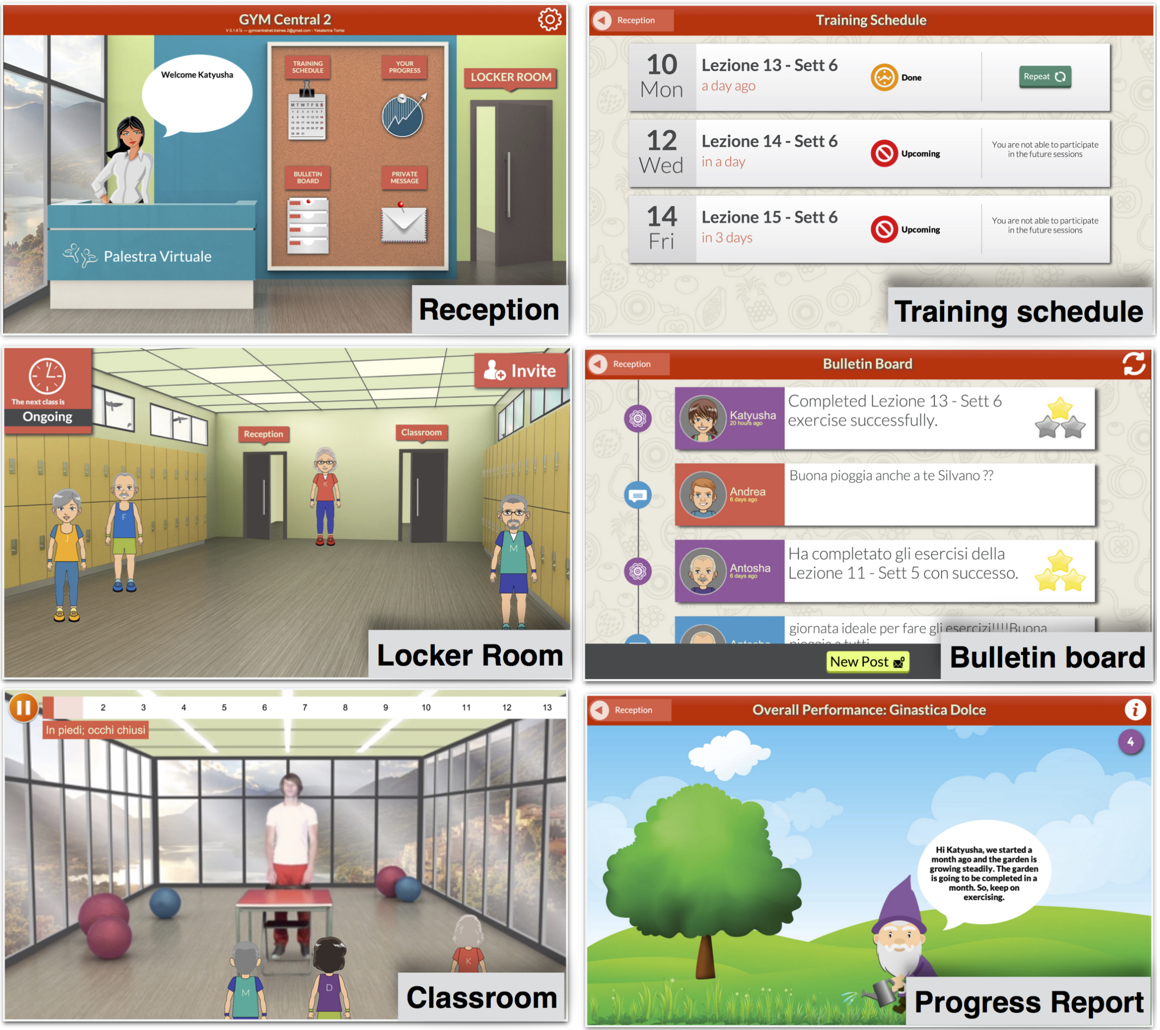


Fig. 2.1 Features of the virtual gym environment of the trainee app.

2.2.2 Research Questions

In this work, we studied the feasibility and effectiveness of the online group exercise intervention and its effects on the well-being of Siberian older adults by addressing the following specific research questions (RQ):

RQ1.Is the online group exercising technology usable and accepted by older adults? We aimed at exploring the perception of older adults toward the technology by measuring the usability and acceptance. More importantly, we also explored how the app was used in practice and how the usage relates to the observed effects.

RQ2. How online group-exercising and baseline measures influence the adherence of older adults to a training program? Previous research suggests that exercising in a group results in higher adherence and preference by older adults [132, 43]. However, research also points to major obstacles when delivering group-exercises to heterogeneous populations, which can make training in this setting difficult and less motivating [43]. In this study we explore how a virtual group environment influences the adherence of older adults, under different measures of known determinants of physical activity.

RQ3. Does online group-exercising affect the well-being of older adults? We explore the effects of physical training via a virtual social environment on subjective well-being and social well-being of participants. By addressing this question, we aim at contributing to the existing research on the association between physical training and well-being [156, 3, 150, 97].

2.2.3 Study Design

We explored the above questions in two pilot studies in Tomsk, Siberian Federal District (Russian Federation) that adhered to the same protocol and conditions, except for the group cohesion setting:

- Tomsk1. (July 2015- September 2015) Participants with high group cohesion, recruited from two organisations, the majority performing shared activities (computer courses, hobbies classes).

- Tomsk2. (April 2016- June 2016) Participants with low group cohesion, recruited from various organizations, with weak/no ties with each other.

We use the term cohesion to describe the differences in group compositions. By cohesion, we determine the fact of knowing each other between participants in a group. Therefore, its a subjectively measured condition. We define only two levels of cohesion: strong and

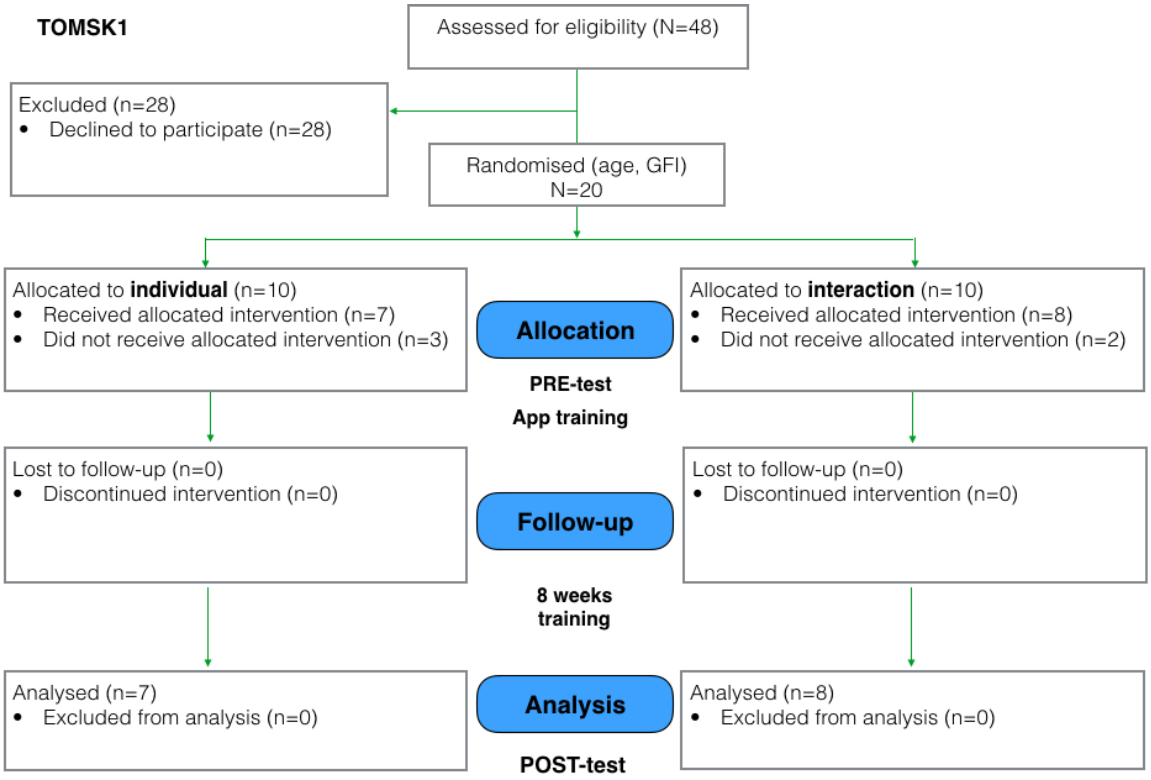


Fig. 2.2 Study flowchart for Tomsk1 (July 2015-September 2015).

low group cohesion. The participants from group Tomsk1 already knew each other before the experiment as they have attended the same hobby club. For this reason, the Tomsk1 group has strong cohesion while the Tomsk2 group has low cohesion since their participants did not know each other before the experiment. As seen above, we explored two group cohesion settings: participants with *strong group cohesion* and participants with *low group cohesion*. Thus, for the reasons explained above, candidate participants from Tomsk 1 had a stronger cohesion than Tomsk 2 at recruitment time, regardless of the treatment they ended up receiving. We did so to understand the effect of the prior connectedness among participants on the observed outcomes.

Both pilot studies are follow-ups to a previous pilot performed in Trento, Italy, and so they follow the same study design [51]. An overview of the study flow in CONSORT-compliant format is shown in Figures 2.2 and 2.3.

In both studies described here, participants were assigned to an interaction group (online group exercise condition) or to an individual group (individual exercise condition) using a random assignment procedure, with age and participants' frailty level as random assignment variables. In Tomsk 1, the process was slightly different as to ensure a high level of cohesion

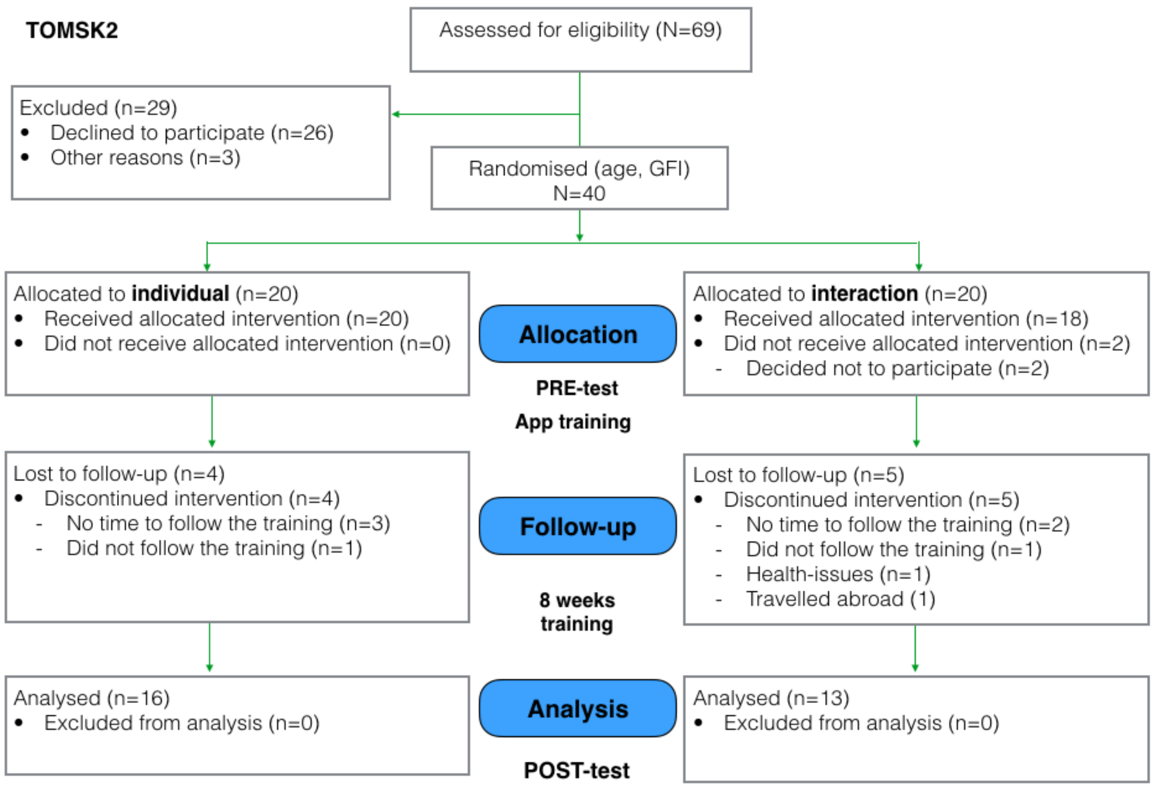


Fig. 2.3 Study flowchart for Tomsk2 (April 2016-June 2016).

after randomization: pairs of friends, identified during the informative meeting, were treated as single elements during randomization. In this modified process, we firstly followed the randomization procedure for participants *without friends*, assigning participants to interaction and individual treatments, and then repeating the process for the friend pair units. Thus, friends were assigned to the same treatments, contributing to the overall group cohesion in Tomsk 1.

The two studies and the two treatment conditions defined four effective groups (see Table 2.1. Participants in the interaction groups have access to online group exercising with social interaction and persuasion features, whereas in the individual groups, participants have access to individual training with persuasion features but with social interactions limited to contacts with the coach. Details about the group cohesion and features available to each group can be seen in Table 2.1.

Both versions of the application implemented the same training program, developed on the basis of the OTAGO Exercise Program [64], which includes a set of muscle strengthening and balance-retraining exercises. In the application, each exercise had 10 levels of intensity based on the duration and the number of repetitions. At the beginning of the study, a personal trainer (who was also the coach in the virtual gym) performed a physical assessment, which was used to set the starting intensity level of the program.

Before the start of the training program, participants joined pretest and technology training meetings: (1) an initial meeting where they signed the informed consent and filled out enrollment questionnaires, (2) a session with a medical doctor to evaluate eligibility, (3) a technology training session in the use of tablets and the assigned version of the app, and (4) a session for the physical assessment with the coach and pretest measures. The technology training followed a workshop format and was done in small groups of 10 participants each. Participants assigned to individual and interaction conditions attended workshops separately as they were provided with different versions of the app.

In the 8 weeks of the training, participants performed the home-based training activity with the monitoring of the coach and of the support staff. The training schedule offered three exercise sessions per week, and participants were required to perform at least two exercise sessions every week. The duration of the training session ranged from 30 to 40 min depending on the intensity level. Participants were free to join the training sessions at any time. Post-test measures took place on the week after the training.

The coach guiding the participants during the training was a practicing doctor with a primary care doctor degree and had over 10 years of experience in gymnastics, rehabilitation exercises, and yoga for older adults. Before the beginning of experiment, the coach was acquainted with the Otago training program and Gymcentral app settings.

During the training period, the coach had the task of progressing the intensity of the exercise program and providing feedback. At the end of every week, the coach could maintain or increase the intensity level of each trainee according to the attendance and completeness of the training sessions in the week. The coach was also instructed to contact trainees at least once a week to provide feedback and to respond to any question from the trainees. The coach was not aware of the difference between the interaction and individual groups, and both received the same amount of technical support.

The pre-test measures included the *Groningen Frailty Indicator* [148], the *Rapid Assessment of Physical Activity Questionnaire* [162], demographic information, questionnaires concerning psychological and social well-being. The post-test measures included the *System Usability Scale* [23], a set of questions on the acceptance of the application, the *Satisfaction with Life Scale* [45, 169], the *MOS Social Support scale* [145, 114], and the 3-item *R-UCLA Loneliness Scale* [140, 76]. The participants filled in all the questionnaires by themselves in pencil-and-paper format.

The study protocol received ethical approval from the CREATE-NET Ethics Committee on ICT Research Involving Human Beings (Application N. 2014-001) in Trento, Italy. The

Table 2.1 Features of the Trainee App available to each study group

Features	Interact.	Control
Tailored exercises program (OTAGO)	x	x
Training with others in the classroom	x	
Invitation to join a training session	x	
Self-monitoring progress (garden metaphor)	x	x
Positive / Negative reinforcement	x	x
Sharing of training activity the in bulletin	x	
Contextual messages in the Locker room	x	
Public messages in the bulletin board	x	
Private messages with other trainees	x	
Private messages with the Coach	x	x

studies reported in this paper — as follow-ups to our previous study — comply with this protocol, with the informed consent and informational materials translated into the Russian language.

2.2.4 Participants

We considered eligible for the study: participants aged 59 years or older, independent living, self-sufficient, and with a non-frail,transitionally frail or a mild frailty level. These criteria were measured by self-reports. All participants had to pass a doctor assessment to ascertain the absence of conditions that would prevent them from performing light physical exercises. Participants wearing pacemakers were considered not eligible as the study required the use of an activity sensor (Misfit shine monitor). The specifics of baseline measures for each study site are described in Table 2.2.

Participants in both studies were contacted through retirement organizations in Tomsk, Russia. In the first study, Tomsk1, participants were mainly invited through organization offering computer-learning classes and hobbies activities for seniors. In the Tomsk2 study, the recruitment was carried out through three organizations organizing social activities and events. We conducted presentations explaining the project and their expected involvement and handed out printed bulletins. Older adults interested in participating provided their phone numbers and were later on contacted by the project coordinator. Details about the retirement organizations and the number of candidates reached can be seen in Table 2.4.

In the Tomsk1 study, 20 participants were found eligible for the study (mean age individual group=65, SD 6.1; interaction group: mean 68.2, SD 7.8; 19 females and 1 male). In the Tomsk2 study, 40 participants were accepted according to the inclusion criteria (mean

Table 2.2 Baseline measures for each study site

	Site	Control	Interact.	p^*
Age, *M*(SD)	Tomsk1	65.0 (6.1)	68.2 (7.8)	>0.1
	Tomsk2	68.8 (7.2)	67.6 (6.2)	>0.1
Females, %	Tomsk1	100%	87.5%	
	Tomsk2	100%	100%	
GFI, *M*(SD)	Tomsk1	4.2 (2.04)	4.5 (2.42)	>0.1
	Tomsk2	3.6 (2.54)	3.56 (2.5)	>0.1
RAPA, *M*(SD)	Tomsk1	5.78(1.79)	5.9 (1.73)	>0.1
	Tomsk2	5.15(2.41)	5.13(1.96)	>0.1
Self-reported, after allocation, *M*(SD)				
PACES, enjoyment	Tomsk1	50.0 (3.5)	50.0 (4.8)	>0.1
	Tomsk2	49.9 (5.4)	47.8 (4.2)	>0.1
R-UCLA, loneliness	Tomsk1	4.2 (1.6)	5.4 (1.4)	>0.1
	Tomsk2	4.3 (1.1)	4.0 (1.2)	>0.1
MOS, social support	Tomsk1	4.0 (1.5)	5.1 (1.6)	>0.1
	Tomsk2	4.3 (1.1)	4.0 (1.2)	>0.1
SWLS, well-being	Tomsk1	4.0 (1.5)	5.4 (1.4)	>0.1
	Tomsk2	4.3 (1.1)	4.1 (1.2)	>0.1
Physical assessment, after allocation, *M*(SD)				
Leg muscle strength	Tomsk1	13.6 (2.2)	12.9 (1.4)	>0.1
	Tomsk2	16.5 (3.8)	16.5 (3.0)	>0.1

[*] Differences computed using independent samples t-test for age and leg muscle strength; all the other variables were analyzed with Mann-Whitney tests.

Table 3. Senior citizen organizations contacted and candidates reached in each study.

Retirement organization	Study	Size of groups reached
Tomsk union of retirees	Tomsk 1	Large organization providing courses to around 600 retirees per year. Four active courses at the time (approximately 20 members each) were contacted, reaching around 80 older adults in total
Veterans council of Tomsk Polytechnic University (TPU)	Tomsk 1	Small organization of around 80 retirees. The invitation was extended to all members
Veterans Council of Tomsk Scientific Center	Tomsk 2	Small organization of around 80 retirees. The invitation was extended to all members
Tomsk region veterans council	Tomsk 2	Small organization of around 100 retirees. The invitation was extended to all members
Veterans council of TPU	Tomsk 2	Small organization of around 80 retirees. The invitation was extended to all members

Fig. 2.4 Senior citizen organizations contacted and candidates reached in each study.

age individual group=68.9, SD 7.2; interaction group: mean 67.6, SD 6.2; all 40 female). The difference in the number of male and female participants is because of the demographics of the study location and the availability of male candidates at the retirement organizations. In Siberia, lifespan gap between males and females is one of the biggest in the world: life expectancy at birth for men is 64.7 years, whereas for women it is 76.3 years [127]. These demographics posed difficulties in recruiting male participants from the retirement organizations. The study flow for Tomsk1 and Tomsk2 is depicted in Figures 2.2 and 2.3. After the recruitment, participants in both studies signed the informed consent before participating in the experiment.

In the Tomsk 1 study, out of 20 participants, 5 withdrew before the start of the study for health problems or personal reasons; therefore, data of 15 participants was included in the analysis. In the Tomsk2 study, out of 40 participants, 2 withdrew before • the beginning of the training because of travel plans. During the training program, 4 participants in the individual group and 5 participants in the interaction group dropped out because of health issues, travels, or reported lack of time for participation. Thus, in the Tomsk2 study, a total of 29 participants were included in the analysis (individual: 16, interaction: 13).

There were no statistical differences between individual and interaction groups in term of initial measures Table 2.2. These baseline comparisons have been performed on participants that finished the training program.

2.2.5 Outcome Measures

Acceptance and usability

We focus on the usability, acceptance of the technology, and preference to train together:

- ***Usability:*** The usability of the application was evaluated by means of the System Usability Scale [23]. This scale includes 10 items rated on a 5-point Likert scale (from 1 = “completely disagree" to 5 = “completely agree"). The SUS score range from 0 (low usability) to 100 (high usability). However, in a pretest of the scale, older adults found difficult to understand two items (“I found the various functions in this system were well integrated" and “I thought there was too much inconsistency in this system"). Therefore, we decided to exclude these two items in the questionnaire we administered to our participants. This mean that the SUS score in our study ranged from 0 to 80.

- ***Acceptance:*** Acceptance was measured with a set of questions designed to evaluate positive (“I enjoy using the app") or negative feelings (“The app makes me nervous") associated with the use of the applications, the response to the communication feature (“It is easy to communicate with other people with the app"), the intention to use it (“I would like to use the app in the future"), and the perceived ease of use (“It is easy to use the virtual gym to perform exercises"). These questions were rated on a 5-point Likert scale (from 1=“completely disagree" to 5=“completely agree"). The questionnaire was developed by our team on the basis of previous literature [131]. Each question has been separately analyzed.

- ***Co-presence:*** Participants had the choice to train at any time, but they could also coordinate to train at the same time via messaging or using the *invite user to join* feature. To capture the preference of users for group training, we logged the attendance to the training sessions to compute for each user whether he or she trained alone (individual attendance) or together with another trainee (joint attendance). We then define co-presence of a group as the ratio of joint attendances with respect to the total number of attendances.

Adherence to the Training

Measured with: Persistence: Persistence was computed considering the ratio between the number of attendances to exercise sessions by a participant and the number of the exercise sessions planned in the program. Participation was measured by logging the attendance to the scheduled training sessions in the virtual classroom. For persistence, a rate equal to 100% was considered as participation in all three sessions per week, for all 8 weeks of training. Participants were not aware of how the persistence was scored but could monitor the individual progress in the garden (self-monitoring feature).

Subjective Well-Being, Social Support, and Loneliness

To measure if there was an improvement in the well-being outcomes as a result of training (secondary outcomes), we relied on the following instruments:

- ***Satisfaction with Life Scale*** (SWLS) [45]: 5 questions rated on a 7-point Likert scale (from 1 "Strongly disagree" to 7 "Strongly agree"). The SWLS was translated and adapted to Russian language by Tucker et al [169]. The total score ranges from 5 to 35, with higher scores indicating higher levels of life satisfaction.
- ***Loneliness*** To measure loneliness, we used a shorter version of the R-UCLA Loneliness Scale [140] developed by Hughes et al. [76]. The scale used includes 3 items scored on a 5-point Likert scale, with the total score ranging from 3 to 15, and higher scores indicating higher levels of loneliness.

Determinants of Physical Activity

In the analyses explained in the following sections, we use the following determinants of physical activity as covariates:

- Physical Activity Enjoyment Scale (PACES) [87]. This scale includes 16 items scored on a 5-point Likert scale (from 1="disagree a lot" to 5="agree a lot"). The PACES total score ranges from 16 to 80 (maximum enjoyment).
- MOS Social Support [145, 114]: 8 questions scored on a 5-point Likert scale (from 1 "None of the time" to 5 "All of the time"). This scale was translated by us according to the international guidelines [13]. It aims at measuring the social support provided by others. The total score ranges from 1 to 8, with higher scores indicating higher levels of social support.
- Leg muscle strength: Measured with the 30 second Chair Stand test [86]. The purpose of this test is to evaluate leg strength and endurance. From seated position, the participant rises to a full standing position and then sit back down again for 30 seconds. The outcome measure is the number of times the participant comes to a full standing position in 30 seconds.

2.2.6 Statistical Analysis

We analyzed the difference between the Interaction and the Control groups in term of SUS score with two Mann-Whitney tests, whereas for the difference in the percentage of co-presence we use t-tests.

We analyzed adherence (measured as rate of persistence) to the training program with an analysis of covariance (ANCOVA) with group (Interaction vs. Control) and study (Tomsk1 vs. Tomsk2) as between-subject factors, and leg muscle strength, social support (MOS score), and enjoyment of physical activity (PACES score) as covariates.

For well-being measures we selected the SWLS score and R-UCLA Loneliness Scale score as dependent variables to be used in two separate repeated-measures analysis of variance (ANOVA). We use the same independent variables in both ANOVAs: time (pre-test vs. post-test) as within-subject factor, and group (Interaction vs. Control) and study (Tomsk1 vs. Tomsk2) as between-subject factors.

The statistical analyses were performed using the open source statistical software R [155], using the ggplot2 package to create plots [183].

2.3 Results

2.3.1 Perception and Adoption of the Technology

A starting point to understand the feasibility of the technology for our target population was to address (RQ1) and investigate the perceived usability, acceptance, and usage of the online group exercising technology.

Usability

Nine participants did not answer to some of the questions of the SUS and thus have been excluded by the analysis on this account. On average, the SUS score (on an 80 points scale, as we excluded two questions) was very similar between the interaction group (mean 63 [SD 9]; N=19; range 48-80) and the individual group (mean 66 [SD 14]; N=15; range 40-80).

From a more detailed perspective, a Mann Whitney test showed that neither in the Tomsk1 study (W=11, $P \geq 0.99$) nor in the Tomsk2 study (W=89.5, P=0.32) the SUS scores were different between the two groups (individual vs interaction) despite the higher complexity of the app assigned to the interaction groups.

Acceptance

Table 2.3 reports the results for the questions concerning acceptance (A). Consistently with the SUS score, trainees showed a high level of acceptance of the app. In fact, as the Table shows, trainees reported high levels of enjoyment (A1) and low levels of nervousness (A2) in using the app. Training with the app was perceived as very easy to do (A4) as well as

Table 2.3 Mean (SD) of the technology acceptance (A) responses for each group and study (range:1-5).

	Tomsk1		Tomsk2	
	Interaction	Individual	Interaction	Individual
A1 (feel joy)	3.9 (1.4)	3.9 (1.6)	2.8 (1.9)	3.3 (1.9)
A2 (feel nervous)	2.3 (1.2)	1.2 (0.4)	1.4 (0.8)	1.1 (0.3)
A3 (easy social)	4.4 (0.9)	3.0 (2.3)	3.1 (1.7)	4.1 (1.5)
A4 (easy train)	4.9 (0.4)	4.6 (1.1)	4.7 (0.5)	5.0 (0)
A5 (future use)	4.9 (0.4)	4.2 (1.8)	4.6 (0.7)	5.0 (0)

communicating (A3), but with a lower score by 1 point. Trainees also reported with a high score their intention to use the app in the future.

Characterization of App Usage

To characterize the usage of the various features of the app, we analyzed the app logs to derive how participants spent their time in the app. Overall, the mean time spent in-app was higher in Tomsk1 (16 hours) compared with Tomsk2 (9 hours), the difference being marked by a higher time spent by the interaction group in the first study (see Figure 2.5).

Not surprisingly, most of the time was spent training in the classroom, as the duration of exercise session ranged from 20-40 mins depending on the intensity level. Looking at the time spent in the classroom relative to the time spent in-app by each participant, we can see that participants of the Control group in both studies spent nearly the same percentage of their time (Tomsk1 = 95.3%, Tomsk2 = 95.6%) in the classroom. Participants in the Interaction groups spent a little less on the classroom - especially in Tomsk2 (Tomsk1=92.5%, Tomsk2=81.4%). The lower use in the Interaction app is due to the presence of extra features, and in the case of Tomsk2, due to the lower time spent training.

Analyzing the usage of the other features, we observe that participants spent a significant percentage of their time messaging, particularly those in the individual groups (see Figure 2.5). We can derive that the individual group not only used the training feature but also the messaging tool to interact with the coach and to check their progress. The bulletin board and the locker room were not available for the individual group.

The interaction group also used the social features (see Figure 2.5). The messaging feature was used to send private messages to other participants and the coach, especially in Tomsk1. The bulletin board was also used, although visits were more related to a *lurking* behavior rather than actual contributions. We attribute this to automatic sharing of the participant's performance (as a 3-star rating based on completeness) on the bulletin board (*social learning* persuasion strategy [51]). The Locker room comprises also an important percentage but it is mostly due to the fact that it preceded the classroom in the navigation. No important interactions or invitation to the join the classroom were registered from this virtual space.

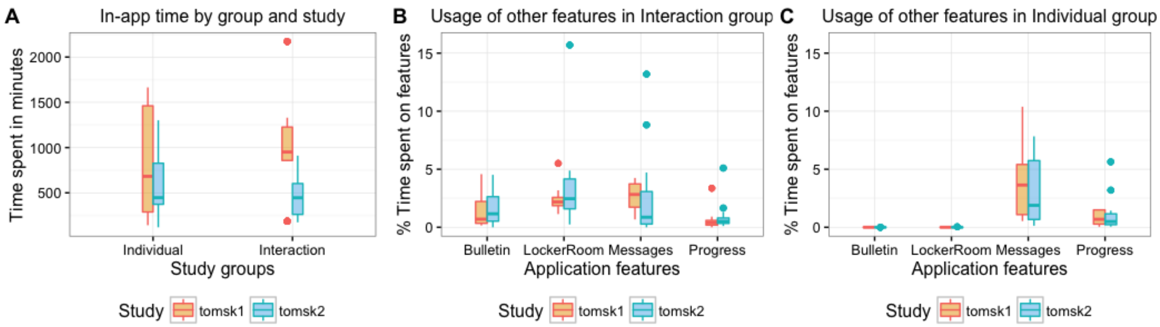


Fig. 2.5 App usage by group and study.(A) Total time (in min) spent by user in the app during the experiment, (B) Usage of the app features in the interaction group, and (C) Usage of the app features in the individual group, % from total time spent in the app.

Online Interactions

Participants in the interaction group had the possibility of exchanging public and private messages either with the coach or other trainees, whereas in the individual group, the interactions were limited to private messages with the coach. Table 2.4 summarizes the exchanges among participants of both groups in the two pilot studies.

Participants in the social condition made significantly more use of private messages compared with public messages. This was the case even for participants in Tomsk1 (strong group cohesion), with 4.4 private messages compared with only 0.6 public messages per user. Not surprisingly, participants of Tomsk1 interacted significantly more among themselves (4.4 messages per user compared with only 0.4 in Tomsk2).

It is also noteworthy the asymmetry between sent and received messages when including messages by the coach. This is because of the scheduled messages by the coach who reached participants on a weekly basis but was not always reciprocated, as well as to the interaction behavior of the coach, that is, sending more than one messages per interaction.

Copresence in the Training

Participants in the interaction group were able to see each other, train together, and coordinate their participations. Participants in the individual group were not. Thus, copresence in the

Table 2.4 Mean (SD) messages exchanged among all users (including the coach) and only trainees.

	Tomsk1		Tomsk2	
	Interaction	Individual	Interaction	Individual
Private messages sent				
All users	8.4 (6)	8.1 (7)	4.3 (6)	5.7 (4)
Only trainees	4.4 (3)	Not applicable (N/A)	0.4 (1)	N/A
Private messages received				
All users	13.5 (2)	13.1 (7)	11.1 (3)	10.9 (1)
Only trainees	4.3 (2)	N/A	0.5 (1)	N/A
Public messages posted				
Trainees	0.6 (1)	N/A	0.5 (1)	N/A

individual group is only an indication of meetings by chance and used for comparisons. The copresence by study and group is shown in Figure 2.6.

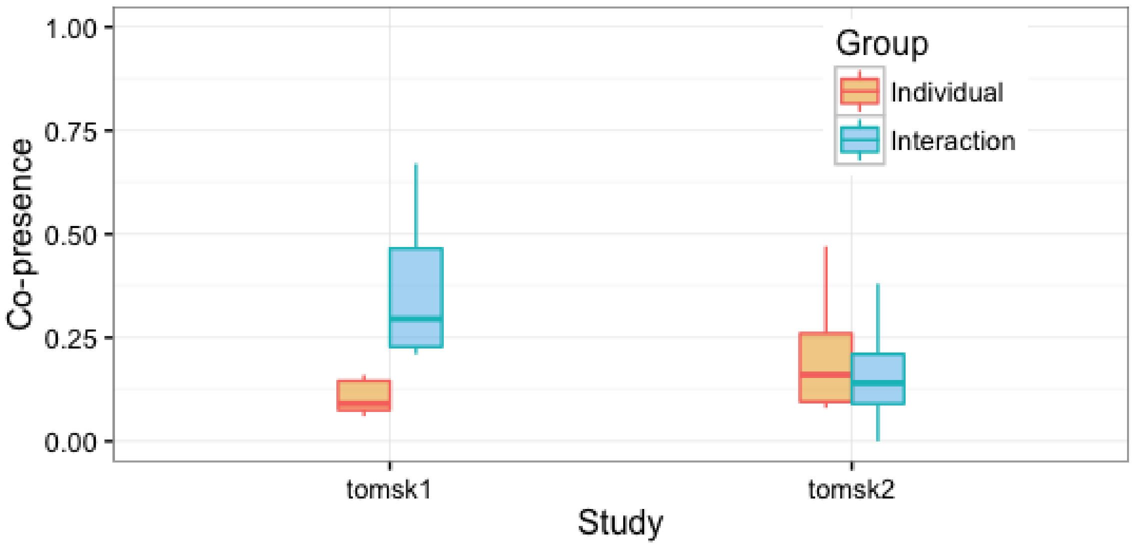


Fig. 2.6 Copresence by study and group.

The copresence in the Tomsk1 study was on average significantly higher in the interaction group: 36.25% (SD 17.25%) in comparison with 10.71% (SD 4.15%) for the individual group.

A t test showed a significant difference between the interaction and individual groups ($t(7.9)=-4.05$, $P=.004$) in favor of the group training condition.

In the Tomsk2 study, the copresence was of 16.38% (SD 11.44%) in average in the interaction group and 19.4% (SD 11.13%) in the individual group. A t test showed no significant difference between groups ($t(25.22)=0.7$, $P=.49$).

2.3.2 Program Adherence

The overall persistence rate was of 74% (SD 27%) when considering the number of sessions available in the 8 weeks of training. Breaking down this number by group treatment, we observe a persistence rate of 75% (SD 28%) for the individual groups and 74% (SD 26%) for the interaction groups, whereas the result by study shows a persistence rate of 82% (SD 24%) or Tomsk1 and 70% (SD 28%) for Tomsk2. In the study Tomsk1, the persistence rate was 77% (SD 25%) for the individual group and 87% (SD 23%) for the interaction group; in Tomsk2, it was 74% (SD 30%) for the individual group and 65% (SD 25%) for the interaction group.

An ANCOVA was performed to compare the persistence of participants of individual and interaction groups in the two studies while controlling for the initial baseline measures of leg muscle strength, social support, and PACES. The results show neither a significant main effect for group ($F(1,18)<1$, $P=.74$) or for study ($F(1,18)=1.46$, $P=.24$), nor interaction between study and group ($F(1,18)=1.15$, $P=.30$).

Considering the baseline measures, the results show a significant interaction between study and the initial social support score ($F1,18=5.23$, $P=.03$). As observed in Figure 2.7, part A, in Tomsk2, participants with higher social support level showed higher adherence to the training, whereas in Tomsk1, the adherence is not significantly associated with by the initial social support score. No significant effects were found for the initial scores of leg muscle strength.

The interaction between PACES score and group was also significant ($F(1,18)=6.001$, $P=.03$). As shown in Figure 2.7, in Tomsk2, participants with higher enjoyment of physical activity had a higher adherence level (Figure 2.7, part B), whereas in Tomsk1, enjoyment of physical activity had a negative effect on the interaction group (Figure 2.7, part C).

2.3.3 Well-Being Outcomes

Eight participants did not answer to one or more questions of the SWLS and thus, have been excluded by this analysis. On the subset of participants without missing answers, SWLS score was analyzed with a repeated measure ANOVA with time (pretest vs posttest) as within-subject factor, and group (individual vs interaction) and study (Tomsk1 vs Tomsk2) as between-subject factors. Only the main effect of time was significant ($F(1,31)=5.85$, $P=.02$).

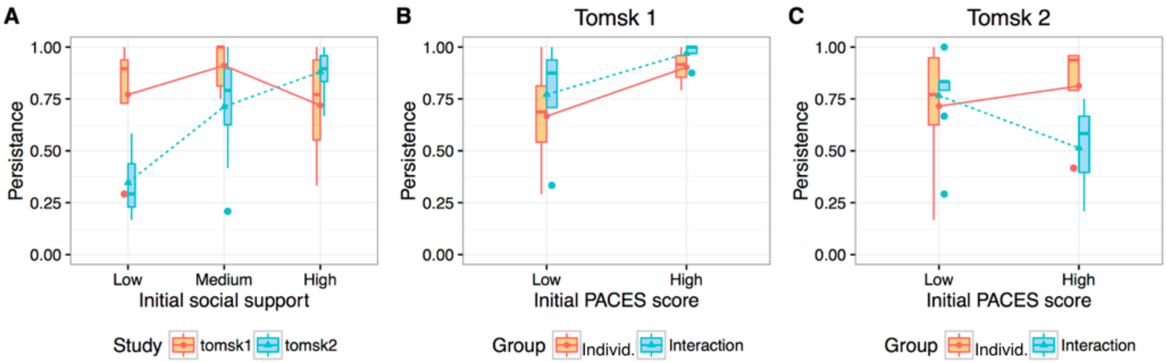


Fig. 2.7 Interaction plots for persistence and baseline measures. (A) Interaction between study and initial level of social support (medical outcomes survey, MOS score has been grouped in three equally distributed intervals: low, medium, and high). (B) Interaction between group and initial PACES score in Tomsk1. (C) Interaction between group and initial PACES score in Tomsk2.

Participants reported high satisfaction in the posttest questionnaire (mean 23.8 [SD 6.2]) compared with the pretest measures (mean 21.34 [SD 5.8]).

The same analysis was performed on R-UCLA Loneliness Scale. Eight participants were excluded from the analysis because of missing values in the pretest or posttest questionnaires. Only the main effect of study showed a tendency toward significance ($F(1,31)=3.55$, $P=.07$). Participants reported a lower level of loneliness in the Tomsk1 study (mean 4.77 [SD 1.7]) compared with the Tomsk2 study (mean 4 [SD 1]).

2.4 Discussion

2.4.1 Principal Findings

Online Group-Exercising Tool Rated as Highly Usable (Research Question 1)

Participants' rating on the usability of the app shows that the group exercise app (assigned to the interaction group) has a high usability and that the added complexity in relation to the more traditional home-based version (assigned to the individual group) did not significantly affect its usability.

When asked in detail, participants reported the training feature as very usable, whereas the messaging as usable but with a lower score (1 point lower), possibly because of the typing. The intention to use the app in the future was also very high, which along with the analysis of the actual usage, points to the feasibility of using the online group-exercising tool for training in a social context. These results are in line with a previous usability study and usage behavior analysis done on the Gymcentral tool [10].

Private Messages as Preferred Interaction Channel Among Trainees, Even in the Strong Cohesion Group (Research Question 1)

As in our previous study analyzing online interactions in a training context among Italian older adults [10], we expected to observe a higher usage of public messages for communication among trainees. Surprisingly, however, participants exchanged more private messages among themselves than public ones, even in the strong cohesion group. The high cohesion setting only accounted for more exchanges per user, not for group-level interactions. This result suggests different attitudes toward group interactions possibly because of cultural differences. In fact, the usage logs suggest mainly a lurking behavior, possibly because of the automatic sharing of the participant's performance—a social learning feature. Thus, further studies are required to design better online interaction tools that would motivate group building in the cultural context of reference.

Copresence Higher in the Strong Cohesion Group (Research Question 1)

The results of copresence show us that participants from the interaction group in Tomsk1 (strong cohesion group) participated in significantly more training sessions with the company of others compared with the meetings by chance in the individual group. We have seen the same effect in our previous study [51] featuring a high-cohesion group of Italian older adults. This effect was not observed in Tomsk2 (low cohesion group), suggesting that training together is not necessarily a preference in groups with low cohesion, and thus, the cohesion level might affect the willingness to train together.

Online Group Exercising Did Not Result in Higher Adherence When Compared With Individual Training With Persuasion Features (Research Question 2)

We have observed a higher adherence for the groups with high cohesion, and in particular, under the group-exercising treatment (interaction: mean 87% [SD 23%]; individual: mean 77% [SD 25%]). However, the ANCOVA showed neither a significant main effect for group or for study, nor interaction between study and group. This suggests that the added group exercising feature did not account for a significant difference in persistence rate compared with the individual training with persuasion features (interaction: mean 65% [SD 25%]; individual: mean 74% [SD 30%]).

In our previous study with Italian older adults [51], we observed a higher adherence to the online group-exercising compared with individual training (with no persuasion strategies). Here, we did not observe the same effect when comparing online group exercising with individual training (with persuasion strategies). We attribute this effect to (1) Persuasion features in the individual training condition that raised the adherence by 10% compared with our previous study [51]. This increase made the difference in favor of the group exercising condition non-significant and (2) Weaker cohesion among participants in Tomsk2, which might have reduced the effect of normative influence and peer support, resulting in a 20% drop in adherence compared with Tomsk1 and our previous study [51].

These results contribute to the ongoing discussion on the differences between individual and group training (see [26] for the most recent meta-analysis on the topic). First, it adds to the evidence that group-exercising in low cohesion groups results in an adherence comparable to that of individual training with contact (with a coach), extending the evidence to online settings. Second, it partially supports the evidence that group exercising in high-cohesion groups results in higher adherence than individual training with contact. On this point, we have seen evidence only when comparing group exercising with individual training with no persuasion strategies, which is indeed closer to the individual condition explored in [26]. The possibility of incorporating persuasion strategies in online setting adds a new dimension that requires further investigation.

Social Support Can Predict Adherence to a Training Program When Social Connections are Weak or Absent (Research Question 2)

In analyzing the effects of social support on adherence, we have seen a significant interaction between study and the initial social support score at baseline. In Tomsk2, participants with higher social support level showed higher adherence to the training. This suggests that higher level of social support is associated with higher levels of adherence when the connection among participants is weak (Tomsk2). This observation is in lines with the literature highlighting the social support structure as an important determinant of adherence [132, 46]. Interestingly, Tomsk1 did not show a significant association between initial social support and adherence. This suggests that low levels of external social support (as measured at baseline) can also be compensated with the social dynamics of an online group with strong cohesion (Tomsk1).

Enjoyment of Physical Activity With Contradicting Effects on Adherence for Groups With Weak and Strong Cohesion (Research Question 2)

Enjoyment of physical activity is described as determinant of physical activity [132, 46] and is associated with positive attitudes toward exercise, intrinsic motivation, and consequently long-lasting adherence to physical activity [176, 136]. We have seen, however, some conflicting effects of this variable—as measured with the PACES scale—on the adherence of the groups with weak and strong cohesion: all groups showed higher adherence for higher PACES score except for the interaction group with low cohesion that showed the opposite effect. This negative effect on adherence in the latter group came as a surprise, and it requires further study to investigate its roots and whether it is because of negative social dynamics in low cohesion settings.

Initial Level of Fitness With Nonsignificant Effect on Adherence of Online Group Exercising and Individual Training With Persuasion Strategies (Research Question 2)

Implementing group exercising can be challenging, especially in heterogeneous populations. Individual differences among older adults can lead to motivational issues and problems in tailoring the training [43]. In addition, perceived barriers such as lack of skills, pain, fear of injuries, and falls can also constitute obstacles to the motivation of older adults to exercise.

In our previous study with Italian older adults [10], we observed that the initial level of fitness could predict the adherence of older adults to an individual training (without persuasion strategies). It was also observed that the online group exercising tool—the same used in the pilots reported in this paper—was effective in mitigating that effect. In lines with this prior study, the results from our two pilots showed that the initial level of fitness did not have a significant effect on adherence of the interaction group but neither on the adherence of the individual group. One potential explanation is the presence of individual persuasion strategies in the version of the app used by the individual group, which might have leveled the effect. This suggests that more studies are needed to better understand the roots of the observed effects of the initial level of fitness, as well as the effects of individual and social persuasion in mitigating them.

Seasonal Fluctuations and Its Influence on Availability of Candidate Participants (Research Question 2)

Seasonal fluctuation has been found to determine the level of physical and social activities of older adults [129], especially in high latitudes where winter can result in a decline of physical functions of older adults [19]. In Siberia, these fluctuations greatly influence the activities of the daily living and the opportunities to engage in activities in general.

Although our studies were set in spring and summer periods, we did experiment the effects of the seasonal fluctuation but at recruitment and for quite the opposite reasons. June to September is gardening season, and independent living older adults usually engage in this activity, spending most of the period in their summer houses (Dacha). This influenced the availability of participants in our study as it created obstacles for some candidates that showed initial interest in participating (eg, finding time to train and worries of bringing tablets with them outdoors or to the Dacha). After this experience, the second study was moved to earlier spring months (April-June) to increase the pool of potential candidates. However, we did not see a significant difference in the program adherence that could be explained by these two different seasons. Further studies are needed, especially to understand the effects of the extreme winter season.

Increase in Life Satisfaction as a Result of the Training, Regardless of the Version of the App (Research Question 3)

Recent history, along with current social, political, and economical factors have impacted negatively in life satisfaction and happiness of older adults in the Russian Federation[2]. Thus, devising and studying solutions aiming increasing the happiness and well-being of older adults in this region is of paramount importance.

In investigating the impact of physical training, we have seen an overall improvement in the SWLS score for all participants, regardless of the version of the tool used. This is consistent with our previous study with Italian older adults [10], where we observed an improvement in the subjective well-being of the participants regardless of being part of the individual or group condition. Furthermore, these results are in line with previous literature on the benefits of physical activity on the QoL of older adults [107, 172], and contribute with additional evidence in favor of technology-supported interventions and their benefit for older adults in the Siberian region.

No Significant Decrease in Loneliness, Despite Social Features (Research Question 3)

Participants did not observe any decrease in the loneliness score as a result of the training, not even those in the online group exercise condition. This is contrary to our expectations, given the social context provided by the group-exercising and the social interaction features. In Trento, Italy [10], we did observe a significant decrease in the loneliness score, but compared with this study, the usage of social interaction tools and adherence to the training was much higher. This difference in the usage of social interaction features, possibly because of cultural differences as reported earlier, could have limited the effectiveness of the medium.

2.4.2 Limitations

Gender Imbalance

The lifespan gap between males and females in the Siberian region is one of the biggest in the world: life expectancy at birth for men is 64.7 years, whereas for women it is 76.3 years [127]. These demographics limit the availability of male candidates in the senior citizen organizations, and therefore, our ability to recruit more male participants. However, previous studies suggest that male and female participants may have the same reactions to sport activities despite differences in motives to participation [136, 92]. Still, further studies are needed to see if these observations can be translated to the intervention described in this paper.

Group Size Difference

The amount of participants in the Tomsk2 study was twice bigger than in the first Tomsk1 study, 40 and 20 participants, respectively.

The difference in the group size between the two studies is because of (1) the complexity of the study design and (2) the difficulty in finding participants of older age willing to participate, given the specific social characteristics of the region (older adults living in Siberia are not used to participate in studies). Therefore, we were able to involve only 20 participants for the study Tomsk 1. The following year, as we built better contacts with various retirement organizations and local organizations became more familiar with the project, we were able to involve 40 people in the study (Tomsk 2).

No Quantitative Measures of Group Cohesion

Group cohesion was defined as a property of the pool of candidates: participants acquainted with each other and engaging in joint activities. This property was maintained during randomization by ensuring that pairs of friends would end up in the same groups. While being a solid definition, the fact that cohesion was not qualitatively measured should be noted as a limitation.

Scales Validation in Russian Language

There is a lack of translations of international standardized measure in Russia. Therefore, except the SWLS (which has already been validate in Russian language), no translation was available for the measures used in the study. These measures were translated and adapted to Russian language and culture by our research group by using the standard translation or bask-translation procedure. During this procedure, we ensured to reach semantic, idiomatic, and conceptual equivalence between the original English and final Russian versions.

Although, without a validation study, we cannot be completely sure that these instruments fully fit the socioeconomic characteristics of Siberia, we believe that the standard procedure adopted to translate these instruments provided reliable results. This should be considered as the limitation of the study.

Validity of the System Usability Scale

Two questions were excluded from the SUS because in the pretest of the prefinal version of the scale (during the translation or back-translation procedure), older adults found it difficult to understand them ("I found the various functions in this system were well integrated" and "I thought there was too much inconsistency in this system"). Therefore, whereas in the original scale the total SUS score ranged from 0 to 100, in our study it ranged from 0 to 80. This is a limitation of our study and could make it difficult to interpret the usability results. However, it is worth noting that no usability scale suitable for older adults existed in Russian language, and our study provides the first adaptation for this culture. Future studies should investigate the validity of this short version of the SUS.

2.4.3 Conclusions

The results point to the feasibility and effectiveness of technology-supported physical interventions, and in particular, of online group exercising among Siberian older adults. High cohesion groups are preferable for group exercising, especially to mitigate effects of low social support on adherence. Cultural differences might explain the preference of private messages over public ones. Results in terms of subjective well-being are promising, but enabling interaction has proved not to be enough to observe a decrease in loneliness. Thus, further research is needed to understand how to better enable community-building interactions.

Chapter 3

Fitness applications for home-based training

3.1 Introduction

Physical activity, especially in the form of structured exercises, not only helps to improve physical function, but has also been linked to positive outcomes in social and mental well-being [1]. However, for several groups of people (such as older adults with physical and cognitive limitations, or postpartum women), regular training - and especially regular training at gym or outdoor - may be inconvenient or impossible.

In this chapter, we review how technology can (and does) facilitate training from home, how it can motivate people to begin and maintain an active lifestyle, and how it can be effective in achieving results (such as better strength and balance). Because older adults represent such a specific and important class of people for which home training may be the most convenient (and sometimes only) option, we specifically analyze research and applications based on their suitability for older adults. Besides discussing current technologies and research, we also underline limitations and research gaps in IT-based home training solutions in general and for older adults in particular.

3.1.1 Home-based fitness applications

With the purpose of analyzing home fitness apps in practice and revealing any emerging classes of applications, we set to analyze the type of support that is currently implemented in commercial fitness applications. In what follows, we describe the selection criteria, the design dimensions and literature considered in the analysis, and the emerging application archetypes.

3.2 Methods. Selection strategy

We screened 524 of the most popular health and fitness applications in the app stores for the following platforms: Android and iOS (mobile, 167 apps; category: health & fitness), Windows and Mac (desktop, 167 apps; category: health & fitness), and Nintendo and Xbox (gaming console, 190 apps; category: music & fitness / kinect). The selection was done based on the implementation of the notion of "most popular" that each app store supports for ranking apps, typically based on number of downloads and of active users (Android: "Top Apps"; iOS: "Popular apps"; Windows: "Top Apps"; Mac: "Popular apps"; Xbox and Wii: all games screened) in Italy as of June 2015, and includes both free and paid applications. We focused on this set of applications as it represents the solutions that users are more exposed to, and which have more visibility.

From this initial set, we excluded i) the applications which were not related to fitness, and ii) older and free-tier versions of apps already evaluated as part of the same list (to avoid duplicates). As a result, **we included 200 fitness applications** (100 mobile apps, 60 desktop apps and 40 console apps) in our analysis, coded by two experts with an inter-coder agreement of 94%.

3.2.1 Design dimensions

The effectiveness of home-based training programs for older adults has been the subject of recent reviews [2, 3, 4]. Chase [3] evaluated physical activity interventions for older adults and demonstrated that interventions do not need to be face-to-face to be effective. Müller et al. [2], in a review of interventions with and without technology, determined that internet-based interventions can be also effective and economically viable. The lessons learned from the literature and our exploration of current fitness applications highlight a set of common design aspects:

- ***Interaction Design*** refers to the technology (software and hardware) used to deliver the training program and to interact with the fitness application.
- ***Coaching and tailoring*** refers instead to the type of instructions, feedback and assessment that are given throughout the training, and how it is customized to fit the trainees needs and abilities.
- ***Monitoring and sensing*** denotes the mechanisms employed to measure performance indicators relevant to the training program.

- ***Persuasion and motivation*** discusses instead, how the various applications and devices encourage trainees to start and continue exercising.

We take the findings from these previous meta-reviews, along with the relevant literature from the HCI community for each of the dimensions under consideration, to derive design recommendations and analyse their adoption in current fitness applications.

3.2.2 Application archetypes

Three general classes of applications emerged when looking at the type of support implemented for each of the design dimensions, each focusing on a specific aspect:

- ***Training apps*** The distinctive feature of this class of applications is exercise prescription. Not all training apps support monitoring or feedback, but they all provide explicit training programs.

- ***Tracking apps*** Applications that don't offer training programs, but rather focus on tracking various aspects of user activity (e.g., steps, distance, elevation; 88%) and physiological indicators (e.g., heart rate, respiration; 44%) are included in this category.

- ***Fitness games*** These apps involve physical activities in a game context, without necessarily following a training program (only 35% have a training program). A distinctive aspect is the use of competition, comparison and cooperation persuasion strategies (91%), which generally is the highest among the three classes of apps.

These archetypes are illustrated in **Figure 1**, and discussed in detail in the following sections. For each archetype, we particularly analyse its specific focus area.

3.3 Dimensions

3.3.1 Interaction Design

The special abilities of older adults, along with level of education and access to technology, are of particular importance in the design of technology-based physical activity interventions. As discussed by Müller et al. [2], the effectiveness of such interventions is ultimately related to the ability of the older adult to follow the training program using the technological instrument, thus calling for a better understanding of the underlying components of the interactive fitness applications.

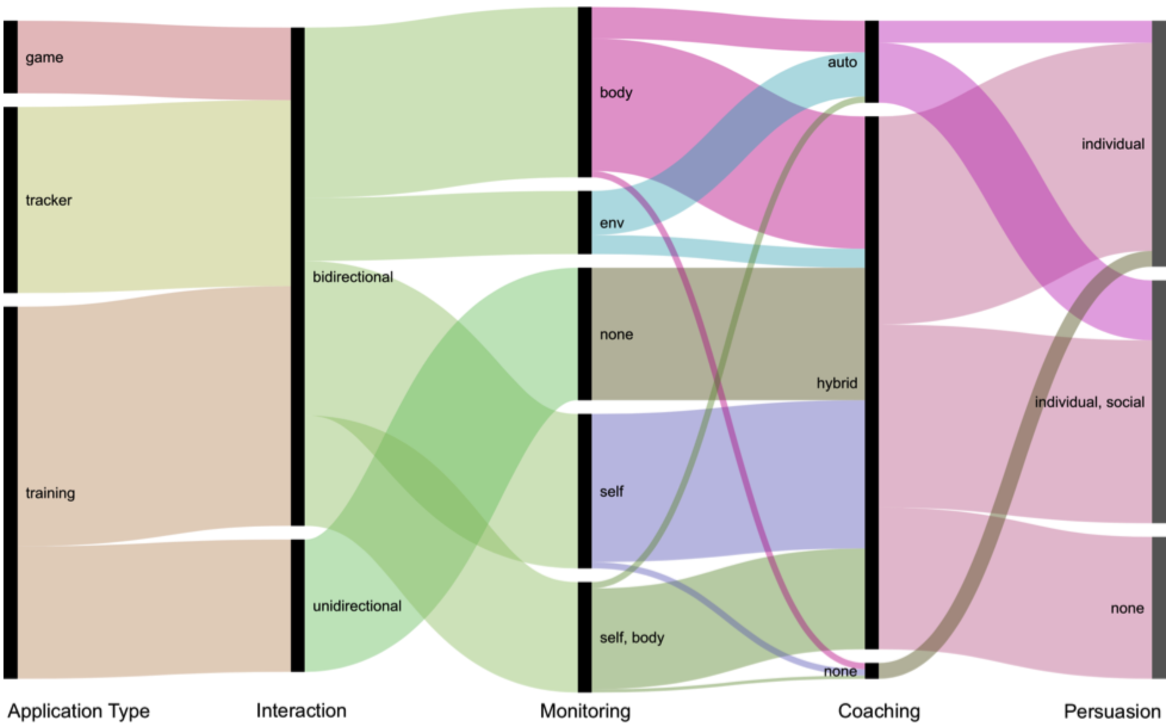


Fig. 3.1 Alluvial chart illustrating common patterns in fitness application (archetypes)

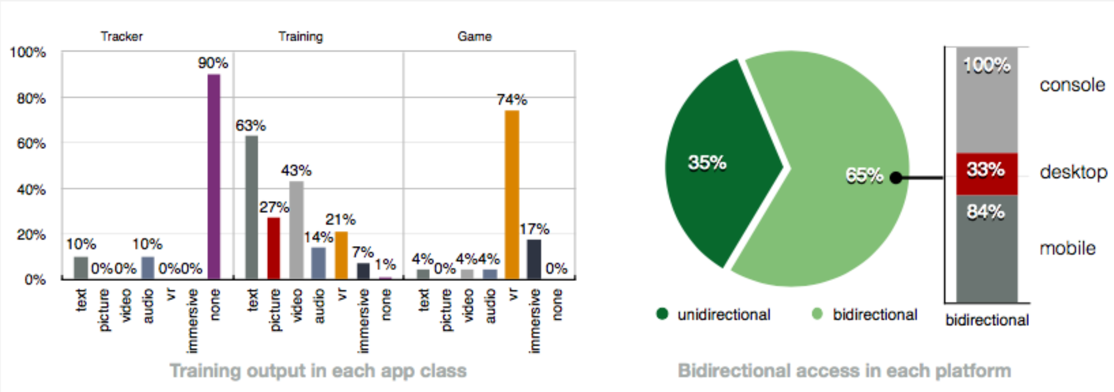
With the notion of design guidelines discussed in the literature, and the top rated apps in the online stores, we identified three different aspects to the design of interactive training applications:

- ***Direction of the interaction:*** "unidirectional" denotes training programs in which information related to training flows only from the application to the user (the application does not measure user activity in any way) and "bidirectional" denotes applications in which information related to training flows in both directions (e.g., the system prescribes the exercises and the users report - manually or via sensors - about their training activity to the the application).
- ***Input type:*** describes how the user interacts with the medium: this is indirect when the user action needs to be translated to provide the input (e.g., using a mouse pointing device); direct, when the action does not need translation (e.g., using a touch-enabled device); natural, when the input components are invisible and the interaction happens by using natural gestures (e.g., posture recognition in MS Kinect); and
- ***Training output:*** refers to how the training is represented (text, illustrations, audio, video and virtual and immersive environments).

Exercise programs delivered via workout DVDs, represent a large percentage of desktop apps (67%) and provide unidirectional access to training programs. In this setting, trainees have access to static exercise instructions with no feedback (from the medium) on their performance. Despite this limitation, in a study with 237 community-dwelling older adults, this class of solutions demonstrated to produce meaningful gains in physical function [5]. On the other hand, training applications in mobile and console platforms rely more prominently on bidirectional access, providing not only training instructions but also the possibility of logging activities and reflecting on training performance (mobile 84%, console 100%).

Research on human computer interaction points to direct input, like the one provided by touch-enabled mobile devices as being more accessible for older adults [6, 7], compared to indirect input found in most desktop applications which rely on mouse and keyboard. Indeed, touch-enabled applications designed especially for older adults have been shown to work in remote training settings [8]. Applications in game consoles, instead, rely on sensors such as the MS Kinect (https://www.microsoft.com/en-us/ kinectforwindows/) and the Wii Remote (http://www.nintendo.com/wiiu/) that offer natural input capabilities. A study by Pham [9] has shown that older adults interacting with the mixed controllers (gestures and buttons) found in Nintendo Wii, require less learning time and perform better, compared to the ones based on gesture-recognition-only found in Xbox Kinect. Nonetheless, the same study reports a preference of older adults towards gesture-recognition controllers due to perceived benefit in performing more physical movements.

In terms of output, there are no formal studies on which representation is more effective. Aalbers et al. [4] discusses this aspect further and concludes that it is not clear what mechanism of delivery can be considered more effective. However we understand from research in multimodal interfaces that combination of formats is preferable (e.g., combinations of visual and audio, or visual and haptic), especially when they compensate declines in perception skills [10, 7]. In fact, mobile fitness applications designed with the specific goal of facilitating remote training do in fact use a combination, with text (73%) and videos (53%) as dominant formats. Instead, applications in game platforms, due to the tracking capability of the sensors, deliver training instructions via virtual and immersive environments. A summary of design considerations and current practices is shown in Figure 3.2.



Design considerations

- Unidirectional training programs can be effective, however there might be differences in adherence (see persuasion and motivation).
- Direct input devices are more accessible to older adults [95, 96].
- Combination of training output formats can help to compensate declines in perception skills [98, 96].

Current trends of gaps

- Training applications offer the most diverse output with text (63%) and video (43%) at the top. Games prominently use virtual (74%) and immersive (17%) environments. Most trackers do not offer training programs (90%).
- Training apps, primarily desktop (67%), are notoriously unidirectional in that the program does not require the user to provide feedback to function.Mobile and console games are often bidirectional, partly due to the use of the sensing capabilities of these platforms.

Fig. 3.2 Interaction design in fitness applications

3.3.2 Sensing and Monitoring

Different types of instruments can be used to capture relevant training data, and at a high-level we can describe them in terms of:

- **The sensing method:** referring to *how* the data is collected, from self-reported data to specialized sensors (wearable or environmental), and

- **The aspect observed:** referring to *what* is being collected, e.g., general activity, physiological indicators or detailed motion patterns.

The choice for the instrument typically depends on the type of activity to be performed (e.g., indoor, outdoor), the aspects to be measured and the level of accuracy needed [177].

Self-reported questionnaires can be used to inquire trainees about their performance and adherence to the training and also their overall physical activity and well-being. Many applications (47% of trackers and training apps) rely on this instrument given its ability to capture training-related data without the need of specialized sensors. Self report is also used in research trials due to its ability to easily collect data from a large number of people without affecting participant's behavior during the experiment [141]. However, self reporting is time consuming and can be a complex cognitive task, especially for target groups with memory limitations such as older adults [141], which could lead to misreporting [117]. In addition, from the trainee's perspective, entering health data manually can lead to a decline in the usage of the application [4].

Sensing technologies have advanced to the point that we can wear and use devices with sophisticated capabilities. Wearable sensors, and other types of body-fixed sensors, can now measure indicators such as general activity level (e.g., Fitibit: http://www.fitbit.com/, Nike Fuelband: http://www.nike.com/us/en_us/c/nikeplus-fuel, Misfit: http://misfit.com/, Striiv: http://www.striiv.com/), quality of sleep (e.g., Fitbit, jawbone: https://jawbone.com), heart-rate (e.g., Polar FT4: http://www.polar.com/en/products/get_active/fitness_crosstraining/FT4), and even breathing quality (Spire: http://spire.io), enabling advanced monitoring capabilities. Such objective measures are desirable to get a more precise picture about the progress in non face-to-face interventions [117]. These sensors usually come with a companion application as well as programmatic interfaces (API) that enable their integration with third party systems. Moreover, built-in sensors (activity tracker and heart-beat) embedded into smart watches (e.g., Android Wear: https://www.android.com/wear/ and Apple Watch: https://www.apple.com/watch/) and on top of wearable operating systems enable developers to add sensing to their apps. Yet, only 20% of the training apps and 90% of trackers which we have analyzed support integration with wearable or built-in sensors and the remaining rely on self-report data. In terms of perception of these sensing technologies, a two-week study with 8 older adults reported no usability issues but a negative change in the attitude (in 5 participants) due to accuracy limitations in measuring some daily activities (e.g., walking in a treadmill), being uncomfortable to wear, and being considered a waste of time [53]. Indeed, the level of accuracy of these devices might not render them appropriate for all scenarios (e.g., clinical trials), as demonstrated in a study [170] with pedometers.

Environmental sensors and advanced motion sensing devices such as MS Kinect and Nintendo Wii Remote enable more advanced performance tracking capabilities and also better suits for trainee's at home. A central aspect of these capabilities is the body motion tracking, which its reliability and accuracy has been demonstrated in a research experiment [165]. Environmental sensors can also measure physiological data. For example, the new MS Kinect provides touch-free heart-rate measurement (by scanning the skin surface of the trainee), with an accuracy within a few beats per minute (http://support.xbox.com/en-US/xbox-one/games/xbox-fitness-faq) under good conditions.

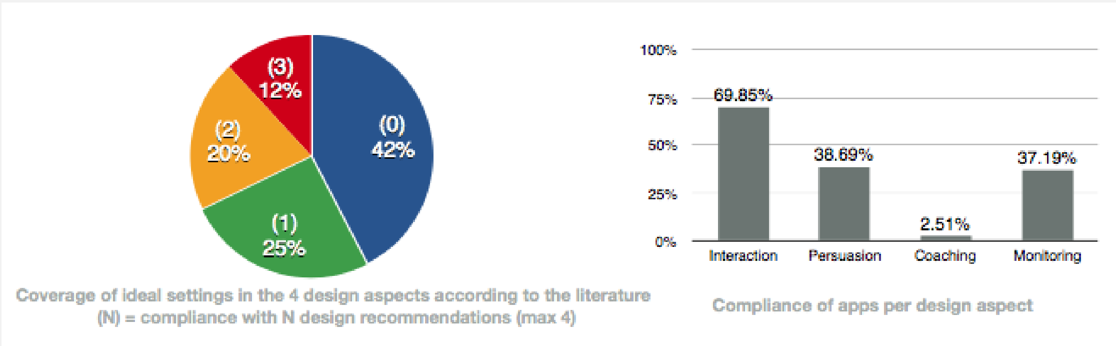
A summary of design considerations and current practices is shown in Figure 3.3.

3.3.3 Coaching and Tailoring

The *coaching* process is commonly described by a series of phases before, during and after the training: i) identifying the needs, abilities, desires and goals of trainees, ii) prescribing a tailored training plan, iii) providing support by monitoring the progress of trainees, and iv) modifying the training plan accordingly [125]. Technology can provide different levels of support in this process, from entirely human to fully automatic (virtual) coaching. For the trainees at home, studies have shown that coaching, either by a human or virtual coach, not only makes the training more effective and safe, but also more engaging [80, 126].

Support for coaching begins with solutions where coaching is provided by a human, and technology essentially acts as a communication tool. An example of such setting was experimented [58] with low-income postnatal women, where SMS messages were used to deliver tailored instructions and to get feedback, resulting in prolonging the duration of physical activity of the target trainees. Indeed, in a review of physical interventions by Muller et al. [117], such low-tech solutions are considered as valid alternatives to increase physical activity in low income older adults. The human coach does not only provide information and feedback, but also exercise knowledge, encouragement and emotional support while the trainee goes through exercising sessions [37].

Technology can also assist the human coach in monitoring and tailoring the training program. Fitness applications nowadays (e.g., Fitbit, Nike+) come with sensors or feedback mechanisms that facilitate the monitoring by a coach or the trainee itself. Tailoring training programs are also supported by dashboards that assist human coaches in tuning the plans based on trainees performance (e.g., http://gymcentral.net). This aspect has shown to be particularly important in a review of 12 internet-based interventions [1], where interventions with tailored information have resulted in lower attrition rate per month (2.7%) compared to those with generic information (6.6%). Furthermore, we should note that the role of the Coach can be played by the trainee itself or an expert. In our analysis, the human factor in



Design considerations

- Self-reports can be used to collect data from users without sophisticated sensing technology. However it is time-consuming and can be a complex cognitive task, especially for target groups with memory limitations such as older adults [100].
- Long term usage of sensing devices can be discouraged by limitations in measuring daily activities of older adults (e.g., walking on a treadmill) and being uncomfortable to wear [102].
- Environmental sensors are effective and can render the interaction more natural [104].

Current trends and gaps

- Body-fixed sensors (wearable) are widely used in trackers (90%), while self-reporting is more popular in training apps (47%). In games, we see body-fixed (57%) as well as environmental (43%) sensors.
- Trackers support mostly training activity (activity, 88%) and physiological data (bio, 45%). In training apps, tracking is more diverse but attendance logging is the most popular (attendance, 40%). Console games extensively use motion sensors to track actual movements (motion, 96%).

Fig. 3.3 Sensing and monitoring in fitness applications the coaching mechanism was the trainee itself in most of the cases. Only 3% of the training applications provided an expert support (a human expert except the trainee itself).

We discovered that a few applications (2%) rely on “virtual coaches” instead of human coaches. Virtual coaches are pre-programmed or smart machines and applications that monitor, prescribe and tailor the training program for the trainees [160]. An example of such solution was experimented by Steffen et al. [147], who developed a personalized exercise trainer for the elderly using a wearable sensor that detects the movement of the user while exercising, automatically tuning the exercise level, and providing audio feedbacks during the exercise. The authors of this study however did not formally evaluated the system but report on positive feedback from the older adults. More formal studies, such as Watson et al. [178] compared the effect of a fitness application with virtual coaching, with respect to applications without coaching, and concluded that the trainees with a virtual coach adhered longer to the training program. However, coaching in this form does not provide the social support that a human coach can provide. Indeed, studies express the need for a real coach [37] , especially when dealing with sensitive trainees such as older adults [67].

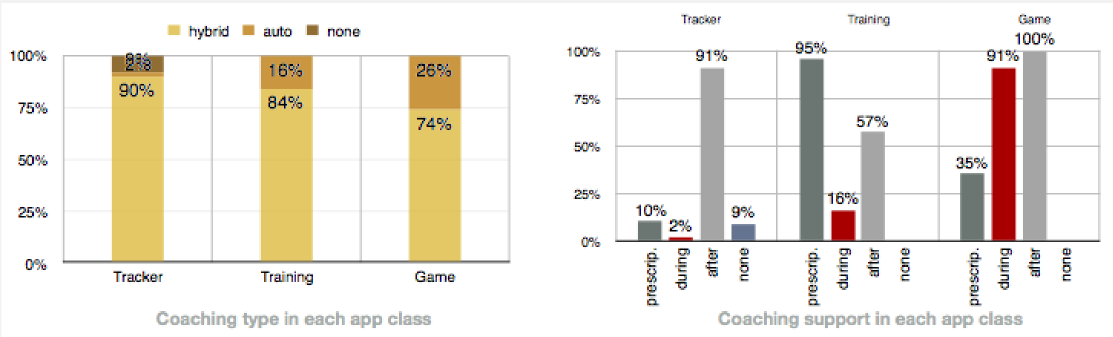
To cope with the aforementioned shortcomings and barriers Thórisson [160] , experimented with “Reactive Virtual Trainer” which, unlike the conventional virtual trainer, is creative and provides emotional and psychological support to the trainees similar to a human coach and tailors the fitness program according to the physical and emotional states of the trainees. However, Thórisson proposes that although the RVT is pre-programmed by a human expert, yet it can not substitute a human coach in critical cases since the precision of such technologies is not accurate enough and longitudinal user studies is required to measure the long term effect of RVT in a training program.

A summary of design considerations and current practices is shown in Figure 3.4.

3.3.4 Persuasive Technologies

Programs aiming at promoting physical activity and lifestyle changes incorporate components aiming at increasing adherence and reducing attrition [117, 1]. In technology-based interventions such components can be described in terms of persuasion strategies. Persuasive strategies for home-based training can be grouped in two major categories: *i) individual*, referring to strategies that leverage the individual wills and natural drive and, *ii) social*, referring to strategies that demand the presence of a community of people with the roles of family, supporters and peer trainees [146]. We use this classification to describe a subset of strategies from the work of BJ Fogg [59] that are applicable to home-based training applications:

Individual Persuasion Strategies



Design considerations

- Coaching, either by a human or virtual coach, is preferable over a solution without coaching [105, 38]
- For sensitive training applications and for the older adults population, the presence of a human expert in the loop is recommended [110].
- Reactive Virtual trainer can provide emotional and psychological support along with the exercise instructions but are not recommended for exercises that require high precision [108]

Current trends and gaps

- Most apps use hybrid coaching (90%), especially covering the prescription and after training phases. Apps in game consoles tend to provide support in real time during the training (91%).
- Training apps in game consoles provide more complete support than mobile apps.

Fig. 3.4 Coaching in fitness applications

- **Reminders and suggestions:** The application reminds trainees of their exercises sessions and suggest better exercising habits.
- **Positive & negative reinforcement:** The application prompts positive or negative comments about the exercising behavior of the trainee to raise awareness.
- **Self-monitoring:** The application provides to trainees, performance monitoring and awareness about their current progress.
- **Rewards:** The application praises the trainee by providing virtual badges, medals, awards and recognition upon completion or success on exercising sessions.

Social Persuasion Strategies

- **Social learning (comparison), cooperation and competition:** The fitness application provides social features that allow trainees to compare their performance with others, collaborate toward common goals, or compete.
- **Social support:** The application provides social features such as messaging and forums enabling trainees to interact with each other and create a community of people supporting each other.
- **Recognition:** The trainees individually or in a group get public recognition on their awards, progress and contributions.

Individual persuasion strategies have been evaluated in several studies (e.g., [146, 139]). For instance, Rodriguez et al. [139] designed a mobile app for the elderly that persuade them to walk. The system incorporates *reminders, notifications, self-monitoring* and a "coin metaphor" as *rewards* for the users. The system rewards the trainee's walking activity with "virtual coins" and their social interaction with "virtual diamonds". Findings after an intervention with older adults reveals that all of them found *self-monitoring* and the reward metaphor motivating to exercise regularly. However, some participants argued that they would have preferred other types of metaphors than "virtual coin" for instance. In fact, researchers (e.g. [81]) emphasis that the heterogeneity in the older population require persuasive strategies, and in particular individual persuasion strategies, to be tailored according to older adults abilities, interests and taste.

However, beyond personalization and usability issues, several studies report on higher level of interest in social persuasion strategies and report that in particular older adults prefer apps that leverage their social activities [81, 24]. Silveira et al. [146] experimented the effectiveness of individual and social persuasion strategies in two groups of individual trainees and social trainees. The result of the study reveals that the long-term adherence was higher in the social group. Moreover, the follow up of the research and further experimentations with individual and social trainees discovered a raise in exercising duration for the social training group.

Despite the evidence in the literature, individual persuasion strategies are more extensively adopted than social strategies, especially with trackers (98% individual, 46% social) and training apps (63% individual, 25% social). A summary of design considerations and current practices is shown in Figure 3.5.

3.4 Findings and Opportunities

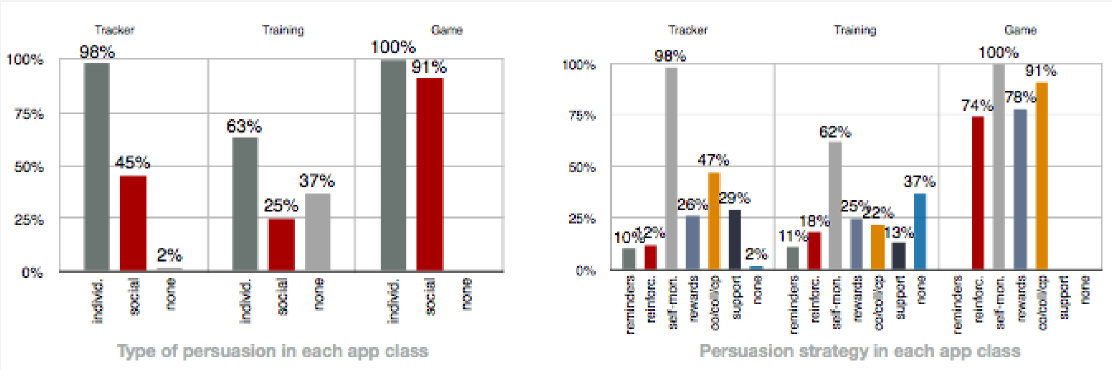
This investigation has shown that home fitness landscape is sprouting with ideas and applications which offers a variety of interaction modes, coaching methods and, measuring techniques. Current solutions provide very good support for the general population, especially for those that do not require expert coaching. However, there are very few applications that explicitly target older adults and are designed for the average interaction skills and physical abilities of the average adult in their late life. This is the case despite the ample evidence of the positive effect of sustained training on mental and physical health and in turn, the benefit that better health has in terms of reduced expenditure by government and families [122].

Thus, sensible groups such as older adults find more limited support in current solutions. This is evidenced in Figure 3.6 and the description below, where we summarized the result of inspecting current applications against the various design considerations motivated in this paper.

The principal lessons learned are specific ingredients that stimulate higher level of engagement and of adherence to training programs:

- The provision for social persuasion mechanisms in addition to individual persuasions.
- The presence of a human coach (as opposed to virtual or no coaches).
- The adoption of sensors that enable automated detection of activity (as opposed to manual data entry), but only as long as the sensors are accurate and do not generate lack of trust in the user as to the reliability of the measures.
- The provision of a multi-modal interaction with the user.

From existing research it is however still unclear which UI representation and metaphors works better in terms of stimulating adherence, making this a topic of further research.



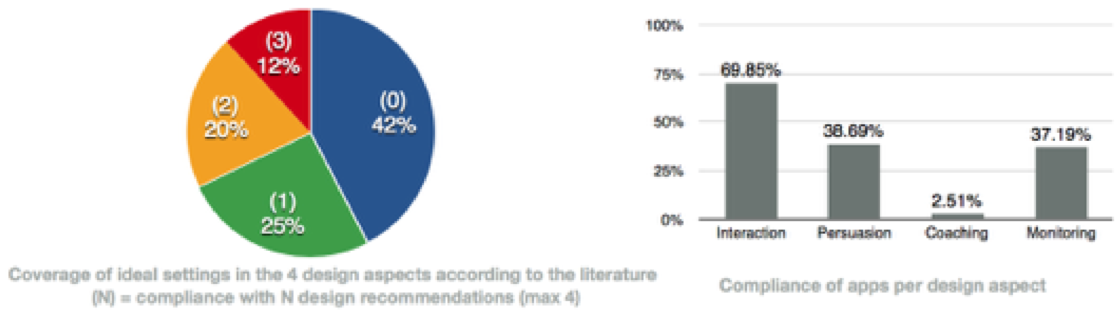
Design considerations

- Individual persuasion strategies, such as self monitoring, are effective in persuading the trainees to adhere to the training plan [70, 111].
- The use of social persuasion strategies can increase adherence of older adults to home-based training programs [70].
- Designing persuasive strategies for older adults requires tailoring and personalization of the strategy, as older adults are heterogeneous and have different taste and abilities [24].

Current trends and gaps

- Except for games (90%), social strategies are just a fraction (25% training apps, 45% tracker apps) of individual persuasion strategies.
- Self-monitoring is the most popular individual persuasion strategy (86%), followed by rewards (43%) and reinforcement (31%).
- Comparison, cooperation and competition (53%) are the social strategies more popular, followed by social support (14%).

Fig. 3.5 Persuasion in fitness applications



Observations

- There are no apps that comply with the design considerations from the literature in the 4 design aspects we review in this paper.
- The most limiting aspects are the very few solutions that provide expert feedback during training.

Summary of gaps

Interaction The suitability of fitness application is in part influenced by the platform. Mobile and game consoles provide the input and output mechanisms that can facilitate interaction. Indeed, we have seen a higher adoption of bidirectional and multimodal interfaces in those platforms. However, how the applications use these ingredients to implement usable interfaces is still a topic for further research.

Sensing Automatic detection of activities are preferable for older adults, especially when it comes to objective measures. Still, self-reporting is widely used as a method for data collection, despite potential effects in adoption and precision of readings. On the other hand, integration of sensors in fitness applications is still limited, with the exception of applications in game consoles that offer motion tracking capabilities during training.

Coaching Training programs supervised by an expert human coach are preferable for older adults. However, most of the applications on the outlined markets rely on hybrid solutions (technology + self coaching), focusing mainly on exercise prescription, while only 3% of the analysed training applications provide expert human coach support.

Persuasion The few studies on persuasion strategies to boost adherence, especially in older adults, show strong evidence on the benefits of using persuasion strategies, and especially social strategies. However, very few of the training application exploit these features, other than allowing users to reflect on their activities. In general, we see this as a clear opportunity for applications that exploit these aspects.

Fig. 3.6 Summary of compliance of fitness apps with design considerations

Chapter 4

What Makes People Bond?: A Study on Social Interactions and Common Life Points on Facebook

4.1 Introduction

Being socially connected can have a significant impact on the quality of life of older adults. Research has demonstrated the association between health risks and the lack of social network diversity, infrequent contact with network members, and the small size of social networks [17, 25, 144].

Social integration with peers is particularly important for older adults transitioning to residential care. Social integration helps in the adaptation, can foster friendships and sense of belonging, and has been found to be one of the key elements contributing to the quality of life in residential care [22]. Instead, failing to socially integrate contributes to feelings of loneliness, boredom, and helplessness, which are commonly regarded as the plagues of nursing home life [158].

The research and practice on technology-supported social interactions in this context has mainly focused on *enabling* social interactions (see, e.g., [32, 40] for a review), and less in addressing non-technological barriers, motivating social interactions and creating bonding. Addressing this gap requires the study and development of solutions that take into account the users' needs, motivations and barriers.

In the previous work done by members of our research group [9] was reported the results from surveys and visits to nursing homes. We identified that i) friendships in nursing homes are difficult, especially in the transition period, and that ii) contact is rather infrequent between older adults and their relatives, especially younger adults, often due to the lack of common topics of conversation and the lack of time. We suggested that technologies should go beyond *enabling* interaction, to aim at creating friendships between people and opportunities for meaningful conversations.

In this Chapter we follow up on these initial results and report on an exploratory study trying to understand the relationship between *connectedness* among friends, *social interactions* and *common life points* on Facebook.

The goal of this study was to understand if, by looking at information of the kind available in people's Facebook profiles and posts, we can predict the feeling of connectedness between two Facebook friends and the intensity of their face-to-face interactions. Specifically, we investigate the following research questions:

RQ1. To what extent can we predict, by looking at profile information on Facebook, the frequency of online and offline communication between two persons? We are interested in understanding if common life points and social interactions are related, and whether certain common aspects can trigger interactions.

RQ2. To what extent can we predict, by looking at profile information and intensity of social interactions, the feeling of connectedness between two persons? This question is fundamental as it will help us understand whether having common aspects and a certain level of interaction is related to connectedness. Connectedness in this context represents the possibility of creating long-term bonds and friendship.

We explore the above questions in the broad population of Facebook users, from younger (18+) to older adults (65+), since we are interested in *intergenerational* as well as in *peer* friend relationships.

In what follows we detail on the motivations, methods and results.

4.2 Background

4.2.1 Technologies to reduce social isolation

Extensive work has been devoted to interventions aiming to reduce social isolation with the help of technology (e.g., [32, 40] for a review). Technology used to enable interactions for older adults include internet and email (e.g., [20]), social networks (e.g., [11]), video chats (e.g., [153]), virtual companions (e.g., [106]), and phone calls (e.g., [31]). Most one-to-one interventions limit the contact to a predefined person, such as a trained interviewer, a trained helper, or a volunteer [31, 106]. However, interventions enabling social interactions with relatives and friends are more common in recent literature [11, 153]. Interactions between participants and new people are also explored in some interventions [60], in particular in those studying the effect of general internet use and social networks [20, 153].

Research on online social interaction with older adults has focused more on “enabling" communication and sharing, and less on creating opportunities for these interactions to happen. This calls for the development of technology that looks into making these interactions more effective.

4.2.2 Studies on friendship and common life points

The notion that similarities among people lead to creating ties between them is known as homophily [108]. In a review, McPherson et al. [108] described it as *“the principle that a contact between similar people occurs at a higher rate than among dissimilar people"*.

Homophily can be defined from two perspectives: i) *value homophily*, which is based on the attitudes, beliefs and values, and ii) *status homophily*, which is based on the major demographic dimensions such as race, ethnicity, sex, age, and characteristics like religion, education, occupation [100, 108]. A review of studies done by Fehr [54] suggests that both status and value homophily are relevant for building friendship. However, a recent survey by Campbell shows that only value homophily affects friendship chemistry (emotional and psychological connection between persons) [30].

There are studies analysing structural properties of friend networks [164] and empirical studies that have explored homophily in social networks. Kwak [96] studied homophily among Twitter users (with 1000 and less followers) and their friends-followers and found the effect for geographic location and popularity. Lewis et al. [103] studied Facebook profiles of 1640 college students in the US and found significant shared interests (movies, music, books) for certain connections (being Facebook friends, picture friends and reciprocal tagging). A similar study by Nick et al. [119] analysed a Facebook dataset of 100 US Universities and concluded that homophily by dormitory, graduation year, and gender is strong.

The above ideas have also been applied to algorithms. In the literature, the approaches used to match friends can be generally classified as content-based and link-based.

Algorithms relying on content, use the similarity of users' profiles in order to make friend recommendations. This implies comparing what users state in their profiles to keywords and tags from other profiles [35]. This general approach has been successfully used for recommending books, movies and web sites (e.g., [112]). Link-based algorithms (e.g., friend-of-friend) use social network information only, relying on the idea that if two persons have a lot of friends in common, perhaps they could be friends. For example, the Facebook feature “people you may know" is partially based on this approach [35].

In this work, we build on the notion of homophily - which has been largely studied - but unlike previous works we focus on predicting the feeling of connectedness and social interactions. Our results could inform approaches for recommending friends and conversation topics.

4.3 Methods

4.3.1 Hypotheses

In this exploratory study, we specifically investigate the following hypotheses:

H1. Common life points are related to the level of online and face-to-face interactions This will help us understand if and how we can predict the frequency of social interactions based on the similarity of people (RQ1).

H2. Connectedness is related to common life points and both online and face-to-face interactions. This will tell us if and how we can predict connectedness on the basis of similarity of users and their frequency of interaction (RQ2).

We should notice that the above corresponds to a preliminary work, in which we are setting the direction for further analysis. We do not assume any causal relationship, which should be tested with a controlled trial.

4.3.2 Data collection

Researchers from our group collected information from Facebook users, both automatically (from users’ profile, with users’ permission) and by explicitly asking users about the frequency and nature of their interactions with friends, as well as the level of connectedness they feel with friends. We analysed profile information (specifically the common aspects between people’s profiles) and interactions to build a model for predicting connectedness and actual face-to-face interactions. In other words, our variables are:

- **Connectedness**. Measured using an adaptation of the Inclusion of Other in Self (IOS) scale by Aron et al. [6], a 7-point scale that relies on pictograms.
- **Social interactions**. Described in terms of *online interactions* and *face-to-face interactions*, both measured on a 5-point frequency scale.

- **Common life points**. Described in terms of *shared relationships* (family ties, having lived in the same places, having attended the same institutions), and *shared aspects* (shared beliefs, activities, and interests).

To collect the information needed for the analysis we developed a Facebook application called FriendRover[1]. The workflow of the application is illustrated in Figure 4.1 and detailed below:

1. Users open the application and instructions are shown as well as the request for consent. After giving consent and logging in, the data on participants' Facebook profile, friends, posts, and interactions on their posts, is automatically collected and anonymised.

2. Each participant is presented with a list of 20 friends who have interacted with the participant's posts (through reactions, comments, and tags). These friends are selected in a way such that they are representative of different levels of interaction (We categorised friends into quartiles according to the interaction with the participant and then took a sample from each quartile). From this list, users report on *connectedness* and *social interactions* (Figure 4.1 A).

3. The 10 friends rated as more connected by the participant are listed, and for each friend participants are asked to specify the traits that better describe this friend. On this interface participants report on *common life points* (Figure 4.1 B).

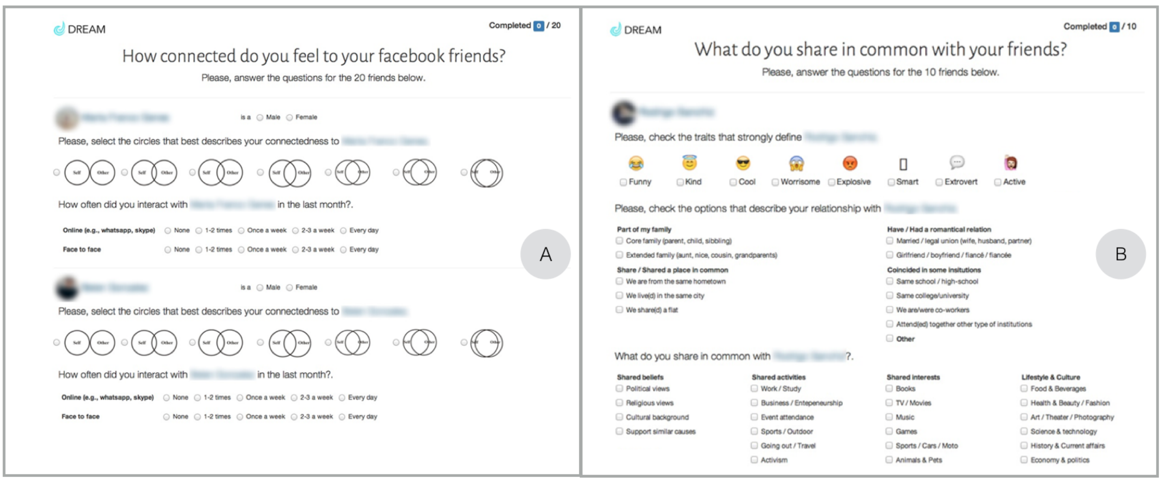


Fig. 4.1 The FriendRover application. Here we show the a) connectedness and b) common life points forms. ©2016 IEEE.

[1] Available at: http://happy.mateine.org/friends

4.3.3 Participants

The study was conducted online with a convenience sample of Facebook users (over 18 years old), obtained by advertising the survey on the Facebook pages of members from the research team. Participants were eligible if they have interacted with at least 20 persons. For this study, we advertised the experiment among Spanish-speaking users.

4.3.4 Resulting dataset

Responses of 33 participants have been collected (age range: 32-65, mean: 33 years old, 45% female), which resulted in 660 friendship relationships. The dataset consists of 660 connectedness samples and 280 reports on common life points out of 330 possible reports, this is because some participants did not complete the second part in full.

4.4 Results

4.4.1 Common life points are related to the level of online and face-to-face interactions

We addressed H1 by testing the association of common life points with online and face-to-face interaction separately.

An analysis of variance was performed to determine a statistically significant difference in the level of *online interactions* for the number of common life points, using the number of *shared aspects* and *shared relations* as independent variables. The results show a main effect for number of shared aspects ($F(1, 279)=57.268$, $p<.001$) and a main effect for number of shared relations ($F(1, 279)=15.251$, $p<.001$), but no interaction effect between both variables.

Analysing the individual components of both dependent variables we see a main effect for *shared activities* ($F(1, 279)=27.535$, $p<.001$) and *shared interests* ($F(1, 279)=31.439$, $p<.001$) but no main effect for shared beliefs ($F(1,279)=0.996$, $p=.319$). We also observe main effects for *common institution* ($F(1, 279)=9.483$, $p=.002$) and *common place* ($F(1, 279)=6.798$, $p=.01$) but no main effect for *family ties* ($F(1, 279)=1.547$, $p=.214$).

The above suggests that shared beliefs (religion, politics, cultural background, and causes) do not significantly help predicting online interactions when controlling for other factors. Likewise, social interactions do not significantly differ for relatives vs. non-relatives (family ties), when other factors are considered. Overall, as seen in Figure 4.2, the relationship suggests that **the more aspects one shares, the more frequent the online interactions are** – especially when there are common interests and when people engage in joint activities. This

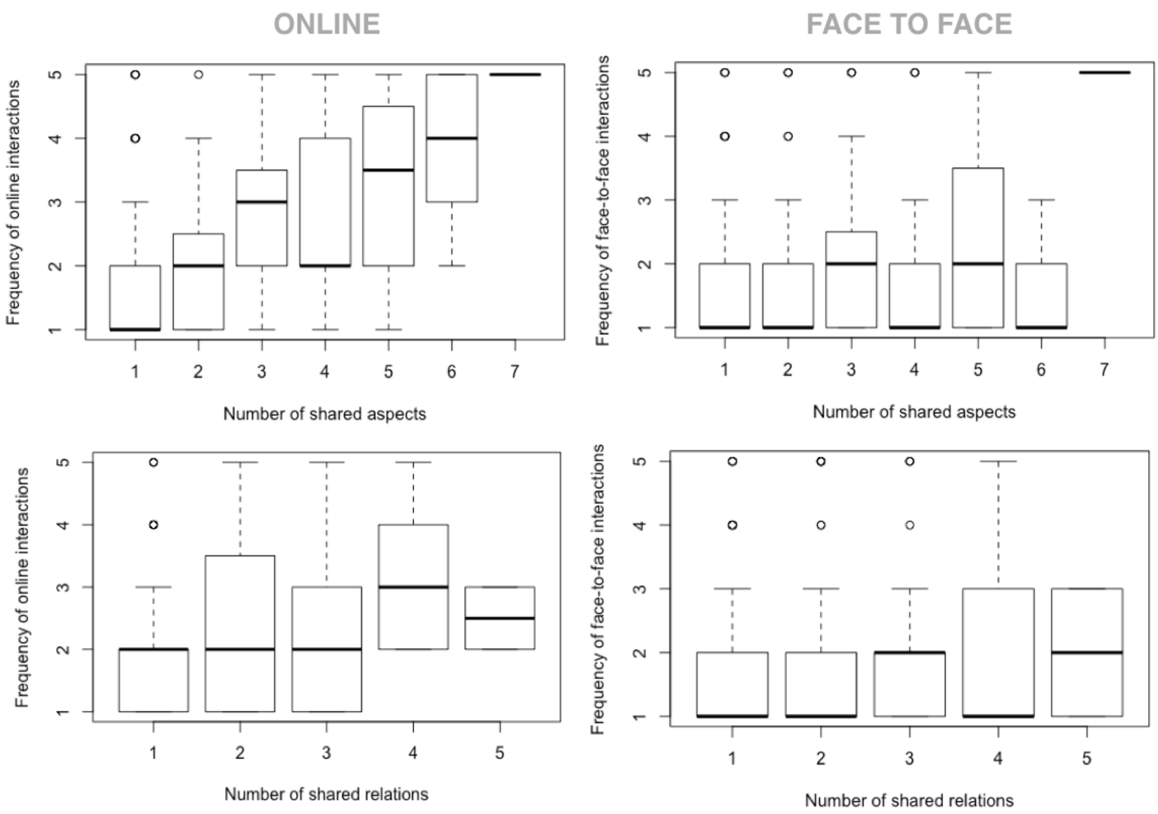


Fig. 4.2 Common aspects and social interactions. ©2016 IEEE.

trend is not present in shared relations, where some relationships might be dominating the effect.

Using the same model with the level of *face-to-face interactions* as dependent variable, the results show a main effect for number of shared relations ($F(1, 279)=8.328$, $p=.004$), a main effect for number of shared aspects ($F(1, 279)=12.587$, $p<.001$), and an interaction effect between both variables ($F(1, 279)=9.420$, $p=.002$).

Replacing the independent variables for their individual components in the model, we see a main effect for *shared activities* ($F(1, 279)=5.388$, $p=.02$) and *shared interests* ($F(1, 279)=11.480$, $p<.001$) but no main effect for shared beliefs. We also observe main effects for *family ties* ($F(1, 279)= 4.940$, $p=.027$) and *common locations* ($F(1, 279)= 4.513$, $p=.034$) but not for *common institutions*. These results are similar to those for online interactions with the difference that **family ties become a relevant predictor of face-to-face interactions**.

4.4.2 Connectedness is related to common life points and face-to-face interactions

To test whether there is a significant difference in connectedness for the various levels of social interactions, we performed an analysis of variance with *connectedness* as a dependent variable and the levels of *online* and *face-to-face* interactions as independent variables.

The results show a significant main effect for the level of face-to-face interactions (F(1, 659)=388.4, p<.001) and online interactions (F(1, 659)=218.5, p<.001), and also a significant interaction effect between both variables (F(1, 659)= 57.8, p<.001).

We illustrate the above relationships in Figure 4.3. For social interactions, the relation suggests a **higher level of connectedness for people interacting more frequently**. The outliers for the lowest levels of interaction correspond to people living abroad but interacting online very frequently (online), as well as people spending time together but not so much of this time online (face-to-face). This is an example of the interaction effect between both variables.

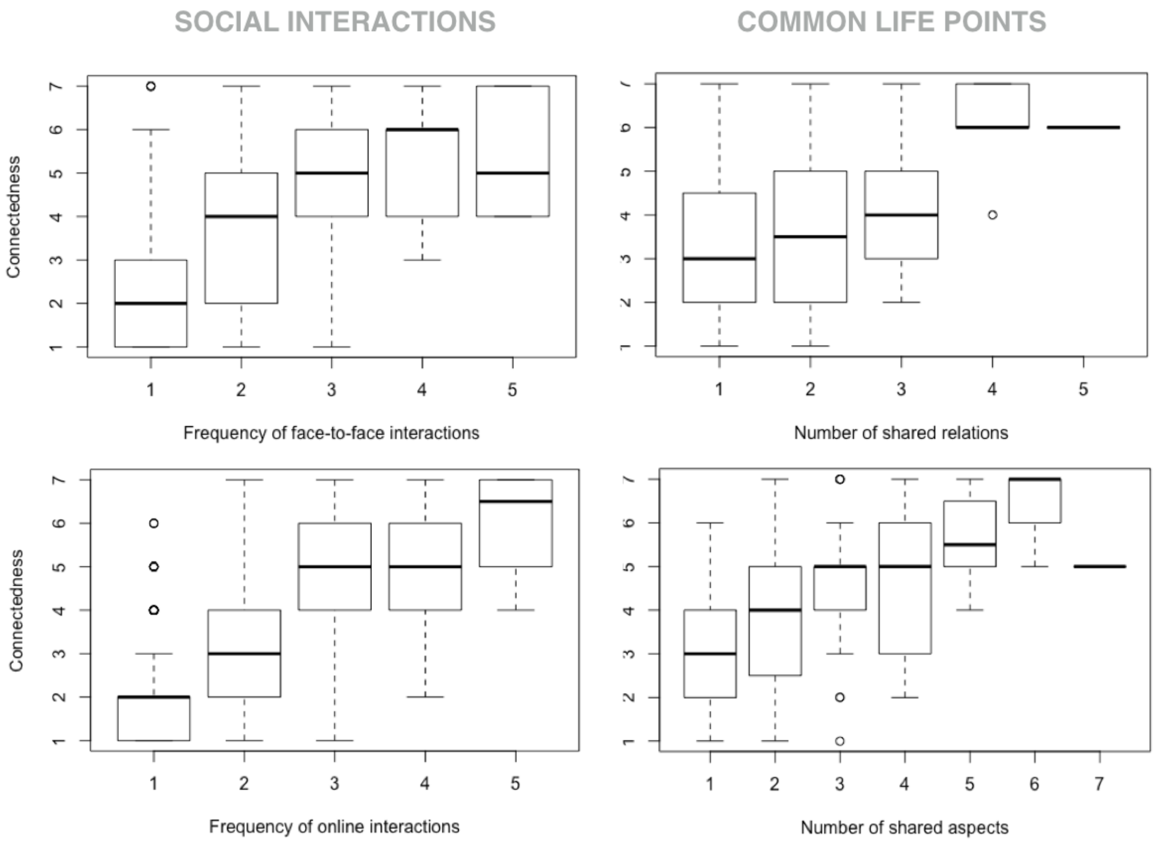


Fig. 4.3 Common life points and connectedness. ©2016 IEEE.

Analysing the relationship with common life points we observe main effects for the number of *shared relationships* ($F(1, 279)=43.48$, $p<.001$) and for the number of *shared aspects* ($F(1, 279)=46.06$, $p<.001$), but no interaction effect between both variables. These relationships are illustrated in Figure 4.3 and suggest that **having more common life points contributes to a higher level of connectedness**.

More details are presented in Figure 4.4, showing the percentage of shared aspects by connectedness level. In the figure we can see a higher percentage of participants reporting sharing common aspects for higher levels of connectedness. The difference is more pronounced for shared interests.

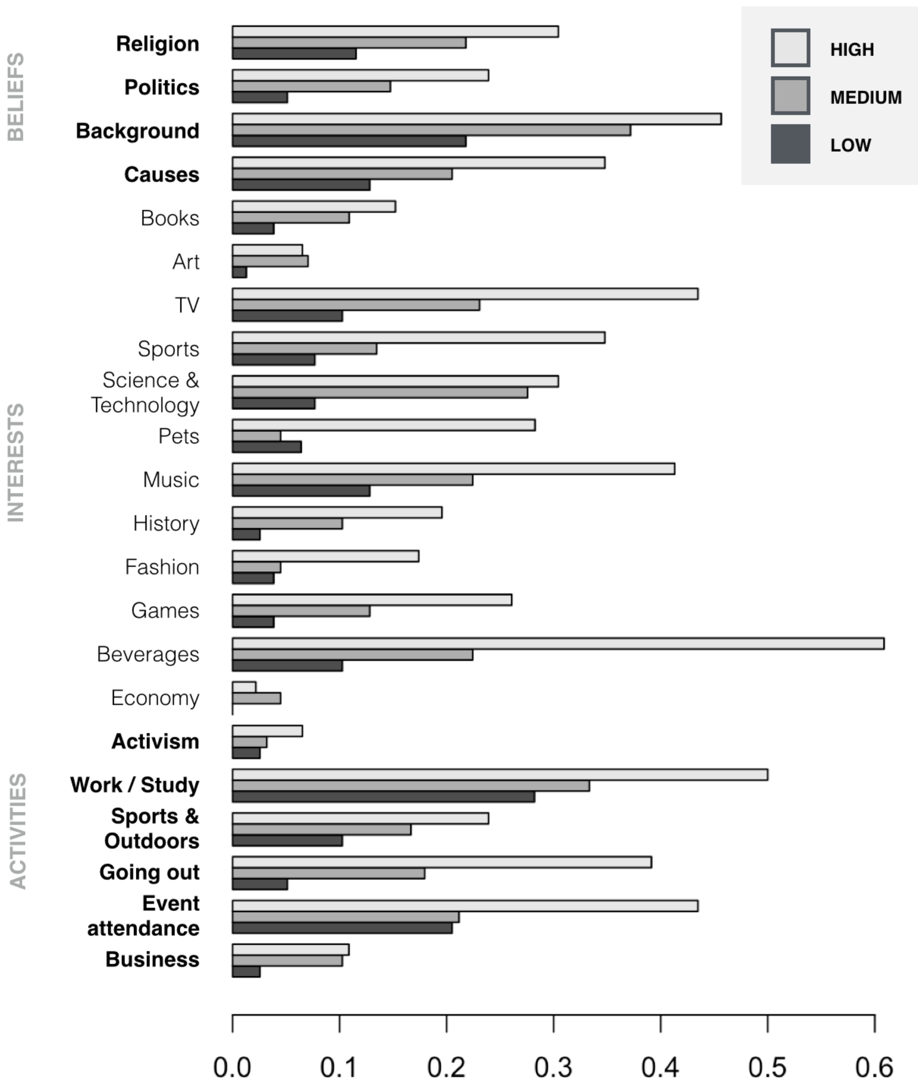


Fig. 4.4 Percentage of relationships featuring each common aspect, grouped and normalized by high, medium and low levels of connectedness. ©2016 IEEE.

4.5 Discussion

In this Chapter we have explored the relationship between connectedness, social interactions, and common life points on Facebook in two main research questions. As an exploratory work, the questions were approached from a general perspective but still bringing some interesting insights.

With respect to the relation between common life points and social interactions, we have seen that the more *common life points* friends share the more frequent their *online social interactions* are, and that shared interests and activities are determinant to this effect. For *face-to-face social interactions* the relationship is more complex, with family ties becoming a relevant predictor.

Interestingly, by exploring common life points we have seen that shared beliefs, as reported by the participants, is not a good predictor of social interactions, even when the literature points to this as a determinant factor [108]. We argue that this might be due to the homogeneity of the participants targeted by the study (Spanish-speaking), or simply the limitation in the type of metadata available on Facebook. Moreover, we have seen that shared activities are strong predictors, which is in line with previous literature stating that accomplishing practical activities together strengthen social ties.

We have also seen that higher levels of interaction and common life points are related to higher levels of connectedness. This suggests that one potential direction to creating bonds is generating opportunities for similar people to have meaningful interactions.

The above gives empirical support to technology aiming at increasing social interactions and creating long term bonds, by - for example - i) seeking to match users based on common life points, ii) generating conversations around shared interests, and iii) engaging users in shared activities.

As for ongoing and future work, we plan to follow up on this study to extend it to other countries (Costa Rica, Italy, Mongolia, Paraguay, Philippines, Russia, United States) and analyse cross-cultural as well as age-group differences. From a technological standpoint, we are currently incorporating these findings in the design of tools to reduce social isolation in older adults, which involve productive activities (crowdsourcing / volunteering) and socializing online. The latter comes from the fact that social interactions are of particular importance when providing productive activities to older adults [79], and it is one example of how the findings of this Chapter can be applied to collaborative systems.

Chapter 5

Conversational agent for reminiscence and re-connection

5.1 Background and motivation

In this chapter we define the requirements and early design of a conversational agent that supports engaging and effective conversation about a person's life and memories. Specifically, we discuss the case of personal and social *reminiscence* sessions by older adults.

Reminiscence is the process of collecting and recalling past memories through pictures, stories and other *mementos* [181]. The practice of reminiscence, and more generally, reminiscence therapy and life review, have well documented benefits on emotional and mental wellbeing [15, 151], interpersonal relationships and interactions [77, 5], and preserving personal identity [42, 57]. These benefits, along with the potential of reminiscence to transfer knowledge, stimulate conversations, reduce boredom, and maintain intimacy [179] makes it a very desirable practice, especially for older adults.

Research on technology-mediated reminiscence have taught us valuable lessons on how to facilitate usage by older adults [182, 133], collect memories and support storytelling [102, 101], stimulate cognitive functions [152, 101], and support conversations [91, 7]. However, these tools are very limited when it comes to creating opportunities for social interactions and for supporting social integration, something that ample and concordant research tells us it is a determinant of wellbeing at any age [68]. In our previous work, we addressed some of these limitations in a reminiscence-based social interaction tool [78] that aimed at stimulating conversations and creating bonds among older adults in residential care. However, our previous efforts and those of the community have reached a limit in terms of what traditional reminiscence technology can do to support older adults.

As we will see, traditional reminiscence technology i) strongly relies on co-located human presence for assisting in the reminiscence process, engaging in conversations or joining social reminiscence sessions; this often requires dedicated personnel, the presence of family member and friends, which greatly limit the practice and potential benefits of reminiscence, especially for those with less access to social contacts; ii) do very little to actively guide reminiscence sessions in a way that is effective and engaging, requiring users (the participant or a guide) to recall important aspects of the participant's life, to rely only on intentional triggers, or to follow predefined templates; thus missing the opportunity to reflect interests and previous stories to make sessions more fun and satisfying, and iii) memory collection and digital storytelling is mostly limited to archiving and browsing, missing the potential benefits of building a rich profile of users, e.g., to improve care practices [159] or facilitate social interactions [142, 27, 142], for example by identifying people with common values, or simply by making one's own family, and especially grandchildren, aware of the rich and often interesting life their grandparents have lived and are living.

Conversational agents and the ever growing number of cognitive services offer a unique platform to engineer the next generation of smart reminiscence systems that are not only more personal and engaging, but that can make those who have less opportunities for co-located interactions enjoy the benefits of the reminiscence practice. Engineering such a system requires insights into the reminiscence process, the development conversational models for reminiscence, the underlying cognitive services and system design.

In this Chapter we describe the requirements, models and a system design for such a conversational agent. From a technological standpoint we investigate the design of a chatbot that can flexibly drive a person (or a family) through a picture-driven storytelling session. The goal of the bot is to be able to sustain a conversation about a person's life, while harvesting useful information and while keeping the session interesting and engaging.

5.2 Related Work

Exploring the use of technology to support the reminiscence practice by older adults has been subject of plenty of research (see [99] for a review on the subject). The work on this area can be summarised in efforts to facilitate usage by older adults [182, 133], collect memories and support storytelling [102, 101], stimulate cognitive functions [152, 101], and support conversations [91, 7, 78]. Despite the valuable contributions, most of these works still rely either on the ability of participants to drive the personal reminiscence sessions or the presence of assistance [7, 78]. Support for social interactions is also mostly focused on co-located settings [171, 7] or offered in the form of online sharing [182]. The efforts in improving data collection are also largely focused on collaboration [102] and optimising the interaction design [101], with little to no intelligence. This represents not only a missed opportunity but also a constraint for those people with less opportunities for face to face interactions.

Another relevant type of technology is that of digital companions and relational agents. The feasibility of social companions for older adults, exploring a wide ranging set of activities was evaluated in [175], using the Wizard of Oz technique, with a researcher selecting the agent responses for a period of a week. An interesting finding from this work is identifying "storytelling" as the type of interaction elders spent more time with the agent. In the follow up study [137], the authors explored the effects of proactive vs passive behavior in conversational agents using a scripted conversation system, identifying that agents capable of initiating interactions (proactive) were more effective in addressing loneliness.

Similarly, a commercial digital companion driven by a human operator was also evaluated in [44]. The companion, an always on digital pet with conversational skills, was positively assessed by the participants, especially in the ability to show pictures (sent by a caregiver) and "remember" things about the older adults (e.g., a scripted reminder that included the birthday of the participant).

Another thread of work focused on identifying requirements as well as older adults attitudes toward autonomous virtual agents (not driven by a human operator). Tsiourti et al. [166] explored the attitudes and perception of older adults towards different aspects of virtual companions during the design phase, pointing to the "usefulness" and "social intelligence" a determinant factors for the success. The authors later conducted a exploratory study in 20 homes [167] with a virtual companion capable of assisting every day tasks, identifying several challenges, especially in the mismatch between expectations of the older adult in what regards the conversational capabilities of the virtual companion.

What the above tells us is that the use of conversational agents is feasible, and that reminiscence is an untapped and desirable application to explore. We also see that the success in previous studies was evident when agents displayed qualities attributed to humans and when proved useful to the task at hand – not surprisingly this was the case in agents operated by humans. All things considered, the use of cognitive services in reminiscence technology still remains in its infancy. This clearly motivates the need for research and development of intelligent conversational agents can assist the reminiscence process.

5.3 Requirements

To summarize the extensive theoretical work on the subject into a practical reference framework for technology-mediated reminiscence, we take the conceptual reminiscence model by

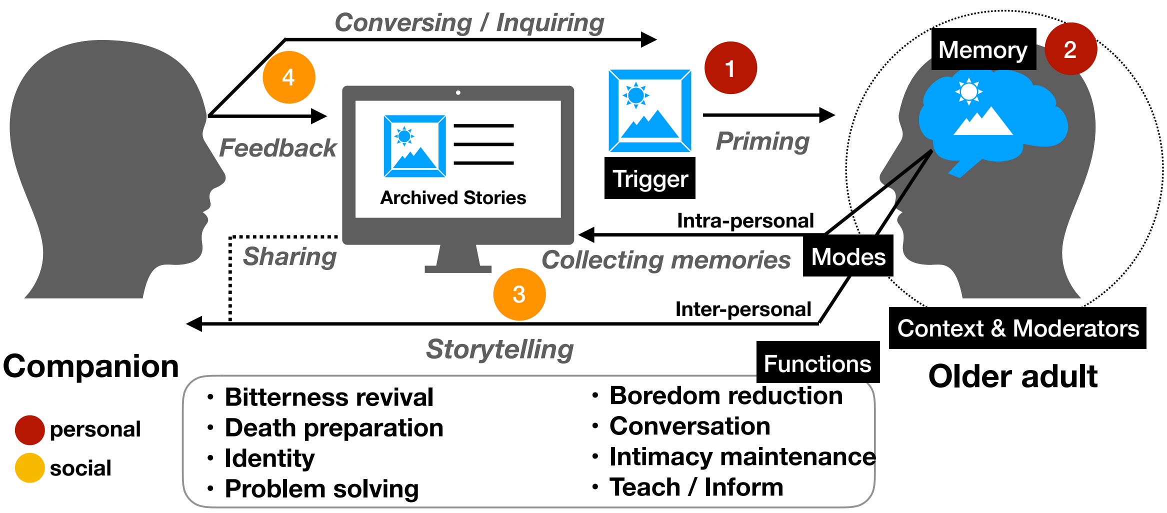


Fig. 5.1 Technology-mediated reminiscence process

Webster et al. [180] and instantiate it in a process that captures the most salient technology roles from the literature.

This summary, illustrated in Figure 5.1, shows how the process is initiated by a *trigger*, usually a memento in the form of pictures, videos or music, or specific questions. The trigger results in memories being primed, which can be then shared (virtually or face to face) via storytelling, or used for personal reflection. These two visibilities of memories define two distinctive *modes* known as interpersonal and intrapersonal. Memories and related stories resulting from this elicitation process can be digitally collected. If memories are used for interpersonal reminiscence in a storytelling session, then companions can engage in conversations, provide feedback, and further inquire about the aspects of the stories told by the person reminiscing. The role of the companion can vary depending on the degree of participation: companions can take a leading role in guiding the reminiscence session, co-participate in the case of social reminiscence, or take a more passive role as recipient of the storytelling. Regulating the entire process we have the *context and moderators*. These are intrinsic properties of the person (and reminiscence session) that can influence the reminiscence process and outcome, such as aspects of personality, age, gender, and relationship with the reminiscing companion. Finally, the practice of reminiscence is related to very specific functions, among which we highlight: i) boredom reduction, referring to the potential to create joyful experiences, ii) conversation, to the opportunity to create a dialog, iii) and teach / inform, to the opportunity to transfer stories and knowledge.

Informed by the above process, and our own experience designing and evaluating reminiscence technology for nursing home residents (on average, 80+ older adults that require

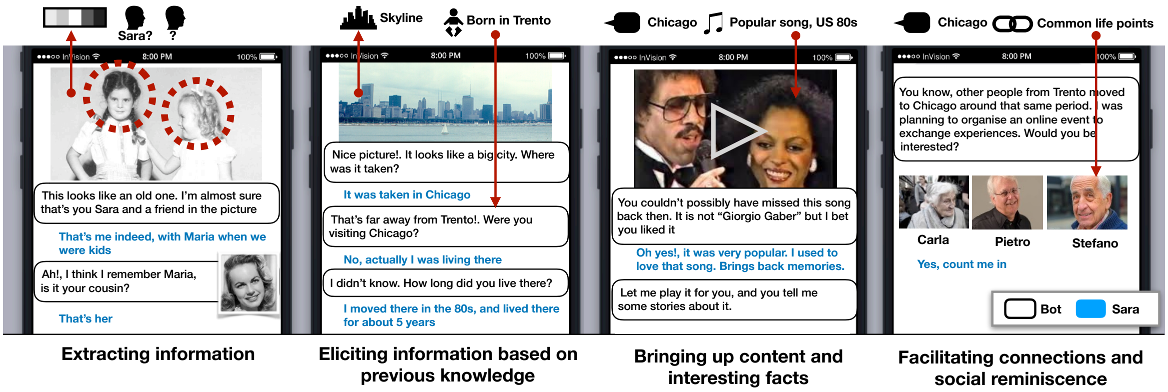


Fig. 5.2 Example conversations illustrating chatbot skills

assistance to perform activities of daily living) [78], we identified the following main requirements:

- **Eliciting memories**, in a way that engage participants in storytelling, in conversations that are sensitive to (changing) interests of the user, the topic, sentiment, and context.
- **Archiving life stories and memories**, to support personal reflection, knowledge transfer or even virtual storytelling.
- **Building representations of the life of a person**, not only to facilitate browsing, but to serve the growing need for historical information about older adults, e.g., to improve care practices [159], facilitate social interactions [142, 27], and build better context and moderators to improve the reminiscence process.
- **Navigating interests and passions**, as to engage users in conversations where they can have fun by navigating their passions (e.g., music, video, movies). This also helps elicit stories, and more importantly collecting preferences.
- **Facilitating (re-)connections**, by identifying potential companions and creating opportunities for social reminiscence and interactions. Interactions are facilitated to avoid stress in initiating conversations.

In addressing the above requirement, technology support will have to take a more proactive role in serving the reminiscence functions and needs of older adults.

5.4 Concept of a social reminiscence service

We propose the design of an intelligent reminiscence chatbot that can play the role of *companion* in reminiscence sessions with older adults. The chatbot acts as the guide of reminiscence session in a way that is engaging and fun, facilitating the tasks of collecting and organising memories and stories, while finding opportunities to connect people and stimulate social interactions. We exemplify these capabilities in Figure 5.2.

The above requires the bot to adopt qualities that go beyond traditional scripted dialog systems [137, 44], to qualities traditionally attributed to humans: i) understand *triggers*, as to formulate meaningful questions in the same way someone looking at a picture would use that information to formulate related questions, ii) learn about the user, and use this *knowledge* to elaborate more natural conversations and improve memory elicitation, iii) *sense* where the conversation is going and guide the sessions accordingly, balancing the objective of collecting specific information (e.g., details on what a picture represents) with the opportunity to follow the person's flow of thinking and be directed by it, even if it takes us away from the picture we want to collect information on.

Note that while use chatbot and text-based interactions in our example, these are only to illustrate the capabilities of the system, and to drive the concept development. Designing interactive technologies for older adults is a challenging task that requires specific attention to their abilities and limitations [123, 70]. Thus, as part of our ongoing efforts we are investigating designs of conversational interfaces that can support individuals of potentially different interaction abilities.

5.4.1 Conceptual model

The chatbot operates over a model comprising three families of concepts: the *life model*, *reminiscence model*, and *conversation model*.

The life model captures the information about the life of a person and the properties that describe them. This includes i) life events, which are well defined moments in the life of a person, such as birthdays, wedding, graduation, among others; ii) habits and skills, describing the person capabilities and routines, such as knitting or reading; iii) values, describing the person beliefs and ideals, such as religion, political views and social causes; iv) preferences, capturing passions and hobbies, such as music, sports, gardening; and v) relationships, referring to familial and non familial connections of the person. All this information is defined from a life course perspective, as they can refer to specific points in life. This temporal dimension cannot be measured in terms of specific dates, but in more flexible terms such periods of life.

Most of the concepts that are specific to the reminiscence practice were introduced in Figure 5.1. The reminiscence model defines: i) mementos, digital artefacts (private or public) that can be used in the reminiscence session to trigger memories, such as pictures, videos or songs; ii) tags, annotations that describe the mementos, such as people, places, dates related to a picture; iii) life stories, anecdotes and memories that result from the reminiscence and

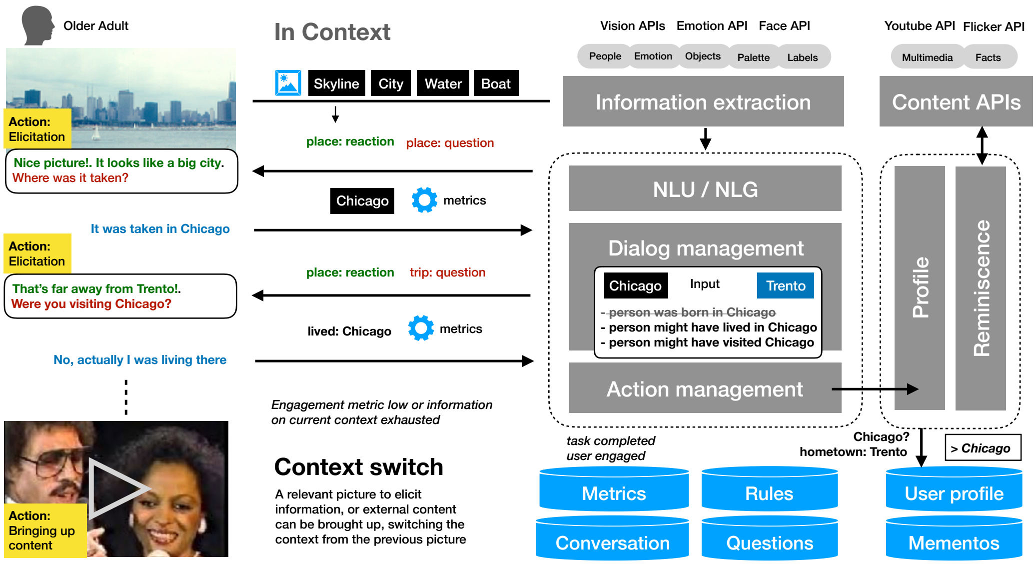


Fig. 5.3 Example of bot execution

storytelling session, possibly in response to a memento. Complementing this model there is the concept of the participant of the reminiscence session, which is captured by the life model, and companion, which we define as part of the conversation model.

The interactions between the user and agent define *conversations*. We organise conversations in *sessions*, which give scope to the interactions during a reminiscence session. An interaction between the user and agent is usually referred to a *turn*, and characterised by an input and a response. A set of input and responses defines a *context*, which provide implicit knowledge to the current and future turns. In the reminiscence bot, the memento under discussion, the questions recently asked, and the information provided by the user are key components of the immediate context. At a more global scale, the knowledge about the user (see life model) and the history of conversations define the *macro context*. The *actions* define the capabilities of our reminiscence bot, such as eliciting information, showing understanding (reactions), bringing up interesting multimedia content, and connecting people. Serving these actions we have a database of *questions* related to specific user knowledge entities (to elicit information), a database of mementos (public and private, to bring up content), and a continuously updated database of user similarity and user clusters (to find potential connections). Orchestrating the decisions of what action to take and how, there is a set of *rules*, and *metrics* that facilitate the self-assessment of the bot.

5.4.2 The chatbot at work

The conversation model together with the reminiscence and life model work with the cognitive services as shown Figure 5.3.

The picture used as trigger is first passed through a *information extraction* service such as Google Cloud Vision that automatically detects its salient features (e.g., color palette, people, artefacts, landscape, emotions). In the example, the service detected a city "skyline" as the most important feature. Being the first interaction, and the context limited to the fact that a picture of a city was provided, the rule system decides to elicit basic information about the picture, embedding this information to elicit the related knowledge (*place*: "Nice picture! It looks like a big city. Where was it taken?"). The user's reply is then processed by an NLP service such as DialogFlow or Google Cloud Natural Language to extract related entities. Metrics are also run to compute engagement and effectiveness of the action and question. The relevant tag (place: Chicago) is then associated to the picture.

In deciding the next action the system consider again the context, which is now comprised by a picture of Chicago, and information about the last exchange. This context combined with knowledge about the user (user: born in Trento) and metrics assessing a positive engagement, makes the following information elicitation rule to kick in ("User was born in Trento, and the picture is from Chicago: the user might have lived or visited Chicago"), requiring the connection of the user with the place to be inquired. This rule combined with the entity in context is paraphrased in a question as ("That's far away from Trento! Were you visiting Chicago?"). As before, the subsequent user responses are used to build the new knowledge that (The user had lived in Chicago in the 80s). Following the conversation, the interest of the user or the gain in eliciting information on the picture might decrease, motivating the bot to bring up new content related to the context to elicit new information, or suggest relevant connections.

It is important to note that the rules and context switching are not predefined, but can be trained using crowdsourcing and machine learning techniques, by crowd of experts and non experts depending on the needs of a specific target population.

Mockups of the chatbot can be seen at http://storygram.net.

5.4.3 Metrics and measures of success

We briefly describe metrics that might shape the chatbot training and ability to self-assessment:

- **Engagement** describes the level of involvement and interest of a person in a conversation. Capturing engagement is important as it has shown to be associated with longer conversations and higher sentiment score [33]. Different engagement metrics have been proposed for conversational agents, such as: i) *conversation length*, in minutes and/or amount of dialog turns [56, 33]; ii) *total number of interactions* with the bot over a session [56]; iii) subjective *user engagement ratings* [33]; and iv) cumulative *average sentiment score*, to see the user sentiment towards a topic [33].

- **Task completion** captures the effectiveness achieved in the execution of a specific action. The metrics that have been recently proposed for the task-based evaluation are: i) *task completion rate* and ii) *completion time*, which can be measured as a number of dialog turns the user talks to bot to complete a task [75].
- **Conversation quality** describes how much consistent the conversation is and the ability of the bot to correctly remember things the user said on the previous turns of the conversation. For evaluation the conversation quality, several metrics have been proposed: i) conversational consistency, ii) memory of past events, and iii) response speed [98].
- **Human-like communication** is proposed as a measure of the conversation quality and naturalness, where conversation logs are analysed to see if a user had conversed with the bot as if with a human, e.g. using greetings or showing "proactivity" [93].

On the other hand, measuring the actual benefits of the such a tool on the mental and social well-being of individuals would require the design of RCTs and the use of validated instruments.

5.5 Discussion

It is clear that conversational agents have the potential to overcome the limitations of traditional reminiscence technology, but further research is needed in order to go beyond scripted or human-operated systems. We did a step forward in this direction and proposed the use of well defined models (life, reminiscence and conversation models) in combination with cognitive services to provide agents with some human qualities.

The use of cognitive services is fundamental in providing bots with a required level of understanding to drive the reminiscence sessions. Recent developments in automatic caption and question generation [115, 163] show promising results, and inspire our future work in building more intelligent reminiscence applications.

Chapter 6

Crowdsourcing for reminiscence bot

6.1 Related work

Crowdsourcing has been used to support all aspects of chatbot design, from holding direct conversations with final users, to supporting conversation design – the latter being the family of approaches under which we position our work. Prior work on crowdsourcing has addressed the *bootstrapping challenge*, investigating strategies to create dialog datasets to train algorithms [154, 104], infer conversation templates [111] or declarative conversation models [118]. It has also been explored to *enrich conversation dialogs* to provide meaning and context, by annotating dialogs with semantics and labels with, for example, polarity and appropriateness [104], extracting entities [73], as well as providing additional utterances for more natural conversations (paraphrasing) [84]. Other approaches incorporate the crowd in the *evaluation of chatbot quality*, making sure crowd contributions are valid and safe [38, 74] and even allowing users to train chatbots directly [38]. Acknowledging that chatbot conversations are not perfect, some approaches explore strategies to *escalate conversation decisions* to the crowd in cases where the chatbot is not able to interpret or serve the user request [14].

The above highlight the potential of crowdsourcing for designing chatbots. We take these approaches as the starting point for exploring the specific challenges of designing and maintaining a reminiscence bot. Previous work in this domain – though valuable in insights – has been limited to human-operated chatbots and Wizard of Oz evaluations, highlighting the complexity of chatbot design in general and in particular for our target population [167, 62, 187].

6.2 Context & Objectives

Reminiscence is the process of collecting and recalling past memories through pictures, stories and other *mementos* [181]. The practice of reminiscence has well documented benefits on social, mental and emotional wellbeing [151, 77], making it a very desirable practice, especially for older adults. Research on technology-mediated reminiscence has advanced our understanding into how to effectively support this process, but has reached a limit in terms of the approaches to support more engaging reminiscence sessions, effectively elicit information about the person, and extend the practice of reminiscence to those with less opportunities for face to face interactions.

In our previous work [120] we made a case for conversational agents in this domain, and proposed the concept of a smart conversational agent that can drive *personal* and *social* reminiscence sessions with older adults in a way that is engaging and fun, while effectively collecting and organising memories and stories. The idea of conversational agents for older adults is not new, and they have been explored to support a wide variety of activities and everyday tasks [167, 175, 66, 168], to act as social companions [137, 138, 44] and even to engage older adults in reminiscence sessions [62].

While these works give us valuable insights into the opportunities of using conversational agents as an instrument to support reminiscence sessions, they also show us how limited our knowledge is in terms of effective strategies to maintain dialogs with older adults. Success stories are mostly limited to Wizard of Oz evaluations [143], in which system functionality is partially emulated by a human operator, or based on fully human-operated agents. The few attempts at autonomous agents highlight issues with the mismatch between user expectations and the actual social capabilities of the agents [167], general challenges with designing conversations suitable to the target population [187], and challenges with engaging older adults in question-based interactions in particular [62].

In this Chapter we aim at identifying effective and scalable crowd-based strategies for designing content, conversation rules, and meaningful metrics for a reminiscence chatbot targeted at older adults. We build on the concept introduced in [120] and identify *where* and *how* crowdsourcing can help to design and maintain an agent-mediated reminiscence process, while addressing the specific challenges posed by the target population.

6.3 Reminiscence Chatbot

The envisioned chatbot is based on the idea of automatically guiding older adults through multimedia reminiscence sessions [120]. It has the dual purpose of i) collecting and organ-

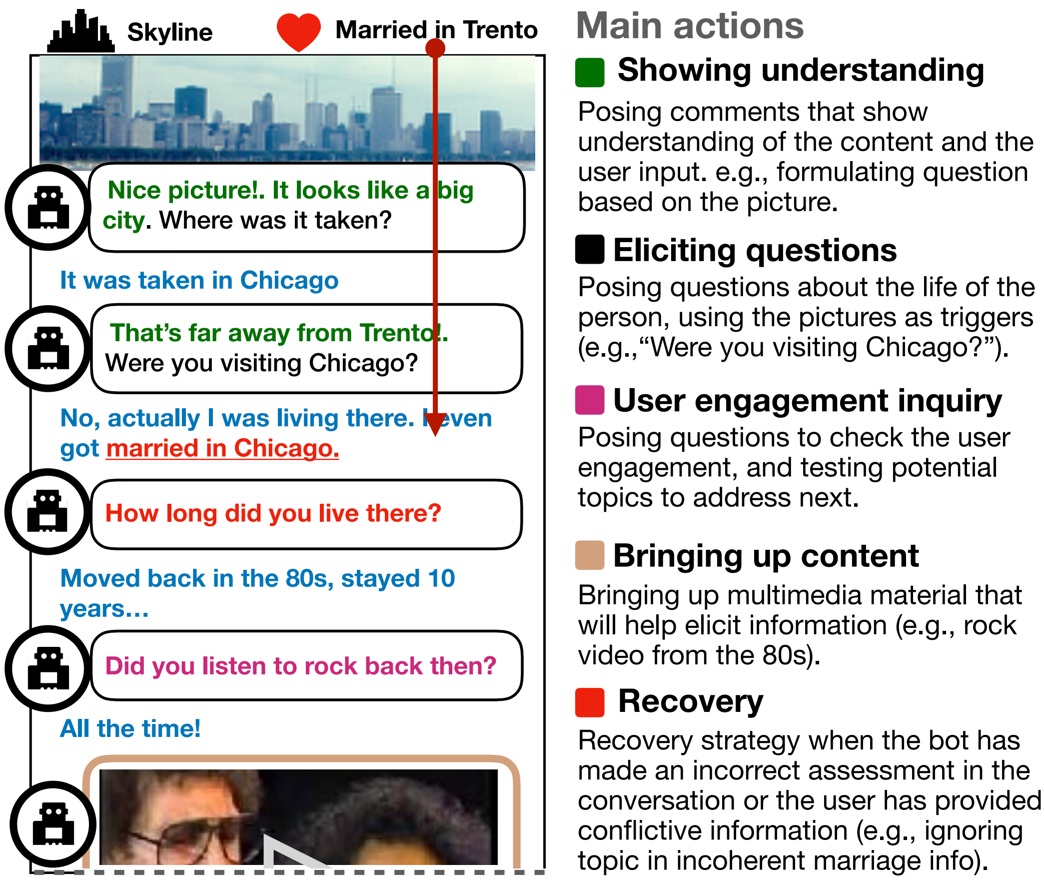


Fig. 6.1 Example reminiscence session with bot actions

ising memories and profile information, and ii) engaging older adults in conversations that are stimulating and fun. In Figure 6.1 we show an example of conversation and related main actions.

The example starts from the subject (the elder) providing a memory in the form of a picture. In response, the chatbot poses a contextual question. In order to do so, it must be able to *understand the theme of the picture* (big city) and to *extract and understand information* from pictures and text. In order to keep the conversation natural, it must further be able to *reference related conversation topics* (the city of Trento) and, in order to show empathy, it must be able to *sense the feelings of the subject* as the conversation evolves (e.g., it looks like the subject likes rock music, so it could be an idea to talk about that for some time). It would also be good if the bot be able to *sense the presence of peers* (e.g., family members or moderators helping with the chat). All this information helps the bot *decide on appropriate next actions* taking into account possible *conversational goals* (e.g., elicit basic user profile data). Among the most complex decisions to be taken is deciding if and when to *change context* in a conversation (e.g., to make the elder laugh).

All these requirements are particularly challenging since special attention must be paid to the subject's abilities and limitations [123, 70]. For instance, it is hard to *cope with user-initiated context switches* or to *keep knowledge about subjects coherent* due to cognitive decline associated with age [128]. Coping with these challenges is difficult even for humans [110].

In the long term, our goal is to develop a crowd-powered chatbot that implements the necessary conversational logic, sensibility and tricks to engage older adults in pleasant and satisfactory reminiscence sessions. The crowd should not be involved in direct interactions with the elderly (like in some real-time crowdsourcing approaches studied in literature [105, 138]), nor should it be used just to train black-box AI algorithms. The idea is to involve the crowd to elicit and represent reminiscence-specific conversation knowledge explicitly in the form of some dedicated model, in order to be able to actively steer the conversation into specific directions (e.g., to elicit health issues or family memories). In this Chapter, we focus on an intermediate set of research objectives: identifying (i) how to *model* the conversational knowledge the chatbot may rely on and (ii) how to use the crowd to *learn and evaluate* the model.

6.4 Crowd-Supported Chatbot Design

6.4.1 Conversational Model Representation

Conceptually, a simple model we can imagine for a chatbot is a state machine (S, A, δ, π, F), where S denotes the states (a state includes the information on the subject and the conversation history), F denotes the final states, A is the set of (conversational) actions, δ is a state transition function (our conversational policy), $\pi : S \times A \rightarrow \{(s, p)\}$ associating to each state and action a set $\{(s, p)\}$ of possible target states s and the probability p with which that action should be chosen (to model that conversations are not deterministic).

In practice however the state space is infinite and the possible conversations are also infinite so this FSM is not the right model. An alternative model is based on Event-Condition-Action (ECA) rules, where the event for example is the sentence by the subject (the elder) and the condition is some expression over what we know about the subject as well as past events. This has however the same limitations just discussed.

We observe that what we really want to have is a definition of the domain and range of the policy function π so that we can learn a useful policy that can be applied to real life conversations. On the action side (the range), we approach the problem by clustering similar actions along several dimensions, such as i) the type of actions (ask information, make a comment, show interesting content) and ii) the topic of conversation (talk about the picture you are showing, or about childhood, or about hobbies). Given the action type and topic, there are many actual conversations and utterances, but at this level we are focused on learning types and topics rather than conducting an interaction within a topic or paraphrasing sentences.

In terms of the domain a policy is defined on, what we wish to have is a description of the characteristics of the state (or event and condition) to which the policy applies. For example, the crowd may tell us that after they learn the date of birth, they show newspaper covers of that year, or famous people born the same day, or songs that where popular when the subject was very young. In this case the trigger of the action is the last conversation element where the subject is notifying the state of birth (or, in terms of events, it is the event of the system, somehow, coming to know the date of birth of the person).

The challenge here is therefore to understand what is the reasoning of crowd workers when they decide to take actions, and based on this reasoning identify the classes of state and event information we need to attach policies to.

6.4.2 Crowdsourcing tasks

The counterpart of the model is the learning process, which has to do with how to design and process the results of crowdsourcing tasks. The objective we have in seeking the proper task designs are the following: (i) identifying *action types and topics* (unless we want to fixe them a-priori), (ii) identifying *when* (based on which state or trigger) a person changes topic or shows specific content, and (iii) identifying *why* (based on which state or trigger) the agent initiates a conversation on a topic.

To do this, we envision crowdsourcing tasks that aim at (i) exploring possible conversations (these can be Wizard of Oz simulations), (ii) reflecting over previous conversations by the same worker or other workers to derive the “rules" that made the worker take a certain course of action, and (iii) aggregating these “rules" into a smaller coherent set that reveals the characteristics that the policy model should have.

For example, the crowd may reveal that they change topic whenever they sense that the person is sad talking about the current topic. This would tell us that an important component of the policy domain is the perceived emotional state, something that therefore the agent should try to detect, and that change in this emotional state should be a trigger to either continue or change topic.

We thus focus on the following research question (**RQ**): *Which crowd-based strategies can help elicit effective conversation logic for conversations (reminiscence sessions) targeting older adults, and how?*

Conversational logic includes understanding of: composition of Dialog State, when and how the State has to be changed, and what are the most important variables that affect the state. That is, given:

- the set of States $S = \{S_1, S_2, ...S_m\}$, where S is the state of the conversation that consists of multiple features (such as user profile info, dialog history, sentiments);
- the set of possible Goals in the conversation $G = \{G_1, G_2, ...G_n\}$, where G is the current goal aimed at (e.g., elicit information, tell a joke, show engagement content); and
- the set of Actions $A = \{A_1, A_2, ...A_n\}$, A being the chatbot action performed, which changes the state and satisfies the current goal (e.g ask question to elicit info);

the aim is to:

- identify the composition of current State; and
- identify the *policy*, i.e., which Action to take given current state S and the Goals G
- such that
$$\pi(G,S) \rightarrow S'$$
where Policy π is a rule that defines the transition from state S to state S' and depends on the *Current State* S and current *Goals* G of the conversation.

The research question is actually of more general nature, and the resulting approach can be applied to any social chatbot. To us, reminiscence is an application domain we have experience with and we want to contribute to.

6.4.3 Success Metrics

Different metrics have been proposed for evaluating the quality of conversations with dialog agents, such as: i) user engagement [34, 56], ii) task completion [75], iii) conversation quality: including dialog consistency and memory of past events [98], iv) human-like communication [93]. The approach to evaluation – and therefore the choice of metrics – is based on the aim of the agent: having an engaging chat or performing a specific task (e.g., booking a flight). In our case, the reminiscence chatbot is a combination of conversational and task-based agent, as it aims at both having an engaging conversation with the user and collecting information while doing so. Therefore, we consider metrics for both types of agents, including: i) engagement (as subjective measure); ii) number of turns of conversation made before it drops; iii) times conversation drops overall; iv) domain-specific metrics like the amount of content which the user has provided during one conversation session (amount of pictures uploaded, amount of data attributes filled about a relevant person), and other task-completion metrics.

6.5 Discussion

Next, we are going to define concrete crowdsoursing strategies to elicit the nature of the states, goals and actions that will give structure to the model. Then, we will focus on tasks to fill the model with data and on algorithms to effectively aggregate and apply the elicited knowledge. At every turn, the reminiscence chatbot required to provide meaningful responses based on the dialog history. It is a difficult task, that needs careful consideration. First of all, we need to understand what are the exact values of the states of a dialog with a given particular goal. For example, if the goal is to collect information related to the school time of a user, some possible states might be: i) First love: Marilyn Monroe, School class: 3, When: 1940 ii) Best friend: Elon Musk, Activities together: fishing, smoking. Crowd workers can help us to get the first insight into the states and values. We can arrange chat sessions for collecting dialogs between two workers where one plays the role of an older adult and another of the reminiscence chatbot. By analyzing the collected dialogues we can infer potential states which are relevant to such type of conversations. In the following, knowing the composition of the states will help to design dialog management and so to decide what is the next question to ask during the storytelling session.

Chapter 7

Understanding the requirements and behaviors for reminiscence chatbots

7.1 Objectives

7.1.1 Opportunities in light of the previous results

Conversational agents (CA) can be effective at interacting with people in a natural and engaging way, while performing a variety of tasks, including *information-providing* tasks (e.g., providing the departure time for the next train), *conversational* tasks (such as having a chit-chat with the user, as in Chorus) [74], or *transnational* tasks, where the agent performs some action such as booking a hotel or buying a train ticket (e.g InstructableCrowd, Calendar.help, Nurtz, SnapTravel) [95].

The objective of our work is to identify the requirements and opportunities for a reminiscence chatbot.

The need for such an instrument became evident during a series of observational studies and interviews run in a set of nursing homes in Italy [78]. Indeed, in the studies researchers in our group have discovered that: i) nursing home (NH) staff members spend a lot of time trying to understand residents, to learn about their life and this is effort consuming; ii) the identification of information that can be leveraged to promote interactions among residents and even re-connection with old friends or formation of new friendships - something that is notoriously hard in later life; iii) elderly residents, family members and NH staff have perceived positively technology-supported reminiscence approaches presented during the studies.

We therefore consider a reminiscence chatbot to be successful not only if i) it is engaging and natural, but also ii) if it is effective in collecting information about the user (e.g values, interests, places) that later can be used for recommending potential friends. Furthermore, since reminiscence is a delicate activity which can be emotionally intense if not run properly, the chatbot needs to follow accepted practices for ensuring an emotionally safe reminiscence session.

In this context, our goal is to formulate design recommendations for a system that can facilitate reminiscence sessions by having a chatbot drive the sessions, stimulating the reminiscence conversations with elderly. We consider in particular two cases, a) when the elderly person interact autonomously with the chatbot, and b) shared-context use, when family members or nurse participate in the reminiscence session together with the elderly, help mediate the interaction (this is helpful in cases where the cognitive or physical abilities of the guest do not enable independent participation).

7.1.2 Research questions

The goals above lead us to the following research questions:

RQ1. Is a chatbot acceptable (both to our target users and according to doctors) as an instrument for storytelling (reminiscing) and collecting information about the person?

RQ2. Which context of chatbot use is considered safe and feasible by doctors and nursing home staff?

If we are successful in answering our research questions, as an outcome we expect to be able to provide UX designers and chatbot's developers with the requirements and recommendations for the design of the interaction, context of use and the interface of conversational agent. Such user experience research should be necessary step before building the any conversational agent for such category of users.

7.2 Background and Related work

7.2.1 Chatbots for older adults

The idea of conversational agents for older adults is not new, and they have been explored to support a wide variety of activities and everyday tasks [167, 175, 66, 168], to act as social companions [137, 138, 44] and even to engage older adults in reminiscence sessions [62]. Recent studies reports that conversational agents have been perceived positively by the category of older adults [167].

In the WoZ study done by Razavi et al. [135] elderly users provided positive feedback regarding the use of virtual conversational agent and have expressed the interest to use it at home settings. The virtual agent designed by the researchers aimed at improving older adult's communication skills by having conversational practice with them, observing and providing feedback to the users regarding their non-verbal communication behaviour. A single- session WoZ study was conducted with 25 participants in lab where the virtual agent's contributions to the dialogue were selected by a human wizard. A second follow-up study was conducted with 8 participants interacting with autonomous fully functioning version of system in home settings. To evaluate the conversations quality, 8 conversations with autonomous system have been compared with 8 WoZ conversations where agent outputs were selected by a human wizard. Transcripts were rated based on 6 features of high-quality conversation proposed by the researchers on : i) naturalness of system contributions, ii) system input encouraging user's participation iii) conversation stayed on track iv) responses relevant to the conversation v) understood what user said vi) responses are polite. The autonomous conversational system had high evaluation ratings which with the Woz version performance.

In the other exploratory study by Tsiourti et al. [167] on the evaluation of the daily life companion conducted over 12 weeks with 20 older adults, the data was collected on acceptance, perceived usability and usefulness of the companion. Regarding the overall acceptance, the companion was well percieved in the living environments of older adults. However, the companion didn't meet some expectations of the users: the misunderstanding of the speech from the companion was increasing the participants stress level and feeling of frustration. Regarding usefulness, participants reported that the system improved their daily life in general and made them more motivated to perform daily life activities.

Success stories are mostly limited to Wizard of Oz evaluations [143], in which system functionality is partially emulated by a human operator, or based on fully human-operated agents. The few attempts at autonomous agents highlight issues with the mismatch between user expectations and the actual social capabilities of the agents [167], general challenges with designing conversations suitable to the target population [187], and challenges with engaging older adults in question-based interactions in particular [62].

What the described above tells us is that the use of conversational agents is feasible, and that reminiscence is an untapped and desirable application to explore.

7.2.2 Evaluating conversational interfaces

Different metrics have been proposed for evaluating the quality of conversations with conversational agents, such as: i) user engagement [34, 56], ii) task completion [75], iii) conversation quality, including dialog consistency and memory of past events [98], iv) human-like communication [93]. The approach to evaluation – and therefore the choice of metrics – is based on the aim of the agent: having an engaging chat or performing a specific task (e.g., booking a flight). Many studies evaluate the user experience with the chatbots by conducting qualitative analysis of comments of the users provided after or during the interaction with tested bots [82, 157].

Cameron et al. [29] have recommended practices for healthcare related chatbots design, where seven categories for bot's evaluation has been proposed: personality, onboarding, understanding, answering, navigation, error management and intelligence.

In the exploration study of Jain et al. (IBM research) [82] on evaluation of first time user's experience with eight different text-based conversational bots over multiple sessions, four the data-driven themes on the bots qualitative user evaluations were proposed: functionality, conversational intelligence, personality, interface. The themes are defined as: i) *Functionality* as an ability to perform its main task; ii) *Intelligence* as the ability of the agent to understand the input, recovering from the failure gracefully; iii) *Personality* as how it suites the domain, human-like behaviour, the start, the end and presence of the humor in the conversation. iv) *Interface* on the presence of interactive elements (buttons, carousel, etc), bot description and main menu. Total interactions time, message count and interaction elements (e.g buttons) has been measured as well.

The users preferred chatbots that either provided a 'human-like' natural language conversation ability or an engaging app-like experience specifically designed for the familiar turn-based messaging interface (e.g click-based "Trivia Blast" bot). The design implications provided by the researchers include recommendation for bot to have "consistent personality with small talk and humor", playfulness, humor and conversational etiquette suggested as a positive traits for the bot.

In the exploratory study from Thies et al. (Microsoft research) [157] three types of conversational agents have been evaluated from the perspective of the chatbot *personality* in order to explore which chatbot personality are most compelling to young, urban users (university students). The study was conducted in Wizard-of-Oz settings when the "Wizard" was playing the chatbot role. After interacting with each of three bot's personalities participants have chosen the most preferable one and provided the comments on their experience interacting with the bots. Participants preferred WoZ chatbot that had friendly personality which can add value to their life by making useful recommendations, have a sense of humor, while being empathetic and non-judgmental [157].

One of the proposed metrics for measuring the quality of conversations is engagement.

Engagement is one of the proposed metrics for measuring conversation quality, it describes the level of involvement and interest of a person in the conversation. Engagement also has shown to be associated with longer conversations and higher sentiment score [33]. O'Brien et al. designed the Engagement scale [124] for measuring the user *engagement* of software applications aimed at online shopping, authors of the scale suggest that it can be also generalized to other types of software apps.

The results of evaluations and comments of the users from several studies suggest that the personality and the style of bot interaction behaviour with the user plays an important role in perceived conversation quality and affects to the ratings from users. Measures of user experience with the chatbots vary in different studies, while the qualitative comments analysis remain one of the main evaluation approach.

7.2.3 Personality in conversational agents

Theoretical background: Personality composition

Personality can be defined as a "dynamic and organized set of characteristics possessed by a person that uniquely influences their environment, cognitions, emotions, motivations, and behaviors in various situations" [109].

Recent research shows that a *personality*, as a set of human-like characteristics, is an important factor in the relation to how humans perceive conversational agents and it affects to the attitude of user whether to use the CA in the future again [28].

One of the most widely used framework for modelling personality is the Big Five model that based on the five main factors: *openness to experience, conscientiousness, extroversion, agreeableness* and *neuroticism*. Every personality according to the framework consists of the set of traits from each of the five factors, expressed to a greater or lesser extent. Since the meaning of the labels of the factors has been often misunderstood, John & Srivastava [85] have provided brief definitions for each of the labels in order to reduce misunderstanding. The definitions are described below:

i) Openness to experience, has been defined as tendency of the person to try new things, “breadth, depth, originality and complexity of an individual’s mental and experiential life" [85].

ii) Conscientiousness is related to goal-oriented behaviour and following the existed rules: “..describes socially prescribed impulse control that facilitates task- and goal- directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks" [85].

iii) Extroversion: "implies an energetic approach to the social and material world and includes traits such as sociability, activity, assertiveness, and positive emotionality" [85].

iv) Agreeableness: "contrasts a prosocial and communal orientation toward others with antagonism and includes traits such as altruism, tender-mindedness, trust, and modesty" [85].

v) Neuroticism: shows the level of emotional instability of the person, “neuroticism contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad, and tense" [85].

From the perspective of the interpersonal interaction, the way how we perceive our interaction partners can be dictated by the social cues. Fogg et. al have proposed 5 social cues that affect to our perception of the computer technology [59]:

i) *Physical*: face, eyes, body, movement;

ii) *Psychological*: preferences, humour, feelings, empathy;

iii) *Language*: interactive language use, spoken language, language recognition;

iv) *Social dynamics*: turn-taking, cooperation, praise, question answering, reciprocity;

v) *Social Roles*: doctor, teammate, opponent, teacher, pet, guide;

7.3 Approach and Method

Described above state-of-the-art suggests us that the use of conversational agents is feasible, and that reminiscence is an untapped and desirable application to explore. We also see that the success in previous studies was evident when agents displayed qualities attributed to humans and when proved useful to the task at hand – not surprisingly this was the case in agents operated by humans.

7.3.1 Studying the bot acceptance and percieved value

We formulate the following research questions:

RQ1. Is a chatbot acceptable (both to our target users and according to doctors) as an instrument for storytelling (reminiscing) and collecting information about the person?

RQ1.1 What are the necessary functionality that a storytelling bot should have from the elderly and NH staff perspective?

RQ2. Which context of chatbot use is considered safe and feasible by doctors and nursing home staff?

Method

We run a study to explore and evaluate: i) the feasibility and acceptance of a reminiscence-based storytelling conversational agent with different stakeholders: elderly, nursing home staff and gerontology doctors, ii) perception of the reminiscence chatbot activities (functionalities) and expected benefits, iii) preferred context of use of the tool by the stakeholders. The goal of conversational agent is conducting reminiscence sessions and collecting relevant information about elderly users,

Design Concept

The study was structured as sessions with users (older adults and NH staff) in which we used the storyboard to present the concept, main supported by the chatbot activities and expected benefits. The storyboard of the concept (Figure 7.1) has resulted from the visits and interviews in nursing homes (in Ticino, Switzerland) and Gerontology department of EOC hospital (Mendrisio, Switzerland).

The storyboard presents several activities to be performed with the storytelling bot:

i) Digitizing pictures and collecting personal memories.

Family members can provide pictures related to the elderly relative to create an archive of old pictures that can be later used during storytelling sessions and help elderly to recall memories.

ii) Driving the reminiscence sessions according to your memories and interests.

As the tool learns more about you, your personal memories, and interests, it can better adapt the conversations to make them more engaging, and collect your memories more effectively (including those you might have forgotten).

iii) Making conversations with your family and friends more engaging.

The tool can also be used in the joint reminiscence sessions between you and your friends and family, in this case supporting the conversations by suggesting questions and content, making the conversations more fun for you and them.

iv) Exploring your interests.

By suggesting and showing the multimedia content that might be related to your hobbies, crafts, favourite artists from TV and music (recommending & showing relevant content - music, pictures, videos).

v) Keeping cognitively active.

Reminiscing is known to bring mental and cognitive benefits as well as storytelling.

vi) Connecting and reconnecting with friends.

The stories and tags on the pictures are used to find common life points among elderly users. We might suggest potential friendship connections for the users based on their profiles info.

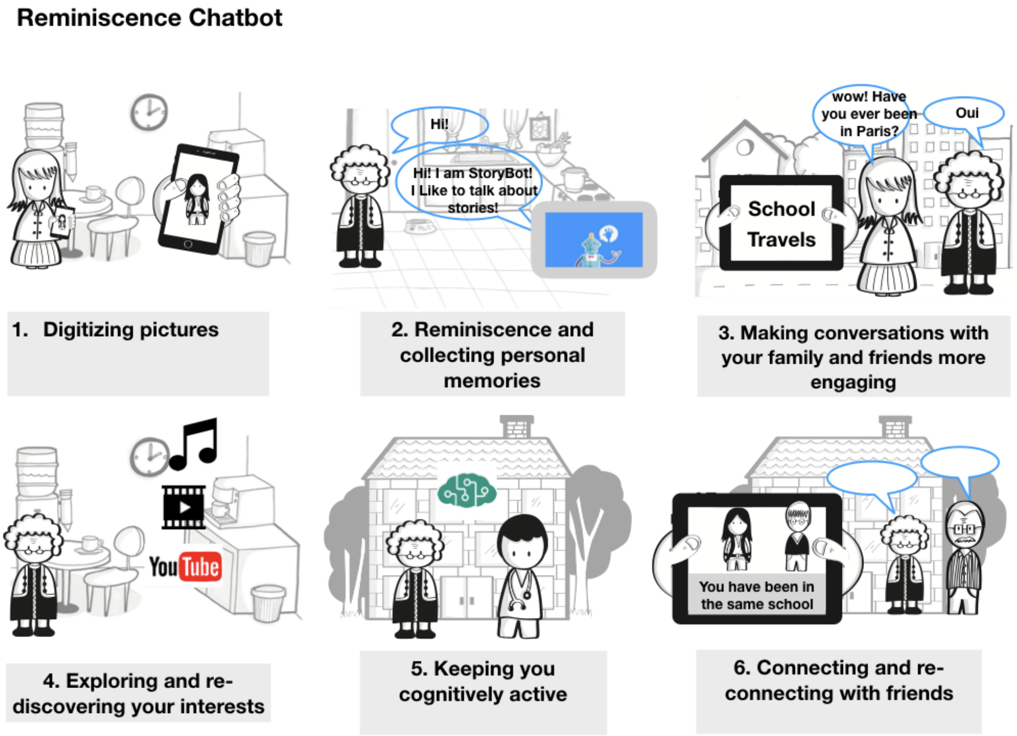


Fig. 7.1 Storyboard with the chatbot concept explained

Study Design

In the first phase of study we have interviewed older adults and nursing home staff in the nursing home and elderly care center in Trento, Italy ("Centro diurno anziani Povo" and "Casa di riposo Margherita Grazioli"). During the interviews the storyboard explaining the chatbot concept, main activities (functionalities) and expected benefits of the chatbot has been presented. The presentation of the chatbot and the interviews has been conducted by us, two researchers (one with IT and one with sociology background). All participants have signed an Informed consent form for the participation in the study. With the permission of the study participants, each interview was audio-recorded for the research purposes.

During semi-structured interviews participants were asked to provide a feedback on the ability and willingness to participate in the reminiscence sessions driven by the chatbot. When presented each chatbot activity in the detail, we have asked participants for the feedback on the ability and willingness to perform each activity and the perceived value of them (rated on the 5 -point Likert scale). Qualitative comments of the users have been collected during the sessions in the written notes and also transcribed from the audio recordings of the interviews.

The Storyboard presented during interviews can be seen in the Figure 7.1.

In the second phase, we have interviewed the neuropsychology doctors (gerontology department, EOC hospital, Switzerland) to evaluate the acceptance of the use of reminiscence conversational agent technology with elderly users from the gerontology doctors' perspective and collected qualitative feedback on what should be avoided and what is preferable for the chatbot-user interaction with a sensitive category of older adults. Interviews with doctors also started with the chatbot storyboard presentation, explanation of the chatbot concept, functionalities and expected benefits, and followed with semi-structured questionnarie (with 5-point Likert-scale ratings on each chatbot activity).

The conducted study is in the context of the bigger project that has been approved by the Ethical Committee of the University of Trento as research protocol 2017-003 on March and July 2017. All the study participants were asked for permission to record audio and signed an informed consent form.

Participants

For this study we have recruited 14 participants (4 elderly residents, 7 nursing home staff members, 3 doctors). To recruit participants, we contacted a reference person from the nursing home, who invited relatives of residents as well as staff members dedicated to animation or other related activities. Gerontology doctors from EOC hospital in Ticino, Switzerland were contacted by the researchers directly and invited for an interviews.

Measures

We evaluate all the presented chatbot's activities for feasibility and perceived value as agreement of the participant with the statements "I will be able to perform this activity on my own" and "This activity helps to accomplish the objectives described". Importance of the tool's objectives was rated on the 5-point Likert scale (from "not important" to "very important").

7.3.2 Chatbot personality and interaction behaviour

We formulate the following research questions: Chatbots can have different personalities (e.g., one can be more "chatty" or humorous or direct, and so on). Which personality - or set of personalities - are considered safe by doctors andcan be effective and engaging?

We try to find out how the set of personalities for such reminiscence chatbot should be designed (combining of the dimensions of bot's behaviour) and what are the acceptable options to be tested from the gerontology doctors perspective.

From the set of acceptable options of the reminiscence chatbot Personalities we aim to choose the most engaging and effective one. The chosen personality should be within the "space" of acceptable options, approved by the gerontology doctors and that are not potentially creating any stress to the patient.

When having a set of designed "personalities" for the storytelling bot we aim to test how elderly users perceive different personality options, if there is any difference in how the interaction perceived, and if so, what is the most *engaging* and *effective* chatbot personality.

Engagement in our case describes the level of involvement and interest of a person in the conversation. *Effectiveness* of the bot is defined as the amount and quality of information collected about the user during the interaction (per time unit and time while user engaged).

Method

We identify three dimensions based on which we plan to evaluate our storytelling chatbot: i) type of interaction ii) presence of humor and engaging comments and iii) response information.

The first two dimensions were inspired by the studies with young people (talkative chatbots which used humor were perceived positively in the studies with young adults). The third dimension is domain-specific as in our case one of the main goals of chatbot is collecting information about the person and life stories, the action of showing related content by the chatbot (text, pictures) might be a stimulating factor for the user to tell more information.

The description of proposed dimensions is provided below:

1) by the type of interaction:

- “following the form” or the form-based approach, where the WoZ conversation flow completely follows the pre-defined set of instances which chatbot needs to ask;

- “following the flow” or the flow-based approach, where the WoZ conversation flow completely follows to the direction which the user gives it, switching topics without much caring about the information to be collected;

- adding the chit-chat elements in the conversation but without routing out from the “filling form” plan, where "Wizard" still follows the form to be completed but meanwhile adds off-topic chit chat comments during the interaction.

2) by presence of humor :

- chatbot that doesn’t use the humoristic comments and jokes;

- chatbot that use the humor;

3) by response information :

Regarding the amount of information the bot tells back and providing to the user.

- the chatbot is only listening to the user without providing any feedback;

- the chatbot tells some content about the things mentioned by the user (e.g some facts about the city/place/year)

- the chatbot shows some multimedia content (picture and or music) about the things mentioned by the user (places/year).

Study design

To answer our research question RQ2 we conduct a study that consist of two parts, first i) the evaluation of the acceptable Personalities options will be done by the gerontology doctors based on the design options proposed by the researchers. In the second part of the study we will conduct Wizard of Oz (WoZ) reminiscence sessions with groups of participants guided by "Wizard" chatbot (a researcher that will follow several different pre-defined chatbot scripts, sitting in another room).

The "Wizard" behaviour will be defined by the combination of dimensions of bot behaviour (form- based vs flow based, with humor vs without humor use, just listening vs showing and telling related content). The baseline condition is the form-based chatbot, that is asking to the user predefined questions without commenting or making any jokes.

The WoZ experiment will be conducted as following. As a prototype of bot- interaction for the WoZ a modern chat system will be used (e.g Skype/Slack), running on the tablet/laptop with the "Wizard" behind the system generating messages. An output from the "bot-side" will be displayed text messages generated by the researcher on the other side of the chat system. An input from the participant will be audio replies (participant talking in open Skype audio call). For simplifying the interaction for the participant written "bot" replies will be displayed with big fonts on the screen, the participant will not need to type anything for the reply.

The time limit of one experiment session will be 20 minutes. After the session time passed the participant will not be interrupted until he finishes telling the life-story/memories.

Other direction that can be also explored if we have enough time and resources available is to interview in the similar way relatives of older adults (young adults age) as family members are potential co-participants of the reminiscence sessions. Preliminary interviews with gerontology doctors and elderly residents also showed that the preferred mode of reminiscence chatbot use is joint participation with the relative/nurse, therefore engagement of the relative can be an interesting aspect to explore.

Participants

Participants of the study are cognitively healthy older adults age over 65 years old. Participants will be contacted via local nursing home organizations in Trento, Italy. Included in the study participants should be retired, being at least one year of retirement at the time of the recruitment.

Measures

As an outcome we expect to have the analysis of how the dimensions of the chatbot behaviour affects to the i) engagement of the user and ii) amount of user information collected per session, and per time unit (information about the user that later can be used for recommending potential friends).

1. Information collected

The information collected about the user will be analyzed from two perspectives:

i) I/time unit - information collected per predefined time term;

ii) I/time engaged - information collected during the whole session while the user was engaged (until naturally finished).

We consider these two measures as we assume if the user is engaged he might be willing to tell more useful information which is an important outcome of reminiscence sessions for the NH staff that helps to learn more about the patient and his/her life.

The information we want to collect about the user will be pre-defined, we select 2 basic topics of the reminiscence sessions: Family composition and Hobbies. For each topic we have a list of data items to be asked: 5 items per topic (e.g city where born, if having siblings/kids).

Preliminary interviews and observations with the gerontology doctors showed that during the reminiscence sessions therapists are trying to have one or at maximum two topics per session as in this case concentration of the patient's attention is higher and it's easier for the patient to recall stories and facts and not get disturbed. While patient's hobbies and family composition are one of the main topics used during reminiscence session with therapists in the gerontology department, questions on this topics helps to learn a lot about the patient and his life story. Phrasings of questions will be taken from the list of "Life review questions" [185]. Topics-related information provided in the replies of the participants will be analyzed by the researchers manually.

2. Engagement

Engagement will be measured with ratings provided by the participants after each session and with qualitative analysis of the participant's comments provided regarding the interaction with WoZ. Ratings will be collected on: i) fun to use ii) frustration level iii) easy to use, iv) future use, v) overall rating for the persona (rated with Likert scale) [82].

7.4 Results

Chatbot feasibility and perceived value

We evaluate all the presented chatbot's activities for feasibility and perceived value as agreement of the participant with the statements "I will be able to perform this activity on my own" and "This activity helps to accomplish the objectives described". Importance of the tool's objectives was rated on the 5-point Likert scale (from "not important" to "very important").

All participants reported that older adults residents would be able to participate in the cognitive stimulation activities. Highly rated was activities such as: exploring interests, sharing conversations with relatives, and connecting with potential friends, which received positive and neutral ratings (with 90.91%, 81.82%, 72.73% positive rating respectively, were agree that activity can be performed by their own). For perceived value (Figure 7.2), (Figure 7.3) "exploring interests", "driving reminiscence sessions" and "stimulating cognitive activity" had only positive scores.

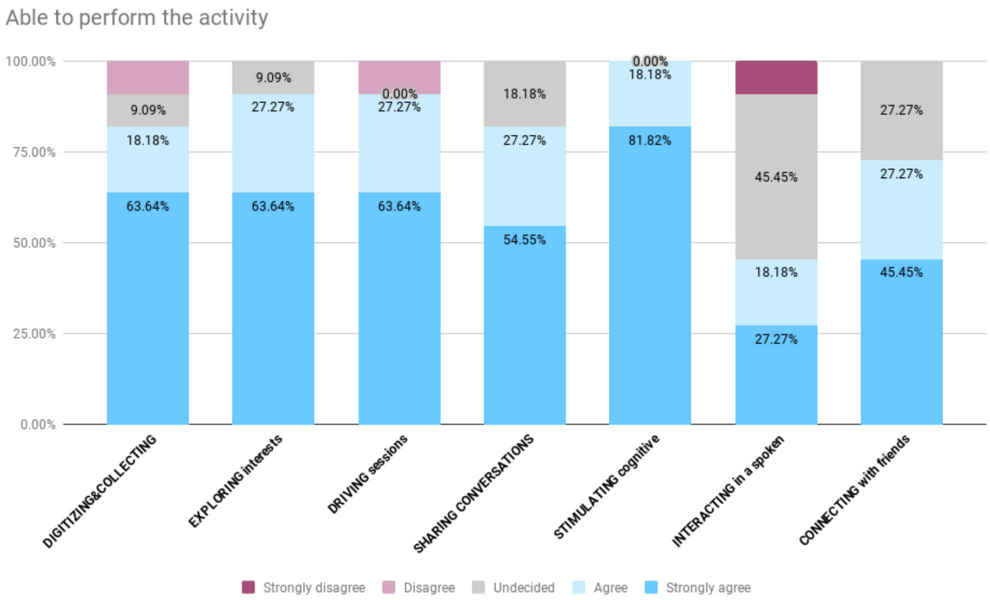


Fig. 7.2 Able to perform the activity

The most Highly rated activities were: i) Exploring interests ii) Sharing conversations with relatives/family members iii) Connecting with potential friends.

The only activities that received lower than neutral ratings regarding the ability to perform the activity : i) digitising pictures ii) driving sessions iii) interacting in a spoken language. In that cases participants gave a low rating for the context of use of the chatbot autonomously on their own, however were open to try it with the help of a relative/staff member.

Participants perceived the objectives addressed by the tool as important or absolutely important. This was the case particularly for "stimulating cognitive activity" objective (100%).

Context of use

The preferred context of use of such conversational agent according to both, doctors and the nursing home staff is the shared context of use - when an older adult participate in a reminiscence session guided by the bot together with the nursing home member (animator) or relative.

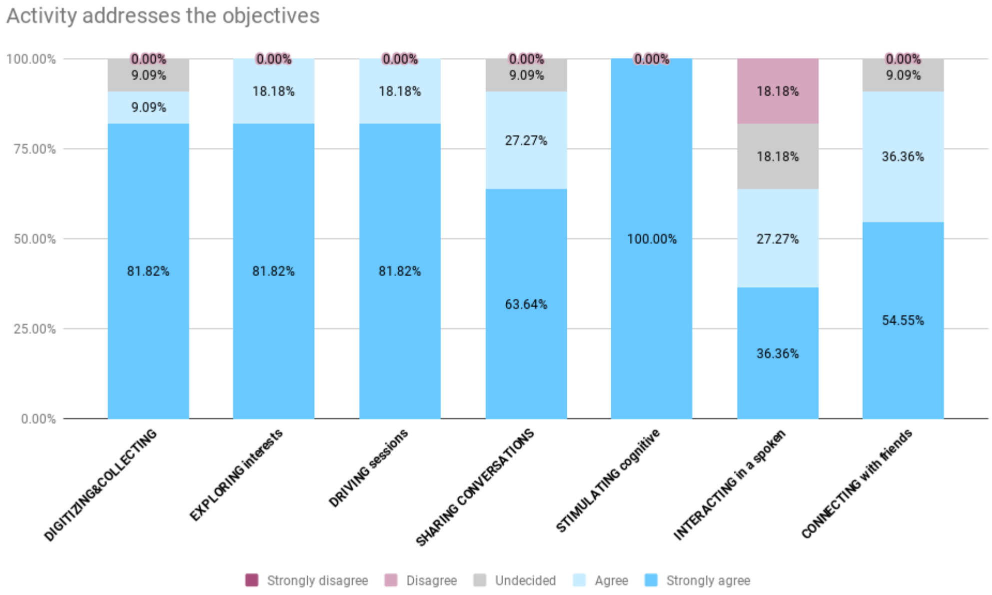


Fig. 7.3 Activity addresses the objectives

Comments of the Neuropsychologists doctors on the context of use: "The interaction with the bot should be done together with somebody, also to motivate the use of technology. Target users should be elderly 60-70 years old, who at least using smartphones, WhatsApp to communicate with family, send pictures and videos. Patients 70+ use tech but less, people 90+ do not use technologies usually." "Elderly need assistance using the chatbot".

Comments of NH staff members on the context of use: According to NH staff opinion, residents are not ready to talk to the chatbot directly. They recommend to use the chatbot mediated with someone, e.g, NH staff member or a relative: "At the moment an autonomous use is too difficult technologically for the elderly residents, better use it (chatbot) intermediated with an animator (staff member)".

Percieved value of activities

Below we provide a summary of qualitative comments from participants on the most highly rated activities.

i)"Exploring interests" was commented as:

by NH staff: NH Operators find exploring interests of the residents a useful part. They share that they already do this activity by themself trying *"to learn the interests of the resident he had also before he entered the RSA structure" ("interessi prima di entrare struttura").*

Operators of the NH conduct sessions of listening and discussing music, showing pictures of the region in groups *("facciamo musica therapia")*. Operators find it important as “some residents with age are losing their interests". "Music was used in the RSA structure also to stimulate the memories of residents".

ii) "Sharing conversations with family members and making it more engaging" was commented as:

by NH staff: NH members shared that in their experience relatives can be collaborative, e.g., relatives were bringing old photos of residents from home for participation in the NH internal project. "It (Sharing conversations) might help to reconnect and strengthen the ties with family". ("Questo (*making conversation with relative more interesting) Sarebbe un miracolo").

iii) "Connecting with potential friends" was commented as:

by NH staff: The idea of connecting and re-connecting potential friends in the settings of nursing home or elderly care center was percieved very positively: "yes, we try do it already now", "That would be very important". In the NH they had cases when elderly residents have re-connected in the NH after losing connection for many years (e.g., "vecchie vicini di casa" - people, who was neighbours or who worked on the same factory), as the "city moves here" (to the NH).

by Residents: Some residents were not tend to search for new friends in the nursing home settings ("non amo buttare via le mie cose"). Some residents instead were open to connect with others: "The more connections you can establish, the better. Different life experiences of another person" ("Piu se stabilisce connessioni .. e meglio, esperienza di vita diversa". "Se che questo simola e importante").

by Neuropsychologists:

Doctors commented that starting connection by themself for the residents is difficult, they need someone to intermediate: "Not sure. Elderly usually need motivation to start something, to connect, they need some intermediate person to stimulate - e.g., can be a nurse".

7.5 Discussion

In this chapter we presented the concept validation of the reminiscence- based conversational agent and its main activities.

The results describe a concept of the tool that is considered useful and that addresses objectives considered as very important by the participants. Results also suggest that for such category of users as older adults the most preferable context of use is shared context, when elderly participates in the reminiscence session together with the NH animator or family member. In this case a chatbot could be a helper for driving reminiscence sessions by suggesting topics for the discussions and storytelling, showing relevant multi-media material(pictures, audio) or questions to ask regarding an old archive picture. This model of shared interaction animator-chatbot- resident should be further explored and tested on the next versions of prototype.

Results also point to the feasibility and willingness of nursing home staff to play an active role, especially in what regards the co-located activities and shared context of use.

Chapter 8

Conclusion and future work

8.1 Contributions

In the first chapter we listed the different studies that were conducted over the years and that helped us build, incrementally, this research work. In this closing chapter we will outline our contributions and provide answers to the thesis research questions.

Our contributions can be summarized as following:

- Study with the tablet application enabling home-based training. From our studies, we identified the effect of social interactions on the motivation for exercising and technology use among elderly users in the high group cohesion settings and the absence of this effect in low cohesion group settings [121].
- Extensive review on the top home-based training applications available in 6 major app markets (i.e., Google PlayStore, Apple AppStore). The findings of this review indicated that social persuasion strategies can be effective for older adults [89].
- Review work on the factors affecting to friendship and feeling of connectedness between people. From the survey on facebook and literature review, we identified the potential of common aspects between people (common interests, joint activities) to increase connectedness. [142]
- A deep understanding of design challenges for tools aiming to encourage social interactions in residential care settings [120].
- Study on acceptance and perceived value of reminiscence conversational agent with stakeholders: elderly, nursing home staff, gerontology doctors.

The study findings highlight positive results on acceptance and importance of the chatbot tool rated by both elderly and doctors.

More specifically, in response to our thesis research questions.

To study the role of social interactions and social conditions in motivation for exercising and technology use among older adults **(RQ1)**.

We conducted two pilot studies with groups of older adults on studying remote physical exercising via Gymcentral application in Russia. A total of 44 older adults (59-83 years) participated in the two pilots and followed a training program for 8 weeks. We analyzed the app usage and the adherence to the training in different group cohesion settings (participants knowing/not knowing each other before the study) in the light of premeasures of social support, enjoyment of physical activity, and leg muscle strength.

One of the interesting study findings presented in the Chaper 2 was that participants who knew each other before the experiments (high group cohestion) and had some common interests (e.g participated together in the same retirement/leisure organizations) were more socially active during intervention and tend to communicate more via the app with other participants compared to people who didn't know each other before the experiment [121].

Another important result was that in the Tomsk studies online group-exercising (social condition) did not result in higher adherence when compared with individual training (with persuasion features). We didn't observe the same affect as in the previous study conducted with italian older adults in Trento where social group condition had higher adherence to the training. This results can be explained by: i) weaker cohesion among participants in Tomsk2 study, which might have reduced the effect of normative influence and peer support, ii) persuasion features in the individual training condition that raised the adherence compared with our previous study in Italy [51].

This results adds to the evidence that group-exercising in low cohesion groups results in an adherence comparable to that of individual training with contact (with a coach). It also partially supports the evidence that group exercising in high-cohesion (where participants know each other and/or share common interests) groups results in higher adherence than individual training with contact.

Results of the study showed us a potential of social interactions and social support factors in motivating older adults for exercising and use of technology in the conditions when participants are familiar with each other and/or sharing common interests. The details of the study, analysis and results are presented in Chapter 2.

To study how common life points are related to frequency of interactions, and how connectedness is related to common life points and interactions **(RQ2, RQ3)**.

In the survey study presented in the Chapter 3 we found that people who interact more often have higher levels of connectedness [142]. People having more common life points (shared beliefs, activities, and interests) also have higher levels of connectedness and in addition, online interactions are more frequent when people have more aspects in common.

After exploring factors the are correlated with people's connectedness and friendship, we focus on studying friendship formation of people in later life, and specifically on how technology can help to collect information about older adults that can be useful for facilitating friendship.

To determine the feasibility and acceptance of the chatbot (for target users and according to doctors) as an instrument for storytelling (reminiscing) and collecting information about the person **(RQ4)**.

We present the Conceptual model of reminiscence chatbot, a smart conversational agent that can drive personal and social reminiscence/storytelling sessions with older adults in the Chapter 4. The conceptual model was informed by the interviews and observations of reminiscence practices conducted in nursing home and gerontology department of EOC hospital.

In the Chapter 7 we present a study conducted with different stake-holders: elderly, nursing home staff (Trentino, Italy) and gerontology department doctors (EOC Mendrisio Switzerland) to evaluate: i) the feasibility and acceptance of a reminiscence- based storytelling conversational agent ii) perception of the reminiscence tool activities/functionalities and expected benefits.

The study was structured as sessions with users (older adults and NH staff) in which we used a storyboard to present the concept, main supported by the chatbot activities and expected benefits. Several main activities were defined and rated by the participants: digitizing pictures, collecting stories, guiding storytelling in engaging way, exploring interests, connecting and re-connecting with potential friends.

In the second part of the study we have interviewed the neuropsychology doctors (gerontology department, EOC hospital, Switzerland) to evaluate the acceptance of the use of reminiscence conversational agent technology with elderly users from the gerontology doctor's perspective and collected qualitative feedback on what should be avoided and what is preferable for the chatbot-user interaction with a sensitive category of older adults.

8.2 Limitations

Our work has a set of limitations, which we consider as areas for future improvement. These limitations are discussed below.

- *Group size.* One of the limitations of this work is the inclusion of few older adults in the studies. Although we must note that some of our participants were over 80 years old (Gymcentral studies described in the Chapter 2); and that we also included older adults who are NH residents in our concept validation study. This is in part due to the difficulty to recruit older adult participants and also because some activities conducted to validate the chatbot concept could be difficult to grasp for participants of this category.

- *Context.* Another limitation is that the results of this work have to be interpreted in context. The study with Gymcentral application (Chapter 2) studying the effects of social interactions to the motivation for the exercising and app use has been conducted in Tomsk, Russia with local older adults and the results are valid for this cultural group. Social interactions might have greater or smaller effects in the other countries with different culture. For instance, previous gymcentral studies conducted by our research group in Trento with italian older adults showed higher effect of social interactions. More studies should be done to explore the effects of social interactions in different cultural context and group cohesion settings.

  The chatbot concept validation studies were conducted in nursing homes in Northern Italy and Ticino region of Switzerland but for instance, challenges and preferences might differ in other countries.

- *Age category.* The results of our survey work (Chapter 3) investigating the factors affecting the feeling of connectedness and frequency of interactions conducted on Facebook should be interpreted taking into account the category of participants: facebook users, age group 32-65 years old.

  More research studies might help to explore better the effects of common life points and values to connectedness between elderly category of people and which common life points/values play grater role especially in the nursing home settings .

- *Platform.* The other limitation of the review work (Chapter 3) is in the type of meta data available on Facebook that has been analyzed to derive the common life points of participants.

8.3 Future work

The ongoing work is focused on the designing the refined prototype based on the findings of the concept validation studies and feedback of the stakeholders. The prototype and its main features should be further tested for usability with elderly residents and nursing home staff. The type of interaction - voice and use of interactive elements suitable or each chatbot activity should be tested.

An interesting direction to study from the hci perspective could be analysing which type of multimedia content (pictures, or music) and on which topics can be the most effective to engage elderly in a reminiscence storytelling. From our studies described in the Chapter 7 we have observed that not only pictures but also music and video content is used by the NH staff to stimulate storytelling sessions in the nursing home settings.

Another possible future work can be also done towards the direction of dialogue design and identifying patterns for the reminiscence-focused conversations. To act a fruitful and engaging conversation session, it is important to initiate a dialog with the right first question to a user. This becomes even more challenging for reminiscence therapy. First, because usually, we do not know enough facts about the patient's life (e. g, hobbies, family). Second, because elder people might be sensitive, introverted, and we would like to initiate a friendly conversation from its very beginning. Therefore, increasing the chance of having a successful and engaging reminiscence session.

In future work, we aim at investigating what is the best suitable first question to ask given a user's photo from the past. Working with only doctors notoriously limits experiments to a low scale. However, we can cope with this problem by adopting crowdsourcing microtasking. It can be realized in two stages. The first step is to run a crowd-based experiment, where workers will provide possible questions to initiate a conversation on a given photo. The second step is given the crowdsourced questions on a set of photos, to analyze and identify possible patterns among questions and content in photos. Thus, we might come up with a range of questions that are more suitable to some groups of content depicted in photos. Afterward, we will discuss our findings with doctors that will enable adjusting the results.

References

[1] Aalbers, T., Baars, M., and Rikkert, M. O. (2011). Characteristics of effective internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. *Ageing research reviews*, 10(4):487–497.

[2] Abbott, P. and Sapsford, R. (2006). Life-satisfaction in post-soviet russia and ukraine. *Journal of Happiness Studies*, 7(2):251–287.

[3] Acree, L. S., Longfors, J., Fjeldstad, A. S., Fjeldstad, C., Schank, B., Nickel, K. J., Montgomery, P. S., and Gardner, A. W. (2006). Physical activity is related to quality of life in older adults. *Health and quality of life outcomes*, 4(1):1.

[4] Ahtinen, A., Mattila, E., Vaatanen, A., Hynninen, L., Salminen, J., Koskinen, E., and Laine, K. (2009). User experiences of mobile wellness applications in health promotion: User study of wellness diary, mobile coach and selfrelax. In *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–8. IEEE.

[5] Allen, R. S. (2009). The legacy project intervention to enhance meaningful family interactions: Case examples. *Clinical Gerontologist*, 32(2):164–176.

[6] Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4):596–612.

[7] Astell, A., Alm, N., Gowans, G., Ellis, M., Dye, R., Campbell, J., and Vaughan, P. (2009). Working with people with dementia to develop technology: The circa and living in the moment projects. *PSIGE Newsletter*, 64.

[8] Atienza, A. A. (2001). Home-based physical activity programs for middle-aged and older adults: Summary of empirical research. *Journal of Aging and Physical Activity*.

[9] Baez, M., Dalpiaz, C., Hoxha, F., Tovo, A., Caforio, V., and Casati, F. (2016a). Personalized persuasion for social interactions in nursing homes. In Orji, R., Reisinger, M., Busch, M., Dijkstra, A., Stibe, A., and Tscheligi, M., editors, *Proceedings of the Personalization in Persuasive Technology Workshop, Persuasive Technology 2016*, pages 101–106.

[10] Baez, M., Ibarra, F., Far, I. K., Ferron, M., and Casati, F. (2016b). Online group-exercises for older adults of different physical abilities. In *2016 International Conference on Collaboration Technologies and Systems (CTS)*, pages 524–533.

[11] Ballantyne, A., Trenwith, L., Zubrinich, S., and Corlis, M. (2010). 'I feel less lonely': what older people say about participating in a social networking website. *Quality in Ageing and Older Adults*, 11(3):25–35.

[12] Beaglehole, R., Ebrahim, S., Reddy, S., Voute, J., Leeder, S., et al. (2008). Prevention of chronic diseases: a call to action. *The Lancet*, 370(9605):2152–2157.

[13] Beaton, D. E., Bombardier, C., Guillemin, F., and Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24):3186–3191.

[14] Behera, B. (2016). Chappie-a semi-automatic intelligent chatbot.

[15] Bender, M., Bauckham, P., and Norris, A. (1998). *The therapeutic purposes of reminiscence*. Sage.

[16] Bentley, F. R., Basapur, S., and Chowdhury, S. K. (2011). Promoting intergenerational communication through location-based asynchronous video communication. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 31–40. ACM.

[17] Berkman, L. F. and Syme, S. L. (1979). Social networks, host resistance, and mortality: A nine-year follow-up study of alameda county residents. *American Journal of Epidemiology*, 109(2):186–204.

[18] Bernard, M.-M., Fruhwirth, M., Brooks, M., Oakley, K., Wang, X., Ouechni, K., and Janson, F. (2011). Intergenerational telementoring for the promotion of social relationships. *Gerontechnology*, 10(1):38–50.

[19] Bird, M.-L., Hill, K. D., Robertson, I. K., Ball, M. J., Pittaway, J., and Williams, A. D. (2012). Serum [25 (oh) d] status, ankle strength and activity show seasonal variation in older adults: relevance for winter falls in higher latitudes. *Age and ageing*, page afs067.

[20] Blažun, H., Saranto, K., and Rissanen, S. (2012). Impact of computer training courses on reduction of loneliness of older people in finland and slovenia. *Computers in Human Behavior*, 28(4):1202 – 1212.

[21] Bower, B. (1997). Social links may counter health risks. *Science News*, 152(9):135–135.

[22] Bradshaw, S. A., Playford, E. D., and Riazi, A. (2012). Living well in care homes: a systematic review of qualitative studies. *Age and Ageing*, 41(4):429–440.

[23] Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

[24] Brox, E., Luque, L. F., Evertsen, G. J., and Hernández, J. E. G. (2011). Exergames for elderly: Social exergames to persuade seniors to increase physical activity. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 546–549. Ieee.

[25] Brummett, B. H., Barefoot, J. C., Siegler, I. C., Clapp-Channing, N. E., Lytle, B. L., Bosworth, H. B., Williams, R. B. J., and Mark, D. B. (2001). Characteristics of socially isolated patients with coronary artery disease who are at elevated risk for mortality. *Psychosomatic Medicine*, 63(2):267–272.

[26] Burke, S. M., Carron, A. V., Eys, M. A., Ntoumanis, N., and Estabrooks, P. A. (2006). Group versus individual approach? a meta-analysis of the effectiveness of interventions to promote physical activity. *Sport and Exercise Psychology Review*, 2(1):19–35.

[27] Caforio, V., Baez, M., and Casati, F. (2017). Viability of magazines for stimulating social interactions in nursing homes. In *IFIP Conference on Human-Computer Interaction*, pages 72–81. Springer.

[28] Callejas, Z., López-Cózar, R., Ábalos, N., and Griol, D. (2011). Affective conversational agents: the role of personality and emotion in spoken interactions. In *Conversational agents and natural language interaction: Techniques and effective practices*, pages 203–222. IGI Global.

[29] Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., Armour, C., and McTear, M. (2018). Best practices for designing chatbots in mental healthcare–a case study on ihelpr. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference*, page 129. BCS Learning & Development Ltd.

[30] Campbell, K., Holderness, N., and Riggs, M. (2015). Friendship chemistry: An examination of underlying factors. *The Social Science Journal*, 52(2):239 – 247.

[31] Cattan, M., Kime, N., and Bagnall, A.-M. (2011). The use of telephone befriending in low level support for socially isolated older people – an evaluation. *Health & Social Care in the Community*, 19(2):198–206.

[32] Cattan, M., White, M., Bond, J., and Learmouth, A. (2005). Preventing social isolation and loneliness among older people: a systematic review of health promotion interventions. *Ageing and Society*, 25(1):41–67.

[33] Cervone, A., Tortoreto, G., Mezza, S., Gambi, E., Riccardi, G., et al. (2017a). Roving mind: a balancing act between open–domain and engaging dialogue systems. *Alexa Prize Proceedings*.

[34] Cervone, A., Tortoreto, G., Mezza, S., Gambi, E., Riccardi, G., et al. (2017b). Roving mind: a balancing act between open–domain and engaging dialogue systems. In *Alexa Prize*, volume 1. https://developer. amazon. com/alexaprize/proceedings.

[35] Chen, J., Geyer, W., Dugan, C., Muller, M., and Guy, I. (2009). Make new friends, but keep the old: Recommending people on social networking sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 201–210, New York, NY, USA. ACM.

[36] Chen, Y.-R. R. and Schulz, P. J. (2016). The effect of information communication technology interventions on reducing social isolation in the elderly: A systematic review. *Journal of Medical Internet Research*, 18(1):e18.

[37] Chi-Wai, R. K., Sai-Chuen, S. H., So-Ning, T. M., Ka-Shun, P. W., Wing-Kuen, K. L., and Choi-Ki, C. W. (2011). Can mobile virtual fitness apps replace human fitness trainer? *The 5th International Conference on New Trends in Information Science and Service Science*, 1:56–63.

[38] Chkroun, M. and Azaria, A. (2018). "did i say something wrong?": Towards a safe collaborative chatbot.

[39] Choi, M., Kong, S., and Jung, D. (2012a). Computer and internet interventions for loneliness and depression in older adults: a meta-analysis. *Healthcare informatics research*, 18(3):191–198.

[40] Choi, M., Kong, S., and Jung, D. (2012b). Computer and internet interventions for loneliness and depression in older adults: A meta-analysis. *Healthcare Informatics Research*, 18(3):191–198.

[41] Cornwell, E. Y. and Waite, L. J. (2009). Social disconnectedness, perceived isolation, and health among older adults. *Journal of Health and Social Behavior*, 50(1):31–48. PMID: 19413133.

[42] Crete-Nishihata, M., Baecker, R. M., Massimi, M., Ptak, D., Campigotto, R., Kaufman, L. D., Brickman, A. M., Turner, G. R., Steinerman, J. R., and Black, S. E. (2012). Reconstructing the past: personal memory technologies are not just personal and not just for memory. *Human–Computer Interaction*, 27(1-2):92–123.

[43] de Groot, G. C. L. and Fagerström, L. (2011). Older adults' motivating factors and barriers to exercise to prevent falls. *Scandinavian Journal of Occupational Therapy*, 18(2):153–160.

[44] Demiris, G., Thompson, H. J., Lazar, A., and Lin, S.-Y. (2016). Evaluation of a digital companion for older adults with mild cognitive impairment. In *AMIA Annual Symposium Proceedings*, volume 2016, page 496. American Medical Informatics Association.

[45] Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, 49(1):71–75.

[46] Dishman, R. K., Sallis, J. F., and Orenstein, D. R. (1985). The determinants of physical activity and exercise. *Public health reports*, 100(2):158.

[47] Edelbrock, D., Buys, L., Creasey, H., and Broe, G. (2001). Social support, social networks and social isolation: the sydney older persons' study. *Australasian Journal on Ageing*, 20(4):173–8.

[48] Edmondson, B. (2010). All the lonely people. *AARP: The Magazine*, page 83.

[49] El-Khoury, F., Cassou, B., Charles, M.-A., and Dargent-Molina, P. (2013). The effect of fall prevention exercise programmes on fall induced injuries in community dwelling older adults: systematic review and meta-analysis of randomised controlled trials.

[50] Elder, A. J. S., Scott, W. S., Kluge, M. A., and Elder, C. L. (2016). Cyberex internet-based group exercise for rural older adults: A pilot study. *Activities, Adaptation & Aging*, 40(2):107–124.

[51] Far, I. K., Ferron, M., Ibarra, F., Baez, M., Tranquillini, S., Casati, F., and Doppio, N. (2015). The interplay of physical and social wellbeing in older adults: investigating the relationship between physical training and social interactions with virtual social environments. *PeerJ Computer Science*, 1:e30.

[52] Far, I. K., Silveira, P., Casati, F., and Baez, M. (2012). Unifying platform for the physical, mental and social well-being of the elderly. In *Embedded and Multimedia Computing Technology and Service*, pages 385–392. Springer.

[53] Fausset, C. B., Mitzner, T. L., Price, C. E., Jones, B. D., Fain, B. W., and Rogers, W. A. (2013). Older adults' use of and attitudes toward activity monitoring technologies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 1683–1687. SAGE Publications Sage CA: Los Angeles, CA.

[54] Fehr, B. (2008). Friendship formation. In Sprecher, S., Wenzel, A., and Harvey, J., editors, *Handbook of relationship initiation*, chapter 2, pages 29–54. Taylor & Francis.

[55] Findlay, R. A. (2003). Interventions to reduce social isolation amongst older people: where is the evidence? *Ageing & Society*, 23(5):647–658.

[56] Fitzpatrick, K. K., Darcy, A., and 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.

[57] Fivush, R., Habermas, T., Waters, T. E., and Zaman, W. (2011). The making of autobiographical memory: Intersections of culture, narratives and identity. *International Journal of Psychology*, 46(5):321–345.

[58] Fjeldsoe, B. S., Miller, Y. D., and Marshall, A. L. (2010). Mobilemums: a randomized controlled trial of an sms-based physical activity intervention. *Annals of Behavioral Medicine*, 39(2):101–111.

[59] Fogg, B. J. (2002). Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5.

[60] Fokkema, D. T. and Knipscheer, K. (2007). Escape loneliness by going digital: A quantitative and qualitative evaluation of a dutch experiment in using ect to overcome loneliness among older adults. *Aging & Mental Health*, 11(5):496–504. PMID: 17882587.

[61] Fratiglioni, L., Wang, H.-X., Ericsson, K., Maytan, M., and Winblad, B. (2000). Influence of social network on occurrence of dementia: a community-based longitudinal study. *The Lancet*, 355(9212):1315–1319.

[62] Fuketa, M., Morita, K., and Aoe, J.-i. (2013). Agent–based communication systems for elders using a reminiscence therapy. *International Journal of Intelligent Systems Technologies and Applications*, 12(3-4):254–267.

[63] Gardner, I., Brooke, E., Ozanne, E., and Kendig, H. (1999). Improving social networks, a research report: improving health and social isolation in the australian veteran community. *Lincoln Gerontology Centre, La Trobe University, Melbourne.[Google Scholar]*.

[64] Gardner, M. M., Buchner, D. M., Robertson, M. C., and Campbell, A. J. (2001a). Practical implementation of an exercise-based falls prevention programme. *Age and ageing*, 30(1):77–83.

[65] Gardner, M. M., Buchner, D. M., Robertson, M. C., and Campbell, A. J. (2001b). Practical implementation of an exercise-based falls prevention programme. *Age and Ageing*, 30(1):77–83.

[66] Hanke, S., Sandner, E., Kadyrov, S., and Stainer-Hochgatterer, A. (2016). Daily life support at home through a virtual support partner.

[67] Hanneton, S. et al. (2009). Coaching the wii. In *2009 IEEE International Workshop on Haptic Audio visual Environments and Games*, pages 54–57. IEEE.

[68] Hartup, W. W. and Stevens, N. (1999). Friendships and adaptation across the life span. *Current directions in psychological science*, 8(3):76–79.

[69] Harvey, J. A., Chastin, S. F., and Skelton, D. A. (2013). Prevalence of sedentary behavior in older adults: a systematic review. *International journal of environmental research and public health*, 10(12):6645–6661.

[70] Hawthorn, D. (2000). Possible implications of aging for interface designers. *Interacting with computers*, 12(5):507–528.

[71] Heikkinen, R.-L. and Kauppinen, M. (2004). Depressive symptoms in late life: a 10-year follow-up. *Archives of Gerontology and Geriatrics*, 38(3):239 – 250.

[72] Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., and Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, 10(2):227–237. PMID: 25910392.

[73] Huang, T.-H. K. (2016). Crowd-powered conversational agents.

[74] Huang, T.-H. K., Lasecki, W. S., Azaria, A., and Bigham, J. P. (2016). " is there anything else i can help you with?" challenges in deploying an on-demand crowd-powered conversational agent. In *Fourth AAAI Conference on Human Computation and Crowdsourcing*.

[75] Huang, T.-H. K., Lasecki, W. S., and Bigham, J. P. (2015). Guardian: A crowd-powered spoken dialog system for web apis. In *Third AAAI conference on human computation and crowdsourcing*.

[76] Hughes, M. E., Waite, L. J., Hawkley, L. C., and Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys results from two population-based studies. *Research on aging*, 26(6):655–672.

[77] Huldtgren, A., Mertl, F., Vormann, A., and Geiger, C. (2015). Probing the potential of multimedia artefacts to support communication of people with dementia. In *Human-Computer Interaction*, pages 71–79. Springer.

[78] Ibarra, F., Baez, M., Fiore, F., and Casati, F. (2017). Stimulating conversations in residential care through technology-mediated reminiscence. In *IFIP Conference on Human-Computer Interaction*, pages 62–71. Springer.

[79] Ibarra, F., Korovina, O., Baez, M., Casati, F., Marchese, M., Cernuzzi, L., and Barysheva, G. A. (2016). Tools enabling online contributions by older adults. *IEEE Internet Computing*, 20(5):58–65.

[80] IJsselsteijn, W., De Kort, Y., Bonants, R., De Jager, M., and Westerink, J. (2004). Virtual cycling: Effects of immersion and a virtual coach on motivation and presence in a home fitness application. In *Proceedings Virtual Reality Design and Evaluation Workshop*, pages 22–23.

[81] Ijsselsteijn, W., Nap, H. H., de Kort, Y., and Poels, K. (2007). Digital game design for elderly users. In *Proceedings of the 2007 conference on Future Play*, pages 17–22. ACM.

[82] Jain, M., Kumar, P., Kota, R., and Patel, S. N. (2018). Evaluating and informing the design of chatbots. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*, pages 895–906. ACM.

[83] Jeurens, J., van Turnhout, K., and Bakker, R. (2014). Increasing family involvement in elderly care. In *International Conference of Design, User Experience, and Usability*, pages 403–411. Springer.

[84] Jiang, Y., Kummerfeld, J. K., and Laseck, W. S. (2017). Understanding task design trade-offs in crowdsourced paraphrase collection. *arXiv preprint arXiv:1704.05753*.

[85] John, O. P., Srivastava, S., et al. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138.

[86] Jones, C. J., Rikli, R. E., and Beam, W. C. (1999). A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. *Research quarterly for exercise and sport*, 70(2):113–119.

[87] Kendzierski, D. and DeCarlo, K. J. (1991). Physical activity enjoyment scale: two validation studies. *Journal of Sport & Exercise Psychology*, 13(1).

[88] Khaghani-Far, I., Nikitina, S., Báez, M., Taran, E. A., and Casati, F. (2016a). Fitness applications for home-based training. *IEEE Pervasive Computing*, 15(4):56–65.

[89] Khaghani-Far, I., Nikitina, S., Baez, M., Taran, E. A., and Casati, F. (2016b). Fitness applications for home-based training. *IEEE Pervasive Computing*, 15(4):56–65.

[90] Khoo, E. T., Cheok, A. D., Nguyen, T. H. D., and Pan, Z. (2008). Age invaders: social and physical inter-generational mixed reality family entertainment. *Virtual Reality*, 12(1):3–16.

[91] Kim, J. and Zimmerman, J. (2006). Cherish: smart digital photo frames for sharing social narratives at home. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pages 953–958. ACM.

[92] Koivula, N. (1999). Sport participation: Differences in motivation and actual. *Journal of Sport Behavior*, 22(3).

[93] Kopp, S., Gesellensetter, L., Krämer, N. C., and Wachsmuth, I. (2005). A conversational agent as museum guide–design and evaluation of a real-world application. In *International Workshop on Intelligent Virtual Agents*, pages 329–343. Springer.

[94] Kozma, A., Stones, M. J., and McNeil, J. K. (1991). *Psychological well-being in later life*. Harcourt Brace.

[95] Kucherbaev, P., Bozzon, A., and Houben, G.-J. (2018). Human aided bots. *IEEE Internet Computing*.

[96] Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 591–600, New York, NY, USA. ACM.

[97] Landi, F., Abbatecola, A. M., Provinciali, M., Corsonello, A., Bustacchini, S., Manigrasso, L., Cherubini, A., Bernabei, R., and Lattanzio, F. (2010). Moving against frailty: does physical activity matter? *Biogerontology*, 11(5):537–545.

[98] Lasecki, W. S., Wesley, R., Nichols, J., Kulkarni, A., Allen, J. F., and Bigham, J. P. (2013). Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 151–162. ACM.

[99] Lazar, A., Thompson, H., and Demiris, G. (2014). A systematic review of the use of technology for reminiscence therapy. *Health education & behavior*, 41(1_suppl):51S–61S.

[100] Lazarsfeld, P. F. and Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. In Berger, M., Abel, T., and Page, C., editors, *Freedom and Control in Modern Society*, pages 18–66. Van Nostrand, New York.

[101] Lee, H.-C., Cheng, Y. F., Cho, S. Y., Tang, H.-H., Hsu, J., and Chen, C.-H. (2014). Picgo: designing reminiscence and storytelling for the elderly with photo annotation. In *Proceedings of the 2014 companion publication on Designing interactive systems*, pages 9–12. ACM.

[102] Lee, H.-C. and Hsu, J. Y.-j. (2016). Picmemory: Enriching intergenerational family interaction and memory collection. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 3715–3718. ACM.

[103] Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., and Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using facebook.com. *Social Networks*, 30(4):330 – 342.

[104] Lin, L., D'Haro, L. F., and Banchs, R. (2016). A web-based platform for collection of human-chatbot interactions. In *Proceedings of the Fourth International Conference on Human Agent Interaction*, pages 363–366. ACM.

[105] López, A., Ratni, A., Trong, T. N., Olaso, J. M., Montenegro, S., Lee, M., Haider, F., Schlögl, S., Chollet, G., Jokinen, K., et al. (2016). Lifeline dialogues with roberta. In *International Workshop on Future and Emerging Trends in Language Technology*, pages 73–85. Springer.

[106] Machesney, D., Wexler, S. S., Chen, T., and Coppola, J. F. (2014). Gerontechnology companion: Virutal pets for dementia patients. In *IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014*, pages 1–3.

[107] McAuley, E., Konopack, J. F., Motl, R. W., Morris, K. S., Doerksen, S. E., and Rosengren, K. R. (2006). Physical activity and quality of life in older adults: influence of health status and self-efficacy. *Annals of behavioral Medicine*, 31(1):99–103.

[108] McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444.

[109] McTear, M., Callejas, Z., and Griol, D. (2016). Affective conversational interfaces. In *The Conversational Interface*, pages 329–357. Springer.

[110] Miron, A. M., Thompson, A. E., McFadden, S. H., and Ebert, A. R. (2017). Young adults' concerns and coping strategies related to their interactions with their grandparents and great-grandparents with dementia. *Dementia*, page 1471301217700965.

[111] Mitchell, M., Bohus, D., and Kamar, E. (2014). Crowdsourcing language generation templates for dialogue systems. *Proceedings of the INLG and SIGDIAL 2014 Joint Session*, pages 172–180.

[112] Mooney, R. J. and Roy, L. (2000). Content-based book recommending using learning for text categorization. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, DL '00, pages 195–204, New York, NY, USA. ACM.

[113] Morris, M. E., Adair, B., Ozanne, E., Kurowski, W., Miller, K. J., Pearce, A. J., Santamaria, N., Long, M., Ventura, C., and Said, C. M. (2014). Smart technologies to enhance social connectedness in older people who live at home. *Australasian Journal on Ageing*, 33(3):142–152.

[114] Moser, A., Stuck, A. E., Silliman, R. A., Ganz, P. A., and Clough-Gorr, K. M. (2012). The eight-item modified medical outcomes study social support survey: psychometric evaluation showed excellent performance. *Journal of clinical epidemiology*, 65(10):1107–1116.

[115] Mostafazadeh, N., Misra, I., Devlin, J., Mitchell, M., He, X., and Vanderwende, L. (2016). Generating natural questions about an image. *arXiv preprint arXiv:1603.06059*.

[116] Mubin, O., Shahid, S., and Al Mahmud, A. (2008). Walk 2 win: towards designing a mobile game for elderly's social engagement. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 2*, pages 11–14. BCS Learning & Development Ltd.

[117] Müller, A. M. and Khoo, S. (2014). Non-face-to-face physical activity interventions in older adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 11(1):1.

[118] Negi, S., Joshi, S., Chalamalla, A. K., and Subramaniam, L. V. (2009). Automatically extracting dialog models from conversation transcripts. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, pages 890–895. IEEE.

[119] Nick, B., Lee, C., Cunningham, P., and Brandes, U. (2013). Simmelian backbones: Amplifying hidden homophily in facebook networks. In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, pages 525–532.

[120] Nikitina, S., Callaioli, S., and Baez, M. (2018a). Smart conversational agents for reminiscence.

[121] Nikitina, S., Didino, D., Baez, M., and Casati, F. (2018b). Feasibility of virtual tablet-based group exercise among older adults in siberia: Findings from two pilot trials. *JMIR mHealth and uHealth*, 6(2):e40.

[122] Norman, E., Sherburn, M., Osborne, R. H., and Galea, M. P. (2010). An exercise and education program improves well-being of new mothers: a randomized controlled trial. *Physical Therapy*, 90(3):348–355.

[123] Nurgalieva, L., Laconich, J. J. J., Baez, M., Casati, F., and Marchese, M. (2017). Designing for older adults: review of touchscreen design guidelines. *arXiv preprint arXiv:1703.06317*.

[124] O'Brien, H. L. and Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69.

[125] of Sports Medicine, A. C. et al. (2013). *ACSM's guidelines for exercise testing and prescription*. Lippincott Williams & Wilkins.

[126] Ofli, F., Kurillo, G., Obdrzalek, S., Bajcsy, R., Jimison, H., and Pavel, M. (2015). Design and Evaluation of an Interactive Exercise Coaching System for Older Adults: Lessons Learned. *IEEE journal of biomedical and health informatics*, PP(99):1.

[127] Organization, W. H. (2015). World Health Organization. Global Health Observatory Data Repository: Life expectancy - Data by country. 2015.

[128] Park, H. L., O'Connell, J. E., and Thomson, R. G. (2003). A systematic review of cognitive decline in the general elderly population. *International journal of geriatric psychiatry*, 18(12):1121–1134.

[129] Perry, T. E. (2014). Seasonal variation and homes: understanding the social experiences of older adults. *Care Management Journals*, 15(1):3–10.

[130] Peter, C., Kreiner, A., Schröter, M., Kim, H., Bieber, G., Öhberg, F., Hoshi, K., Waterworth, E. L., Waterworth, J., and Ballesteros, S. (2013). Agnes: Connecting people in a multimodal way. *Journal on multimodal user interfaces*, 7(3):229–245.

[131] Phang, C. W., Sutanto, J., Kankanhalli, A., Li, Y., Tan, B. C., and Teo, H.-H. (2006). Senior citizens' acceptance of information systems: A study in the context of e-government services. *IEEE Transactions on Engineering Management*, 53(4):555–569.

[132] Phillips, E. M., Schneider, J. C., and Mercer, G. R. (2004). Motivating elders to initiate and maintain exercise. *Archives of physical medicine and rehabilitation*, 85:52–57.

[133] Piper, A. M., Weibel, N., and Hollan, J. (2013). Audio-enhanced paper photos: encouraging social interaction at age 105. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 215–224. ACM.

[134] Piper, A. M., Weibel, N., and Hollan, J. D. (2014). Designing audio-enhanced paper photos for older adult emotional wellbeing in communication therapy. *International Journal of Human-Computer Studies*, 72(8-9):629–639.

[135] Razavi, S. Z., Schubert, L. K., Kane, B., Ali, M. R., Van Orden, K., and Ma, T. (2019). Dialogue design and management for multi-session casual conversation with older adults. *arXiv preprint arXiv:1901.06620*.

[136] Richard, M., Christina, M. F., DEBORAH, L. S., Rubio, N., and KENNON, M. S. (1997). Intrinsic motivation and exercise adherence. *Int J Sport Psychol*, 28(4):335–354.

[137] Ring, L., Barry, B., Totzke, K., and Bickmore, T. (2013). Addressing loneliness and isolation in older adults: Proactive affective agents provide better support. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 61–66. IEEE.

[138] Ring, L., Shi, L., Totzke, K., and Bickmore, T. (2015). Social support agents for older adults: longitudinal affective computing in the home. *Journal on Multimodal User Interfaces*, 9(1):79–88.

[139] Rodríguez, M. D., Roa, J. R., Morán, A. L., and Nava-Muñoz, S. (2012). Persuasive strategies for motivating elders to exercise. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 219–223. IEEE.

[140] Russel, D., Peplau, L. A., and Cutrona, C. E. (1980). The revised ucla loneliness scale: Concurrent and discriminant validity evidence. *Journal of personality and social psychology*, 39(3):472–480.

[141] Sallis, J. F. and Saelens, B. E. (2000). Assessment of physical activity by self-report: status, limitations, and future directions. *Research quarterly for exercise and sport*, 71(sup2):1–14.

[142] Sanchiz, E., Ibarra, F., Nikitina, S., Báez, M., and Casati, F. (2016). What makes people bond?: A study on social interactions and common life points on facebook. In *Collaboration Technologies and Systems (CTS), 2016 International Conference on*, pages 26–30. IEEE.

[143] Schlögl, S., Doherty, G., and Luz, S. (2014). Wizard of oz experimentation for language technology applications: Challenges and tools. *Interacting with Computers*, 27(6):592–615.

[144] Seeman, T. E., Berkman, L. F., Blazer, D., and Rowe, J. W. (1994). Social ties and support and neuroendocrine function: the macarthur studies of successful aging. *Annals of Behavioral Medicine*, 16(2):95–106.

[145] Sherbourne, C. D. and Stewart, A. L. (1991). The mos social support survey. *Social science & medicine*, 32(6):705–714.

[146] Silveira, P., van de Langenberg, R., van het Reve, E., Daniel, F., Casati, F., and de Bruin, E. D. (2013). Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: a phase ii preclinical exploratory trial. *Journal of medical Internet research*, 15(8).

[147] Steffen, D., Bleser, G., Weber, M., Stricker, D., Fradet, L., and Marin, F. (2011). A personalized exercise trainer for elderly. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 24–31. IEEE.

[148] Steverink, N., Slaets, J., Schuurmans, H., and Van Lis, M. (2001). Measuring frailty: Developing and testing the gfi (groningen frailty indicator). *Gerontologist*, 41:236–237.

[149] Stickley, A., Koyanagi, A., Leinsalu, M., Ferlander, S., Sabawoon, W., and McKee, M. (2015). Loneliness and health in eastern europe: findings from moscow, russia. *Public health*, 129(4):403–410.

[150] Stuart, M., Chard, S., Benvenuti, F., and Steinwachs, S. (2008). Community exercise: a vital component to healthy aging. *Healthcarepapers*, 10(1):23–8.

[151] Subramaniam, P. and Woods, B. (2012). The impact of individual reminiscence therapy for people with dementia: systematic review. *Expert Review of Neurotherapeutics*, 12(5):545–555.

[152] Subramaniam, P. and Woods, B. (2016). Digital life storybooks for people with dementia living in care homes: an evaluation. *Clinical interventions in aging*, 11:1263.

[153] Széman, Z. (2014). A new pattern in long term care in hungary: Skype and youth volunteers. *Anthropological notebooks*, pages 105–117.

[154] Takahashi, T. and Yokono, H. (2017). Two persons dialogue corpus made by multiple crowd-workers. In *Proceedings of the 8th International Workshop on Spoken Dialogue Systems (IWSDS)*.

[155] Team, R. C. et al. (2013). R: A language and environment for statistical computing.

[156] Thibaud, M., Bloch, F., Tournoux-Facon, C., Brèque, C., Rigaud, A. S., Dugué, B., and Kemoun, G. (2012). Impact of physical activity and sedentary behaviour on fall risks in older people: a systematic review and meta-analysis of observational studies. *European Review of Aging and Physical Activity*, 9(1):5–15.

[157] Thies, I. M., Menon, N., Magapu, S., Subramony, M., and O'neill, J. (2017). How do you want your chatbot? an exploratory wizard-of-oz study with young, urban indians. In *IFIP Conference on Human-Computer Interaction*, pages 441–459. Springer.

[158] Thomas, W. H. (1996). *Life worth living: How someone you love can still enjoy life in a nursing home: The Eden Alternative in action*. VanderWyk & Burnham, Acton, MA.

[159] Thorgrimsdottir, S. H. and Bjornsdottir, K. (2016). Reminiscence work with older people: the development of a historical reminiscence tool. *International journal of older people nursing*, 11(1):70–79.

[160] Thórisson, K. and Jonsdottir, G. (2008). Towards a Reactive Virtual Trainer. *Intelligent Virtual Agents*, 5208:131–138.

[161] Tilvis, R., Routasalo, P., Karppinen, H., Strandberg, T., Kautiainen, H., and Pitkala, K. (2012). Social isolation, social activity and loneliness as survival indicators in old age; a nationwide survey with a 7-year follow-up. *European geriatric medicine*, 3(1):18–22.

[162] Topolski, T. D., LoGerfo, J., Patrick, D. L., Williams, B., Walwick, J., and Patrick, M. M. B. (2006). Peer reviewed: the rapid assessment of physical activity (rapa) among older adults. *Preventing chronic disease*, 3(4).

[163] Tran, K., He, X., Zhang, L., and Sun, J. (2016). Rich image captioning in the wild. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2016 IEEE Conference on*, pages 434–441. IEEE.

[164] Traud, A. L., Mucha, P. J., and Porter, M. A. (2012). Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165 – 4180.

[165] Tripette, J., Murakami, H., Gando, Y., Kawakami, R., Sasaki, A., Hanawa, S., Hirosako, A., and Miyachi, M. (2014). Home-based active video games to promote weight loss during the postpartum period. *Medicine & Science in Sports & Exercise*, 46(3):472–478.

[166] Tsiourti, C., Joly, E., Wings, C., Moussa, M. B., and Wac, K. (2014). Virtual assistive companions for older adults: qualitative field study and design implications. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pages 57–64. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[167] Tsiourti, C., Moussa, M. B., Quintas, J., Loke, B., Jochem, I., Lopes, J. A., and Konstantas, D. (2016a). A virtual assistive companion for older adults: design implications for a real-world application. In *Proceedings of SAI Intelligent Systems Conference*, pages 1014–1033. Springer.

[168] Tsiourti, C., Quintas, J., Ben-Moussa, M., Hanke, S., Nijdam, N. A., and Konstantas, D. (2016b). The cameli framework—a multimodal virtual companion for older adults. In *Proceedings of SAI Intelligent Systems Conference*, pages 196–217. Springer.

[169] Tucker, K. L., Ozer, D. J., Lyubomirsky, S., and Boehm, J. K. (2006). Testing for measurement invariance in the satisfaction with life scale: A comparison of russians and north americans. *Social Indicators Research*, 78(2):341–360.

[170] Tudor-Locke, C., Sisson, S. B., Lee, S. M., Craig, C. L., Plotnikoff, R. C., and Bauman, A. (2006). Evaluation of quality of commercial pedometers. *Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique*, pages S10–S15.

[171] Uriu, D., Shiratori, N., Hashimoto, S., Ishibashi, S., and Okude, N. (2009). Caraclock: an interactive photo viewer designed for family memories. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 3205–3210. ACM.

[172] Vagetti, G. C., Barbosa Filho, V. C., Moreira, N. B., Oliveira, V. d., Mazzardo, O., and Campos, W. d. (2014). Association between physical activity and quality of life in the elderly: a systematic review, 2000-2012. *Revista Brasileira de Psiquiatria*, 36(1):76–88.

[173] Van Der Bij, A. K., Laurant, M. G., and Wensing, M. (2002). Effectiveness of physical activity interventions for older adults: a review1 1the full text of this article is available via ajpm online at www. elsevier. com/locate/ajpmonline. *American journal of preventive medicine*, 22(2):120–133.

[174] van Dijk, B., Dadlani, P., van Halteren, A., and Biemans, M. (2010). Life changes, connection stays: photo sharing and social connectedness for people with special needs. In *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*, pages 135–142. ACM.

[175] Vardoulakis, L. P., Ring, L., Barry, B., Sidner, C. L., and Bickmore, T. (2012). Designing relational agents as long term social companions for older adults. In *International Conference on Intelligent Virtual Agents*, pages 289–302. Springer.

[176] Wankel, L. M. (1993). The importance of enjoyment to adherence and psychological benefits from physical activity. *International Journal of Sport Psychology*.

[177] Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., and Vanhees, L. (2010). Assessment of physical activity–a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the european association of cardiovascular prevention and rehabilitation. *European Journal of Cardiovascular Prevention & Rehabilitation*, 17(2):127–139.

[178] Watson, A., Bickmore, T., Cange, A., Kulshreshtha, A., and 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.

[179] Webster, J. D. (1993). Construction and validation of the reminiscence functions scale. *Journal of Gerontology*, 48(5):P256–P262.

[180] Webster, J. D., Bohlmeijer, E. T., and Westerhof, G. J. (2010). Mapping the future of reminiscence: A conceptual guide for research and practice. *Research on Aging*, 32(4):527–564.

[181] Webster, J. D. and Gould, O. (2007). Reminiscence and vivid personal memories across adulthood. *The International Journal of Aging and Human Development*, 64(2):149–170.

[182] West, D., Quigley, A., and Kay, J. (2007). Memento: a digital-physical scrapbook for memory sharing. *Personal and Ubiquitous Computing*, 11(4):313–328.

[183] Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media.

[184] Wilson, R. S., Krueger, K. R., Arnold, S. E., Schneider, J. A., Kelly, J. F., Barnes, L. L., Tang, Y., and Bennett, D. A. (2007). Loneliness and risk of alzheimer disease. *Archives of general psychiatry*, 64(2):234–240.

[185] www.legacyproject.org (2019). Life interview questions. Last accessed 24 April 2019.

[186] Yach, D., Hawkes, C., Gould, C. L., and Hofman, K. J. (2004). The global burden of chronic diseases: overcoming impediments to prevention and control. *Jama*, 291(21):2616–2622.

[187] Yaghoubzadeh, R., Pitsch, K., and Kopp, S. (2015). Adaptive grounding and dialogue management for autonomous conversational assistants for elderly users. In *International Conference on Intelligent Virtual Agents*, pages 28–38. Springer.

[188] Yang, K. and Victor, C. (2011a). Age and loneliness in 25 european nations. *Ageing and Society*, 31(08):1368–1388.

[189] Yang, K. and Victor, C. R. (2008). The prevalence of and risk factors for loneliness among older people in china. *Ageing and Society*, 28(3):305–327.

[190] Yang, K. and Victor, C. R. (2011b). Age and loneliness in 25 european nations. *Ageing and Society*, 31(8):1368–1388.