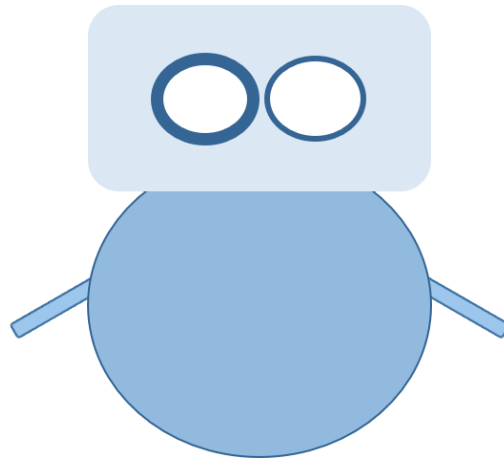


Hi I am Habit Trainer: Self-empowering virtual assistant that facilitates implementation intention and rehearsal



Project exam group: hci103f20
Supervisor: Lefteris Papachristos
Students: Anders Wiese Christensen
awch18@student.aau.dk
IT Design and Application Development, Aalborg University
June 12th, 2020



AALBORG UNIVERSITY
STUDENT REPORT

Summary

This master's thesis is an exploration of how a virtual assistant can support habit development. The virtual assistant was accessed via the Google Assistant Application and interacted with through voice. This enabled creation and formulation of an implementation intention, which is a simple if-then plan to link a specific cue with a goal oriented response. To evolve this plan into a habit, it has to be repeated. We chose rehearsal as a strategy to explore how this augmented repetition and habit development. The main motivation is exploring user empowerment in terms of Digital Behavior Change Interventions. Whereas other research focuses on technology that becomes part of the context, we suggest to empower the users own capabilities and not rely on the technology to support the behavior change in the long term.

Participants were split into two different rehearsal groups and asked to train their habit with the virtual assistant for a 4 week period. Our results show that there was an acceptance of the concept in general. This includes creating and rehearsing an implementation intention to develop a habit. The interviewed participants also explicitly said that they felt like the habit had been developing. Another important finding is that the participants needed support in figuring out which habit to choose. This includes both the cue and the target behavior of the habit. It is also discussed how to measure habit development or automaticity of the target behavior, because implementation intention is a conscious effort in the beginning. Furthermore, the findings show many opportunities for improving the experience of using a virtual assistant for habit development. The primary way to enhance the experience is through the virtual assistants ability to store information. This enables adaptation of the length of conversation based on the amount of experience with rehearsal and type of habit formulated.

Our contributions comprise a set of design recommendations for developers and future researchers that are interested in addressing habit development through use of implementation intention and/or rehearsal. It gives insight into methodologically decisions and considerations when measuring habit development in such a context, and, lastly, it shows the possibility of creating Digital Behavior Change Interventions without creating a reliance on the technology that is used. The set of design recommendations are specific for interventions focusing on implementation intention and/or rehearsal using a virtual assistant. The most important design recommendation is to slowly fade out the rehearsal as the target behavior becomes automatic. This suggestion needs further exploration into how much or little rehearsal is needed in the beginning to until it becomes automatic. This is related to the second contribution which is a discussion of how to measure habit development for research and in practice using the virtual assistant. The most important contribution is the example of trying to support behavior change without a design that leads to reliance on technology, and showing that it is a possible direction for future investigation.

Hi I am Habit Trainer: Self-empowering virtual assistant that facilitates implementation intention and rehearsal

Anders Wiese Christensen

Aalborg University, Department of Computer Science
Aalborg, Denmark
awch18@student.aau.dk

ABSTRACT

Habit development is a central theme when exploring long term behavior change. Implementation intention can support conscious creation and formulation of a habit, while rehearsal can support the need for repetition of the target behavior in a consistent context until the habit becomes automatic. This paper describes an exploratory study in which a virtual assistant is utilized to support implementation intention and rehearsal over a 4 week period. We describe the findings and experiences with creating and formulating habits with a virtual assistant and rehearsing it over time to develop it as a habit. Habit development was measured throughout the study with the Self Report Behavioral Automaticity Index and self-reported adherence. The assessment was further supported with in-depth interviews at the end of the study period. We found an overall acceptance of both the concept of creating and rehearsing an implementation intention with a virtual assistant. Our results suggest that there are many opportunities for improving the use of virtual assistants for habit development. We summarize the most important findings and present a set of design recommendations for developers and future research.

Author Keywords

Implementation intention (IMI); Digital Behavior Change Intervention (DBCI); Rehearsal; Habit Formation; Virtual Assistant (VA)

INTRODUCTION

An increasingly interesting question in Human Computer Interaction (HCI) is how technology can be used to support long term behavior change. Factors that can influence the success of behavior change interventions are individual differences in motivation, context, and situational circumstances that are difficult to predict and measure. Using technology to support behavior change has shown to be effective but most studies are reporting short term effects [18]. The question is difficult to address due to the complexity of behavior change and the fact

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '20, April 25–30, 2020, Honolulu, HI, USA

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-6708-0/20/04...\$15.00

DOI: <https://doi.org/10.1145/3313831.XXXXXX>

that it should be evaluated long term. In addition, researchers do not address habit as a key part of long term behavior [18]. This is important since habits by definition ensure that the behavior will have long lasting effects [15]. Mobile health behavior change interventions not only lack commitment to one theory but also grounding in behavior change theory in general [7]. The far-reaching problem of behavior change is also reflected in app-stores where hundreds of applications address the issue in their communication but often fail to ensure behavior change. The applications often focus on tracking and self-monitoring and lack connection to behavior theory in general and certainly to the habit literature [4, 24]. Although tracking and self-monitoring can support behavior change in some sense, this is not due to the development of habits [22].

A habit is defined as a repeated behavior within a consistent context. It is unconsciously triggered by a cue, or a set or cues, in a context and can be described as automatic [15]. One way to address habit is through the concept *implementation intention* and repetition. An implementation intention (IMI) is a simple if-then statement that links a certain cue with a response in relation to one's goal [9]. It is written in the format of:

"After X happens, I will Y"

An example of this could be "After I brush my teeth, I will floss". Gollwitzer and Sheeran [10] show in a meta-review study that IMI has medium to large effect at goal attainment. However, this effect cannot be translated into repeated behavior, such as habits which are developed through repetition of the behavior in a consistent context. The studies included in the meta-review focus on the effectiveness of IMI at any given single event and not a repeated behavior. For IMI to become automatic it has to be repeated consciously in the beginning, and the more repetition the more automatic it becomes [18].

The purpose of the present study is to explore how a virtual assistant (VA) can be used to support habit development in a sustainable way without having the user become dependent on technology. In a 4 week study, participants ($n = 11$) were asked to use a VA to create and formulate a habit, and to rehearse it consciously to support the development of automaticity in the behavior. A VA is interesting for several reasons. It can be used on any smartphone, and therefore in any given context, and one can interact with it through text or voice. This creates high accessibility. Moreover, the simple format of an IMI

allows for a simple conversation structure, which aligns with what commercially available VAs are capable to provide when developing the VA.

This article has three primary contributions to the field of habit development and HCI. First, a set of design recommendations for both researchers and developers of VAs who strive to help people change behavior. Second, it gives insight into how researchers can set up Digital Behavior Change Interventions (DBCIs) in the future and measure the intended effect. Third, it elaborates and shows the possibility of creating DBCIs that empower the user to sustain their behavior change on their own without continual use of technology.

RELATED WORK

The strategy of using IMI in an intervention facilitated by technology has been explored primarily through SMS as triggers for a certain behavior [22] and mobile applications to support the creation of IMI and triggering of behavior [19]. In the cited papers, it is hypothesized that this strategy can address the process of developing a habit that ultimately leads to long term behavior change. The first part is the creation and formulation of the IMI and the second part is facilitating repetition. Pinder et al. [19] explore both the creation phase and how an application can be used to create context-aware reminders to help the user remember their behavior in the desired moment until it becomes automatic. They suggest further investigation into using Bluetooth stickers that can increase the context awareness of the phone. Their idea is that this will increase the user possibilities in regards to choosing cues in more contexts to trigger the wanted behavior. Several other studies explore the notion of calendar based reminders to increase short term goal attainment [20, 22].

One important issue that has been pointed out is that even through reminder and tracking based strategies can be useful for supporting behavior change, the problem is that the technology creates a scaffolding effect. The approach creates a technology dependency. This issue has been mentioned in a review of mobile applications designed to help users remember their medication [21], and in another study where it is suggested that context-aware reminders should be faded out [19]. Yet, the suggested new approaches have not been explored practically in experiments.

Stawarz et al. [22] explore the effect of reminders on habit development as indicated by automaticity [8]. It seems that even though reminders have a positive effect on adherence to a desired behavior, they also create reliance on technology. In their empirical study, there was no evidence that using reminders helped people develop automatic behavior. Similarly, even through the use of activity trackers has shown to support behavior change, it has failed to do so by supporting habit development. Instead the technology has become part of the environment and again created reliance on technology [11].

Rehearsal has shown positive effect on creating new behavior. In a study by Veiling et al. [23], participants created an IMI for dieting behavior change and trained it with an online program. This helped people lose weight by consciously thinking about their behavior once a week during the four week intervention.

Another way of rehearsing consciously is through role playing. We know this for example from fire drills, where we act as if there is a fire to learn the process for a real emergency [12].

MOTIVATION

A main goal of HCI and habit development is long term behavior change. This study is the first step in exploring both the feasibility of using a VA for habit development and the potential benefits of supporting self sustaining behavior change. This study provides insight into some of the many open questions about the effectiveness of such an approach.

Our main motivation comes from the interest in exploring user empowerment in terms of DBCIs. As technology and its capabilities advance, there is a tendency in DBCIs to explore the potential of technology and investigate how it can change behavior. The nature of behavior change is a wicked problem as the requirements for the any given system are incomplete, contradictory and constantly changing [18]. Due to its complexity, the context of several different users is something that one system or technology cannot comprehend [1]. Cues for behaviors can consist of emotions and feelings, which can have just as much impact on outcome behavior as visual, time based or location based cues. This suggests that no matter how sophisticated the technology gets, tracking and context-aware systems will only have an incomplete picture of the relative importance of cues to individual users in specific situations.

Instead we suggest to empower the user and allow behavior change to only partially rely on technology in the beginning while gradually fading out the dependency. This study is a concrete example of an exploration to develop technology that actually supports the users own capabilities of behavior change. In this study it is explored how Nickerson et al.'s [17] way of facilitating formulation of IMIs with questions works in the context of using a VA. Further, we look into how a rehearsal scheme similar to Veiling et al. [23] and Holmes and Jones [12] works in the context of IMI facilitated by a VA.

METHOD

Participants were asked to formulate and rehearse a habit with the help of a VA for 4 weeks. During the period, participants were asked various questions regarding themselves, such as previous experience with habits, how their habits developed in this study, and their experience with this intervention. Data about their habit development were gathered continuously throughout the study. In the end, a subset of the participants shared their experience in an in-depth interview. These interviews were analyzed using thematic analysis to support the continuously gathered quantitative information.

Participants

The participants (n = 15) were primarily recruited at social media networks both through personal relations and public groups. Three participants never initiated the study (reason unknown) and one was unable to install the VA. Most of the participants (n = 11) completed the study and rehearsed their habit for 4 weeks. The participants did not start the study at the same time but all the data was matched up on how long

each participant had been training. The age distribution ranged between 24 and 58 years, with an average of 30.9 years. 7 out of 11 were female. Participants expressed their interest in the study after reading the recruitment text on social media and shared their email.

Procedure

We here give a high-level overview of the procedure. The most relevant steps will be elaborated in the following sections.

1. Participants received an email with information about the study and instructions on how to get the VA.
2. After creation of the IMI with the VA, it informed them about rehearsal at the end of the conversation. Afterwards they answered a questionnaire with information about themselves and baseline information about the habit.
3. It was recommended to rehearse the habit three times per week in the beginning and then less as the habit developed.
4. Each week they answered a questionnaire to keep track of habit development.
5. At the end they answered a more descriptive questionnaire and some participated in interviews.

Study context

The study was communicated as a help to develop corona outbreak related hygiene habits, such as washing hands or coughing and sneezing in the sleeve. A number of the participants wanted to create their own habit for different reasons. To increase the potential amount of participants, it was allowed to pick a habit of one's own choice. The participants were asked to use the VA through the google assistant application on their smartphone.

Creation of implementation intention

Habits and behaviors changes must be as simple and easy to carry out as possible. In the study, predefined hygienic habits were formulated to consist of only simple behaviors, and the participants who were choosing their own habits were carefully guided to ensure a simple behavior that was easy to incorporate in their daily life. The simplicity of a habit affects the time it takes to develop [15]. Hence, all participants were engaged in behaviors that can become automatic relatively fast. Another reason for choosing simple habits is to lower the impact of motivation on habit development. The 'stage-of-change', which indicates readiness to change behavior, also seem to affect habit development [11]. The simpler habits the less these factors come into play and affect the result. Participants were encouraged to find a noticeable cue as a trigger for the behavior according to habit theory [9].

The participants were instructed to engage with the VA and create their habit. Inspired by questions used in the article by Nickerson et al. [17] who created an IMI indirectly to increase voter turnout at an election, the VA was designed to ask questions to support the formulation of the IMI. After creating the IMI, the VA explained the importance of rehearsal and recommended rehearsing three times a week at the initial stages of habit development, as done in similar studies [23, 14].

Rehearsal groups

First and foremost, the study was conducted to examine the feasibility of using a VA to create and formulate an IMI. Second, but just as important, it was conducted to investigate the effectiveness of the type of rehearsal to train the IMI. Based on literature, speaking out loud, acting and writing down were identified as interesting types of rehearsal, in terms of supporting conscious memorization [3, 12, 25]. Speaking out loud and acting (i.e., role playing) were most suitable for the format of a VA and were thus chosen as experimental groups. The participants were split up into the two experimental groups (n = 4 in each group) and a control group (n = 3) that did not rehearse the habit. All groups were instructed to create an IMI, and the two experimental groups were also instructed to rehearse with the VA. The first group downloaded an assistant that facilitated speaking out loud to rehearse the habit. The second group downloaded an assistant that facilitated a role play for the rehearsal. The creation phase was identical for both groups. The speaking out loud assistant asked questions and made the user answer them. The role playing assistant guided the user through acting out the habit, i.e., instructed them to pretend to be in the situation of the chosen habit, noticing the cue and then consciously carry out the target behavior.

Designing the Habit Trainer

In this section, we describe the most important parts of the development process and the tools we used. From the first draft of the conversation, through different types of testing and validation of the design, to the last implementation and deployment.

The first draft of the conversation flows for creation and rehearsal of IMI were created. The development of the conversational design was done in an iterative manner. Initially, the questions formulated by Nickerson et al. [17] to facilitate creation of IMI needed adjustment to fit our context with VA-person conversations rather than person-person conversations. To develop and create the conversation flows, the tool Mindomo (www.mindomo.com) was used which helped create a tree structure of the different flows of conversation. Keeping track of the conversation in Mindomo was favorable over DialogFlow which does not support an overview of the conversation needed to iterate and improve the text for each phrase.

The content of the conversation was developed through several iterations and sparring with other people, as suggested in Google's development documentation. Inspired by wizard of oz testing [5], this was done with a Facebook site appearing as a chatbot. The people using it believed that they were interacting with a chatbot, but in fact they were just chatting with us without being aware of it. We convinced users that they were talking to an already developed chatbot and that there was no human interaction going on. This also gave insight into what type of sentences the users were using when thinking that they were in dialog with a system.

The assistant was developed in DialogFlow, and integrated with google assistant to distribute it to the participants. The inline editor, which is a part of the tool, that allows for simple

coding, was used to develop logic that guided the conversation in relation to the habit that was chosen.

The implementation was tested to fix unexpected errors in the final version and afterwards the VA was deployed through the alpha testing functionality in DialogFlow.

Data gathering

In this study both quantitative and qualitative data was collected. Participants answered questionnaires both before and after the study, as well as recurrently during the four week period. In addition, a subset of the participants were interviewed after the research period. The following sections describe how and what information was gathered.

Weekly recurring questionnaires

The weekly questionnaire functioned as a measure of automaticity, adherence and rehearsal. Automaticity was measured on the Self Report Behavioral Automaticity Index (SRBAI) [8] to be used as an indicator of habit development. Adherence was also measured as self-report and later calculated as percentage of how many days of the week the habit was completed. The number of times the participants rehearsed was both measured as self-report and validated through usage data.

Pre-study questionnaires

Initially the participants answered a pre-study questionnaire with demographic data such as age and gender. They also answered self-assessment questions about how motivated they felt to change their behaviors in the context of this study. They further answered how difficult the task seemed to them, and, lastly, they answered questions about automaticity to establish a baseline automaticity score before deployment to track habit development.

Post-study interviews

In the post-study interviews, the users were asked to elaborate on three topics; their overall experience, creation of their habit and rehearsal of the habit. The interview was completed in a semi-structured manner with broad questions. For their overall experience they were asked to evaluate their progress and perceived impact of the entire study. This naturally lead to more specific questions about the process of creating and rehearsing the habit.

Data analysis

Thematic analysis was done according to the guide from Braun and Clarke [2]. The in-depth interviews were used as data input to the analysis according to the following process. First, we got familiar with the data and got an overview. Each question and answer was set up in a spreadsheet for each interviewee in an organized manner. This enabled the initial coding of the text. The first walk-through of the data included coding of topics of interest. Afterwards, overall common themes were found based on this initial coding, and, at last, the themes were defined and named. Overall there were 71 interesting comments from the in-depth interviews. The top 4 themes included 92 percent of the comments and between 1 and 3 subcategories as seen of Table 1.

FINDINGS

This section describes our findings from the questionnaires. Primarily, this consists of measurements of adherence, automaticity and rehearsal presented as simple graphs with accompanying descriptions. Findings from the in-depth interviews are presented in the form of most mentioned and described topics based on the thematic analysis. Furthermore, each topic is often exemplified with quotes or paraphrases from the interviews.

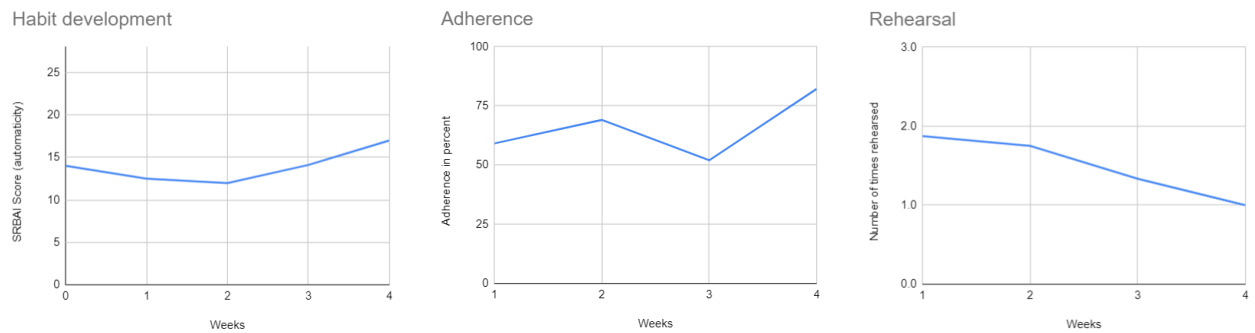
Interview data show that participants in general liked the Habit Trainer and that they were surprised with the effect of the intervention as a whole. Well expressed by the quote referring to an event near the end of the intervention "...the first time I completed the habit without even realizing... That was a fun experience". 18 percent of the mentions included positive feedback on the VA, as seen in Table 1. The first theme 'Memorization of IMI' includes a general positive attitude towards the concept of IMI and how well participants think that it worked. The negative feedback or indications of possible improvements of the VA in relation to this intervention is included in the themes 'Issues with conversation design' and 'Difficulties with rehearsal context'. These will be described as separate subsections later on.

Themes	Mentions	% of total	Subcategories	Mentions	% of total
Memorization of IMI	33	46%	Liked IMI	13	18%
			Knowledge about their IMI	11	15%
			IMI relevance	9	13%
Liked VA	13	18%	Like talking to VA	13	18%
Issues with conversation design	12	17%	To long conversation	7	10%
			To long text	3	4%
			Rehearsal became tiresome/obsolete	2	3%
Difficulties with rehearsal context	7	10%	Context hindered talking	4	6%
			Context prevented rehearsal	3	4%

Table 1: This table shows the most mentioned themes in the thematic analysis sorted by relevance, including subcategories, their mentions and percent of total mentions.

Habit development and automaticity

Figure 1a shows an increasing trend in automaticity from the baseline measurement in the beginning to the last measurement in the intervention based on measurements from the 11 participants across the study period. SRBAI results indicate an increase in automaticity when performing the intended behavior [8]. All the interviewed participants reported that they believe that their habit will stay part of their life in the future, ranging in certainty from "probably" to "definitely". The interviewed participants scored relatively high in the SRBAI and when they were asked directly in lay man's terms if they would consider it a habit, all of them said "yes", with varying degree of certainty. The habit development graph for all participants illustrate the overall increase, but also an unexpected small drop from the baseline measurement compared to the following two weeks. Furthermore, some participants expressed



(a) Habit development across the study period for all participants indicated by automaticity, i.e., SRBAI-score. (b) Average adherence across the study period for all participants, measured in terms of number of days the habit was completed in a week as reported in the weekly questionnaires. (c) Rehearsal across the study period measured as the average number of times the habit was rehearsed in a week. This was self-reported and validated with usage data.

Figure 1

confusion towards the questions for measuring automaticity and had difficulty with answering them.

Adherence

Figure 1b shows adherence measured throughout the four week period with a tendency to increase over time. It is measured as self-report through the weekly questionnaires with a question about how many days throughout the week they performed the target behaviour. Based on the answer, a percentage of days during the week was calculated. Overall, the interviewed participants felt quite confident about the correctness of their answer about the weekly measurement of days they completed the habit.

Memorization of implementation intention

All participants that were interviewed remembered their IMI when they were asked. In the thematic analysis, 18 percent talked about how they liked the concept of formulating a habit in the form of an IMI, 15 percent of mentions indicated that the participants had a good understanding of the theory behind IMI, and 13 percent of mentions showed that the chosen IMI had a specific relevance to the participant. Several interviewees were surprised of how well this intervention helped them remember their habit and complete it. Some thought this was because of the logic behind the formation that links the cue and behavior or because of the rehearsal or both.

Creation of implementation intention

Overall the concept of IMI was well received by the participants, and, as Table 2 shows, everyone formulated good IMIs, using salient cues and simple behaviors as instructed. Talking out loud to the VA also worked effectively, as one described it *"saying it out loud feels like making a promise to myself"*. This was also described in other ways with similar meaning by other interviewees. All interviewees would strongly consider using this format to develop new habits in the future, which also indicates an overall acceptance of the concept and a belief or proof of its effectiveness.

Difficulty with choosing a habit

Almost all participants had a lot of difficulty with choosing a cue or behavior. Most participants chose a cue or behavior

Cue	Behavior
When I brush my teeth	I do five squats
After I enter my front door	I will wash my hands
When I drink coffee	I will do 5 minutes of yoga
After I brush my teeth	I will make my bed
When my nose tickle	I will sneeze in my sleeve
After I brush my teeth	I will do 10 push-ups
After I am finished eating	I will wash the dishes
After I touch a door handle	I will wash my hands
After I come home from work	I will wash dishes
After I come home from work	I will go for a walk
When I come home from shopping	I will wash my hands

Table 2: List of habits in terms of cue and related behavior. Bold marking indicates the habits that are inspired by the given examples, either from the VA or recruitment text.

inspired by examples in either the recruitment text or from the VA as seen on Table 2 (illustrated as bold text). Most of the participants (7 out of 11) used examples as inspiration for both the cue and the behavior, while only one participant was completely original in terms of finding a cue and behavior in the creation of the habit. One of the interviewees expressed a need for help with the creation of other habits as well and said *"I would like to be able to create and train habits on other things"*.

Rehearsal

The findings and feedback on rehearsal was wide ranging, from *"the VA made me laugh"* to *"the app kept interrupting me"*. As seen on Figure 1c the data suggest that, on average, the participants rehearsed regularly with a tendency to rehearse less and less over time exactly as the VA recommended. The rehearsal was measured as self-report and validated with usage data, with little deviance. The interviewed participants had a lot of opinions about the rehearsal and the VA, but most participants agreed with the following message expressed by one interviewee as *"It really made sense to rehearse it, I feel like it helped a lot"*.

Issues with conversation design

In the thematic analysis, 17 percent of mentions pointed at issues with the conversation design with 10 percent at 'too long conversations', 4 percent at 'too long text' and last 3 percent mentions of rehearsal becoming tiresome and obsolete after some training. One participant expressed it in a single sentence: *"I would have liked a short rehearsal after I have gotten used to it. It is too long"*. One described the rehearsal as tiresome and followed up with, *"I sometimes just rehearsed it without the VA because I knew what to do and understood the concept"*. Someone also suggested a quick rehearsal when the user knows the concept.

Difficulties with rehearsal context

Some of the interviewees mentioned the context as being an issue in relation to the rehearsal. One participant in the speaking out loud group was unable to speak at some point because there was someone else in the room, which made it uncomfortable. It was expressed as, *"It was intimidating to speak to the VA, what if someone heard me"*. Another had issues with acting out because the participant was unable to act out the habit due to external constraints. The room in which the habit was triggered was occupied which did not allow for rehearsal in the right context where the cue was located. Someone else had difficulties with remembering to rehearse and suggested *"Maybe if it could set reminders to practice in my phone by itself, that would help getting the practice done."*

General issues with the Virtual Assistant

Other issues were also described which were more general to the VA than the specific case of IMI and rehearsal. One complained about the way the VA interrupted and said *"It is annoying that it interrupts when I am in the middle of answering. And even though the answer is very incomplete, it still said 'good job' or something like that" and "it continued even though I wasn't ready to continue, because the microphone listened every time after it had talked"*.

Others described the need for human touch in a supporting and value reinforcing manner with feedback expressed as *"Maybe give it a more human touch, with encouragement and such. For example if you complete the habit 3 times in a row, have the assistant say something like: 'Good job, I am proud of you' or something similar."*

DISCUSSION

Overall, participants liked the experience of creating IMI and rehearsal with the VA called Habit Trainer. Even though the results showed an increase in the automaticity score, the small number of participants does not allow us to draw determinate conclusions or make generalizations. The participants felt like they have developed and were supported in developing a habit. Participants gained a sufficient understanding of IMIs after the intervention and they created good IMIs with salient cues and simple behaviors. It seems that the concept of IMI is easily explainable and easy to facilitate even with rigid conversation structures delivered by an VA. The combination of IMI and rehearsal for habit development facilitated by a VA seems promising. This includes both the creation of the IMI in the beginning, and rehearsal of the IMI in two different

groups. The feedback from both rehearsal groups had great similarities.

Creation of implementation intention

The IMI creation through the VA was expected to be more difficult as Nickerson et al. [17] did it through a phone with a real person facilitating the IMI. But one should keep in mind that the participants were also introduced to the concept of IMI in the recruitment text and preliminary communication, meaning that they already had familiarized themselves with the concept before interacting with the VA. Nevertheless, it turns out that creating IMI with speaking out loud works very well, and, as one participant described, *"it feels like a promise"*. But, as mentioned in the findings sections, some of the participants found it intimidating to speak out loud in certain unsuitable contexts. Thus, it could be interesting to see the effect of two groups creating an IMI either by speaking out loud or by writing it to a chatbot.

Findings also suggest that people do not know what behavior and cue they should pick. This is despite the fact that the participants decided to join the study because they were interested in developing habits. Most of the participants picked examples or suggestions directly from the VA. The finding from Pinder et al. suggested to *"Guide users during the implementation intention formation phase"* partially to support appropriate goals. It seems that users not only need support in the creation of the IMI, but *also* support in which cue and behavior to pick. This finding is only an initial indication, but it suggests that when helping people change habits, they might need help choosing both the behavior and they definitely need help choosing the cue that triggers the behavior.

In this regard it could be interesting to see the effect of integrating a semi complete list of the most popular goals and related behaviors combined with cues for each behavior. If the VA was connected to a relatively simple database with the above mentioned attributes, the user could search around for interesting habits based on their overall goals or areas of interest. The conversation design for creating an IMI could start with an overall area or goal such as dieting, exercise, working or reading. Then the VA could share the top three behaviors related to the specific goal. If a behavior is picked, then it could share the most used or recommended cues for the specific behavior or group of behaviors. This would benefit the users that would get the help they need to create proper IMIs. Further, data could be gathered about the most popular IMIs and how well they work over time. We return to this in later paragraphs.

Progress in implementation intention

The findings also suggest that within a couple of weeks some participants wanted to create more habits or adjust their habits. One of the participants found the behavior to become too easy or uninteresting and changed it. The interesting part here is that the participant really found the cue to work well and kept the cue, but only changed the target behavior to something more difficult. This is also an interesting finding pointing to the fact that people should get help with how they can progress with their habit and when.

Rehearsal

The findings suggest that people generally remembers their IMI when they were asked, which was partly the purpose of rehearsal. In the acting out group, participants shared great experiences with the rehearsal. Similarly, the speaking out loud shared excitement to this way of supporting conscious momentary remembrance when noticing the cue. At the same time, people generally thought that practice provided by the VA was tiresome and that it was too long. Both of these findings were expected, as rehearsal has been used for memorization in other studies showing similar effects [23].

One of the participants suggested to make both the interaction, and the text within each message in the VA shorter. This will be a way of addressing some of the issues found. Making each of the texts messages shorter but also having less text messages can work when the user know how the rehearsal works. This would be possible to adapt based on the times the user has rehearsed, but that might not be favorable. Another way of addressing it could be through evaluating the progress in general for each user. Rehearsal should be elaborate enough to induce conscious memorization in the moment of noticing the cue, while the rehearsal itself should not lead to less engagement through a tiresome conversation with the VA. Based on this study, we do not know how much rehearsal or how long rehearsal is enough to have this effect.

Virtual assistant storing user information

To improve the conversation design, the VA needs to store user information in regards to cue, behavior and goal in a database in order to have more knowledge about the context of the user. Their IMI and how many times they have rehearsed would be interesting data points to save initially in the database. The storing of IMI can be used in order to recommend most popular habits for each type of goal for new users. This could both be based on how any times it has been chosen, or how well other users in the past have developed their habit if this is measured as well. If the number of times rehearsed can get logged, it is possible to design in accordance with this. An example of this could be that if the user has rehearsed the same habit for more than 3 times, some of the examples or additional text can be removed from the messages. Similarly, the length of the conversation could also be adjusted. Yet, it can be difficult to know if three times is enough for everyone to remember. Another way to decide when to shorten the rehearsal could be based on how the habit develops if this is measured properly and directly in the VA.

Type of rehearsal and context

Several participants commented on not being able to rehearse in the specific context for various reasons. Furthermore, based on this study, there is no conclusion whether speaking out loud or acting out works better in terms of long term habit development. One can easily come to think that the user should be allowed to choose their own type of rehearsal based on what fits best with the momentary context. This way, they could decide whether speaking out loud, acting out or writing with the VA seems more suitable. This can become problematic for a couple of reasons. Users would probably not know what to choose if they were asked, or they might choose what feels

most comfortable in the moment. The problem is that we do not know if, for example, acting out is the most effective way of rehearsing. Future research should therefore investigate group differences and explore the context of users thoroughly and see what type of IMIs fits in which contexts.

Another enhancement to the VA could be the feature of setting reminders in users' calendars for rehearsal. Not to remind people of their IMI similar to other research [22], but to remind people to rehearse the habit. The reminders could be recommended based on the progress of each user.

Measuring habit development

In our study, there were several issues with measuring habit development, both in terms of automaticity and adherence. Automaticity might not be the best indication of habit development in the context of IMI, and adherence is difficult to measure without involving the participant or too much.

Do we dare to talk about automaticity based on SRBAI?

What is actually measured with Self Report Behavioral Automaticity Index in an intervention with IMI? One of the four questions is formulated as *"The behavior is something I do without thinking about it"* and another is *"The behavior is something I do without having to consciously remember"*. SRBAI is not developed directly in the context of IMI and it might be problematic to use in this context. The purpose of IMI and rehearsal is to help people think consciously about their habit before it becomes automatic. This is done with the salient cue that shows itself, which in the beginning should trigger conscious thinking of the IMI. After some repetition, the amount of thinking between cue and behavior will decline, but not in the short term. Even though it happens automatically at some point, there is still conscious thought about the IMI because the entire strategy has been a really conscious effort. The parsimonious instrument was developed to measure automaticity in behavior related to energy-balance, physical activity and dieting [8]. There is no direct evidence that this works as an indicator for habit development in interventions containing IMI, even though other studies have used it [22]. Our findings and line of logic suggest that the use of IMI lowers the result of the SRBAI score in the first couple of weeks, as the participant is instructed to think more about their habit. The automaticity baseline, on the other hand, was measured before the participants started their training and therefore scored relatively high compared to the following weeks. We argue that this should be kept in mind, in general, when using SRBAI to measure the development of habit in an intervention where participants are asked to think about their behavior over a period of time. Questioning people several times gets them thinking about their behavior more consciously which will affect their response to the questions. This means that one should in fact *expect* to see a drop initially. Our results show that after the first two weeks SRBAI starts trending back up which indicates an increase in automaticity. Hence, in long term evaluation we believe that SRBAI works sufficiently well as a measure of automaticity. We further remark that as long as we only compare differences between SRBAI scores in our interventions groups, the skewness of SRBAI is not as big an

issue as it would be if used directly as an indicator of habit development.

Adherence as an indicator for habit development

The best indication of habit development might be to measure adherence directly over time, as a habit is defined as repeated behavior in a consistent context. It is up for discussion whether a behavior is automatic and happens without thinking about it, but, as for behavior change, the completion of the behavior itself must be the primary goal. There are not a lot of ways to measure this depending on the behavior that needs to be measured. The issue is to measure behavior without requiring too much effort from the participant in the study. In this study, adherence was measured as self-report each week similar to automaticity. Generally, self-report may not be the ideal way of measurement, since participants do not remember precisely what happened during the previous week. An alternative could be to have the participant write down each day when they complete their habit or to have an easily accessible technology to register completion of behavior [6]. The issue with this is that this may affect the result of the intervention too much, as it becomes part of the context.

Measuring progress

Researchers in this area should be careful about using the SRBAI to measure the development of habit in an intervention where participants are asked to think about their behavior over a period of time. Questioning people several times gets them thinking about their behavior more consciously which will affect their response to the questions. It should be used long term and supported by other measures.

The issue with using adherence as a measure is that it does not measure the automaticity aspect, and technology such as activity trackers could be mistakenly perceived as good for long term behavior change.

A solution to track the progress practically in the VA could be to simply let the perception of the participant or user guide the progress. One of the participants asked to give the Habit Trainer a more "human touch". The VA could ask questions like *"How is it going with your new habit?"* or *"Do you think that your habit is developing?"* in the initial greetings. If the user feels good about the habit, the VA could save this information and adjust the rehearsal conversation design accordingly. If the user is in doubt, the VA could be programmed to take into account the number of times the user has practiced to further evaluate the progress.

Technology dependence

Another important part of the general philosophy behind using rehearsal and a VA was to minimize the use of technology and rely more on the human mind and tendencies. Thus, it makes sense to start small and create or formulate the habit in the form of an IMI before it gets developed and become automatic. This strategy minimizes the time the user needs to rely on technology to change the behavior. Think of a health tracker which should be used over a longer period of time to work properly. Users usually stop using a fitness trackers, thus, it is unable to support long term behavior change [13, 16]. In the same way, using context-aware technological reminders to

support IMI, the technology has to stay part of the context to support behavior unless the reminder can be built in a way to not become part of the context.

Limitations

The overall goal is to create a DBCI that leads to sustainable long term behavior change. Long term behavior change can be based on habits, habits are created through repetition, repetition increases as a result of rehearsal and the needed rehearsal can be facilitated with a VA like Habit Trainer. In the study, however, although the results point in this direction, participants were not measured long enough to conclude long term behavior change. An ideal study should follow participants for more than 4 weeks.

In our study we assumed that automaticity, as measured with the Self Report Behavior Automaticity Index (SRBAI), can be used as an indicator of habit development. We believe that one could get a better measure of behavior change directly though adherence, i.e., completion of the habit when using IMI in the intervention.

Our study constitutes a first step in exploring the feasibility of using a VA to create and rehearse IMI. There are still many aspects that need further investigation before we can conclude to see long term behavior change. As mentioned, behavior change is a complicated domain. We believe that all components in any intervention are too entangled to be analyzed separately and that these must be explored with one single intervention in isolated studies to conclude anything specific. Therefore, we need many new studies that measure behavior change directly as a result of a given intervention.

Another limitation in our study the small number of participants to evaluate the outcome with statistically sound evidence. The expected differences across the groups of rehearsal are small, and a power calculation suggests more than twenty times more participants to find significant effects. However, in previous studies both IMI and rehearsal have been shown to give significant effects [17, 23].

Design recommendations

Our analysis gives insight into how a VA can support creation and rehearsal of IMI. We suggest that VAs facilitating IMI and rehearsal should:

1. Not only guide users through the creation and formulation of an IMI with examples, but also allow for users to pick goals, behaviors and cues that are predefined.
2. Encourage users to talk to the VA about their IMI out loud or to formulate the sentence to make it *"feel like a promise"* to themselves.
3. Allow for progress in the behavior when the down-scaled target behavior has become a habit.
4. Fade out gradually by, for example, keeping track of rehearsal to shorten both text, length of conversation and decrease recommended times per week for the rehearsal, before it gets tiresome. This should be in a way such that the VA explains the rehearsal in depth for participants to understand what to do.

5. Help users accept that the VA is suboptimal in terms of accidentally interrupting and not understanding what the users says.

FUTURE WORK

The most important step forward is to figure out which type of rehearsal works best. This should be further explored qualitatively to find out which type of habits work and do not work for different types of rehearsal. It would be interesting to get statistically sound evidence for habit development and rehearsal group, and to explore associations between habit development and other factors. It might be favorable to find out if certain types of rehearsal fits certain types of contexts and therefore habits. It may be irrelevant to check the difference between rehearsal groups if the type of rehearsal is mutually exclusive for different types of habits. In this case, the type of rehearsal should be based on the type of behavior or cues the user picks to develop as a habit.

Furthermore, the fading out of rehearsal in terms of length of text, length of conversation and necessary amount of rehearsal should be explored. The amount of information that is needed to understand the concepts and instructions to fulfill the rehearsal, compared to how tiresome they get over time, is not addressed in this study. Fading out rehearsal should be based on habit development, but measuring this properly without too much trouble for the user is difficult and needs further investigation. Based on this study, it seems as if people are capable of assessing their own habit development. In the short term, this might be as good an indicator for habit development as SRBAI. We still have no insight into habit measurement in the long term.

If researchers were to do a statistically sound study for long term behavior change, they may reconsider the way to measure the change in behavior directly. It may be easier to discover effects, if a more direct measure than the SRBAI is used in relation to IMI.

When creating an IMI it would be interesting to do a momentary exploratory study with a VA that is linked to a database of habits. This could be done with a simple user test, to see how people respond to navigating through a goal, to a behavior, to a cue, compared to choosing their habits based on more generic examples.

The most important lesson from this study is the general direction of avoiding technology dependence as much as possible. This study indicates that it is possible to find an intervention that ultimately empowers the users own capabilities to change behavior, without using technology that needs to stay part of the context.

CONCLUSION

In this study, we found that participants are open to change behavior using a VA that helped them create and formulate a habit and rehearse it. The creation of the IMI was found to work better than expected and the rehearsal got both positive and negative feedback. A set of design recommendations is

shared based on these findings which can be used as a guideline for further research in this area and for development of VAs that support habit development through IMI and rehearsal.

Additionally, our analysis gives insight into methodological issues regarding measurement of habit development. The most important take away in this regard is the difficulty of using SRBAI as an indicator for automaticity, especially in the short term, when using IMI.

Several questions have been addressed and can be used as an inspiration for other concrete examples of interventions for behavior change. Yet, several questions regarding long term behavior change remain unanswered as well as unmeasured. We argue that long term behavior change can be addressed with an intervention such as considered in this paper, focusing on minimizing the reliance of technology as much as possible.

REFERENCES

- [1] Victoria Bellotti and Keith Edwards. 2001. Intelligibility and accountability: human considerations in context-aware systems. *Human-Computer Interaction* 16, 2-4 (2001), 193–212.
- [2] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [3] Ronald L Cohen and Kjell Granström. 1968. The role of verbalizing in the memorizing of conventional figures. *Journal of Verbal Learning and Verbal Behavior* 7, 2 (1968), 380–383.
- [4] Logan T Cowan, Sarah A Van Wagenen, Brittany A Brown, Riley J Hedin, Yukiko Seino-Stephan, P Cougar Hall, and Joshua H West. 2013. Apps of steel: are exercise apps providing consumers with realistic expectations? A content analysis of exercise apps for presence of behavior change theory. *Health Education & Behavior* 40, 2 (2013), 133–139.
- [5] Richard C Davis, T Scott Saponas, Michael Shilman, and James A Landay. 2007. SketchWizard: Wizard of Oz prototyping of pen-based user interfaces. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*. 119–128.
- [6] Maria Angela Ferrario, Will Simm, Adrian Gradinar, Stephen Forshaw, Marcia Tavares Smith, Thomas Lee, Ian Smith, and Jon Whittle. 2017. Computing and mental health: intentionality and reflection at the click of a button. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 1–10.
- [7] Caroline Free, Gemma Phillips, Louise Watson, Leandro Galli, Lambert Felix, Phil Edwards, Vikram Patel, and Andy Haines. 2013. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. *PLoS medicine* 10, 1 (2013).
- [8] Benjamin Gardner, Charles Abraham, Phillipa Lally, and Gert-Jan de Bruijn. 2012. Towards parsimony in

- habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity* 9, 1 (2012), 102.
- [9] Peter Gollwitzer. 1999. Implementation Intentions: Strong effects of simple plans. *American Psychologist* 54, 7 (1999), 493–503. DOI : <http://dx.doi.org/10.1037/0003-066X.54.7.493>
- [10] Peter M. Gollwitzer and Paschal Sheeran. 2006. Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. *Advances in Experimental Social Psychology*, Vol. 38. Academic Press, 69 – 119. DOI : [http://dx.doi.org/https://doi.org/10.1016/S0065-2601\(06\)38002-1](http://dx.doi.org/https://doi.org/10.1016/S0065-2601(06)38002-1)
- [11] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2015. How do we engage with activity trackers? A longitudinal study of Habito. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1305–1316.
- [12] Glen A Holmes and Russell T Jones. 1996. Fire evacuation skills: cognitive-behavioral versus computer-mediated instruction. *Fire technology* 32, 1 (1996), 50–64.
- [13] Evangelos Karapanos. 2015. Sustaining user engagement with behavior-change tools. *interactions* 22, 4 (2015), 48–52.
- [14] Ilya Kreynin, MohammedShabbar Manek, and Chirag Variawa. 2019. CREATING A VIRTUAL CHATBOT TO SCAFFOLD SKILLS DEVELOPMENT IN FIRST-YEAR ENGINEERING EDUCATION. *Proceedings of the Canadian Engineering Education Association (CEEA)* (2019).
- [15] Phillippa Lally and Benjamin Gardner. 2013. Promoting habit formation. *Health Psychology Review* 7, sup1 (2013), S137–S158.
- [16] Dan Ledger and Daniel McCaffrey. 2014. Inside wearables: How the science of human behavior change offers the secret to long-term engagement. *Endeavour Partners* 200, 93 (2014), 1.
- [17] David W Nickerson and Todd Rogers. 2010. Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychological Science* 21, 2 (2010), 194–199.
- [18] Charlie Pinder, Jo Vermeulen, Benjamin R Cowan, and Russell Beale. 2018. Digital behaviour change interventions to break and form habits. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 3 (2018), 15.
- [19] Charlie Pinder, Jo Vermeulen, Adhi Wicaksono, Russell Beale, and Robert J. Hendley. 2016. If This, then Habit: Exploring Context-aware Implementation Intentions on Smartphones. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)*. ACM, New York, NY, USA, 690–697. DOI : <http://dx.doi.org/10.1145/2957265.2961837>
- [20] Marco; Hurling Robert Prestwich, Andrew; Perugini. 2010. Can implementation intentions and text messages promote brisk walking? A randomized trial. *Health Psychology* 29, 1 (2010), 40–49. DOI : <http://dx.doi.org/https://doi-org.zorac.aub.aau.dk/10.1037/a0016993>
- [21] Katarzyna Stawarz, Anna L Cox, and Ann Blandford. 2014. Don't forget your pill! Designing effective medication reminder apps that support users' daily routines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2269–2278.
- [22] Katarzyna Stawarz, Anna L. Cox, and Ann Blandford. 2015. Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2653–2662. DOI : <http://dx.doi.org/10.1145/2702123.2702230>
- [23] Harm Veling, Guido M van Koningsbruggen, Henk Aarts, and Wolfgang Stroebe. 2017. Targeting impulsive processes of eating behavior via the Internet: Effects on body weight. In *The Goal Conflict Model of Eating Behavior*. Routledge, 160–178.
- [24] Joshua H West, P Cougar Hall, Carl L Hanson, Michael D Barnes, Christophe Giraud-Carrier, and James Barrett. 2012. There's an app for that: content analysis of paid health and fitness apps. *Journal of medical Internet research* 14, 3 (2012), e72.
- [25] Liu Yu. 2011. Vocabulary recognition and memorization: A comparison of two methods. (2011).