University of BRISTOL

Department of Computer Science

Harry's Habits: A Chatbot Investigating the Impact of Positive Reinforcement Rewards on Habit Formation.

Harry John Mumford-Turner

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Master of Science in the Faculty of Engineering

September 2017 | CSMSC-17

Declaration:

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of Master of Science in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

Harry John Mumford-Turner, September 2017

Executive Summary

Habit formation technologies use rewards and points as means for providing positive reinforcement, often in the format of visual or audio rewards such as jingles, badges or animations. Providing the right reward increases the chances for developing a new habit; yet, research on how these rewards should be delivered and the impact that this has on the process of habit formation is scarce. In this thesis we investigate how three types of positive reinforcement (visual, auditory, visual-auditory) influence habit performance and automaticity. Sixty people participated in a 4-week study where a custom built chatbot was used to deliver different types of positive reinforcement rewards for completing a new daily habit. The results reported higher habit performance rates when a reward was present without necessarily increasing behaviour automaticity. This has implications for the design of habit formation technologies that rely on visual and auditory rewards as means of positive reinforcement.

This research presents four main contributions:

1. A review of existing habit formation techniques, impact of rewards from different modalities and behaviour change interventions used in technology.
2. Design guidelines for building chatbots that deliver rewards based on different modalities.
3. Technical specification and implementation details of *Harry's Habits* — a chatbot to track habits and deliver rewards from three modalities: visual, auditory and visual-auditory.
4. A description and results of a 4-week user study where sixty people participated to test the impact of each reward on habit automaticity and habit performance.

Acknowledgements

I would like to thank Dr. Oussama Metatla and Dr. Katarzyna Stawarz for their patience, guidance and supervision through this project.

Contents

1 Introduction

Understanding how to design systems that support behaviour change is important to human computer interaction (HCI) — the field of computer science that studies how people interact with computers — to ensure that designers build these systems to have the maximum impact [1]. Habits play an important role in behaviour change by making the changed behaviour permanent [2]. However, the full habit formation process within the HCI domain needs further research due to the difficulties in evaluating the long-term effects of technology on habit formation [1]. Therefore, the need for HCI researchers to design technology that encourages people to change their behaviour and form new habits is still a main focus.

Habits are automatic actions that require little concious effort [3]. To develop new habits people must keep to a strict routine and perform the action repeatedly to strengthen the automaticity of the action [2]. When strong habits are developed the likelihood of behaviours persisting is higher [4] and habits are more effectively developed when specific and measurable goals are set [5]. The habit should perform automatically in response to a trigger that has been actioned repeatedly in the past [3]. The more automatic the response the higher habit automaticity. For example, a simple action such as turning on the light when you enter a room, happens automatically, even if the light is already on. However, forming a new habit is difficult and people are more likely to give up due to their lack of routine [2, 3].

Technology can help people stick to a routine by sending repeated messages [6] and encouraging people [7]. However, these techniques do not always work and may build repetitive actions rather than habit automaticity [8]. Therefore, technology should be designed to avoid building repetitive actions and instead build habit automaticity. Repetitive actions lead people to become dependent on technology and when the system is eventually removed, habit performance decreases [9, 10]. Habit automaticity is a measure of habit strength [11] and is the key that removes this dependency [12]. Automaticity can be increased by building motivation to complete the action [13, 14] and motivation can be encouraged by giving people positive reinforcement rewards after they complete an action [7]. However, how the reward is delivered and the type of reward used is also crucial to success.

The method of delivery should suit each individual user and a choice of delivery should be available. For example, a survey on feedback systems [15] advised that delivery of interaction should span different sensory modes to increase retention and better suit the needs of users. In the context of HCI, a modality or mode is the classification of a single independent channel of sensory input or output between a computer and a human [16]. Although interaction across modalities is important, this project does not have configurable feedback, but aims to compare how each type of feedback can affect motivation. Monetary (extrinsic) rewards can hinder motivation [14], whereas, satisfaction-based (intrinsic) rewards can be beneficial to motivation and should be preferred. This project investigates the impact of intrinsic positive reinforcement rewards on habit formation when delivered by a chatbot — a method of communicating with a computer system via a conversation. *Harry's Habits* was built to help conduct a research trial and to achieve the following aims and objectives. The source code is available and open-source (`www.github.com/harrymt/harryshabits`).

1.1 Aims and Objectives

The project aims to answer three research questions (RQ).

RQ1: What is the impact of modalities on habit formation?

RQ2: What are people's attitudes towards using a chatbot?

RQ3: How effective are chatbots for supporting habit formation?

A mixed-methods approach will be used to answer the research questions. This approach is common in HCI research [17] and combines qualitative and quantitative methods [18]. Therefore, to answer the three research questions the following objectives must be competed that use quantitative and qualitative methods:

1. Present an overview of existing research into:
   (a) Habit formation and behaviour change techniques.
   (b) Different types of rewards focused on positive reinforcement and how they impact motivation within habit formation.
   (c) Visual, auditory and visual-auditory feedback on habit performance and habit automaticity.
   (d) Existing technology used to form habits.
2. Construct theory-based design requirements that focus on positive reinforcement rewards for habit formation.
3. Create a prototype based on the requirements that is grounded in theory, to track habits and provide positive reinforcement using visual, auditory and visual-auditory modalities.
4. Conduct a 4-week study to answer RQ1 by using the following two hypotheses during a 4-week study.
   (a) Hypothesis 1: positive reinforcement is effective at supporting habit formation by increasing automaticity and regular habit performance
   (b) Hypothesis 2: rewards from multiple modalities are more effective then singular mode rewards
5. Conduct and analyse interviews with participants after the 4-week study to answer RQ2.
6. Analyse the data from the study with participant interviews to answer RQ3.

1.2 Added Value

Evaluation from real people in a 4-week study reveals insights and adds value in four places. First, the positive and negative aspects of using a chatbot to track habits and collect data during a research trial are revealed. Second, the evaluation of Harry's Habits provides value on how to build a chatbot to deliver rewards from different modalities to support habit formation. Third, analysis about the impact of the chosen visual, auditory and visual-auditory rewards on habit automaticity and habit performance gives insight into using these modes for behaviour change interventions. Finally, the project opens up new research avenues for investigating the use of chatbots as vehicles for promoting behaviour change.

2 Background

2.1 Habit Formation

Habits are automatic actions that require little concious effort [3]. To develop new habits people must keep to a strict strategy and perform the action repeatedly to strengthen the automaticity of the action [2]. When strong habits are developed the likelihood of behaviours persisting is higher [4] and habits are more effectively developed when specific and measurable goals are set [5]. Changing behaviour requires the formation of new habits to make this change permanent. Three elements are needed to form a habit: positive reinforcement, repetition and contextual cues [19]. Positive reinforcement rewards the person by encouraging them to perform the action again until it forms into a habit which significantly increases intrinsic motivation and increases the persons perception of their own performance [7]. Context from a cue acts as a trigger with constant repetition (depending on the complexity of the behaviour) from 18 to 254 days [20]. However, people still fail at forming new positive habits and give up, often due to their lack of routine [2, 3].

Contextual Cues and Repetition

Context from information around the action, serves as a cue to trigger events to push the person onto performing the action. For example, if you wanted to adopt a habit to floss your teeth, you could attach it onto an existing habit like brushing your teeth. The contextual cue of brushing your teeth will trigger you to floss them. Behaviour change literature [21] shows that attaching habits onto existing event-based cues are easier to remember, when compared with time-based habits, e.g. take a break from your computer screen every `x` hours, mainly due to changing environments [8]. Event-based cues help connect the contextual information with the action and builds habit automaticity [22]. Further research into the design implications of contextual cues shows how multi-cue routines are more effective than a single cue [23]. Finally, the process of creating a new habit takes on average 66 days of repetitive use [20]. The easier the action, the shorter time before the action turns into a habit, from drinking water (18 days), to going to the gym (254 days). But, rewards are needed to motivate the desired action, before the action develops into a habit [24, 25].

Rewards

Rewards give motivation, fuelling the belief in success and self-efficacy, which plays a large part in forming habits. Some researchers [13] suggest it is the main part of behaviour change. Variable types of rewards have been shown to increase dopamine in a laboratory study on rats [26]. This technique has proved to be an effective method of increasing repetition as shown in slot machines that vary their reward payout [27]. However although variable rewards increase habit repetition, they hinder habit automaticity [26], which is key to creating permanent and long-lasting behaviour change. Rewards are a good form of external motivation because they don't change the ability to perform a behaviour, unless the reward itself is a tool that increases ability [28]. Rewards provide a strong motivational source, but like all extrinsic motivators, these are less effective for changing behaviour in the long run, because externally motivated behaviour lasts as long as the external motivator exists [12]. One type of reward is positive reinforcement.

2.2 Positive Reinforcement

Positive reinforcement rewards the person by encouraging them to perform the action again until it forms into a habit. This reward increases intrinsic motivation by giving the feeling of satisfaction [2] and increases the persons perception of their performance [7] especially when the task is interesting [14]. Technology can use different type of rewards from different modalities to deliver positive reinforcement.

2.3 Modalities

Positive reinforcement feedback can be delivered by technology from different modalities. Visual, auditory and visual-auditory modes are the most common for issuing feedback in HCI [29]. Feedback from each of these different modalities has been shown to improve task performance for small tasks [30]. Combining modalities can be useful for issuing user feedback, enabling systems to be accessible to varying ages [15]. Individual modes can improve the effectiveness of reminders [31, 32]. However, the effectiveness of each type of modality as reward feedback and how combined modalities impact interaction when used by technology has little research. Therefore, technology to support behaviour change should explore a range of modalities for feedback. Research into visual, auditory and visual-auditory feedback was identified to understand how they impact behaviour and habit formation.

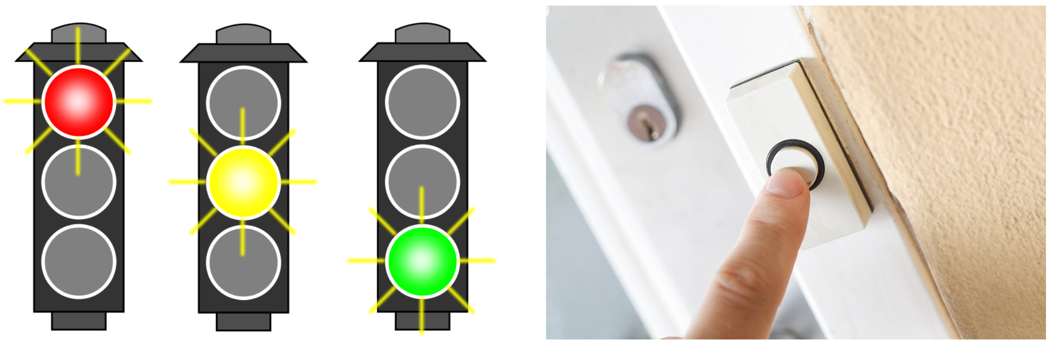

Figure 1: An example of visual (left: traffic light) and auditory (right: doorbell) feedback.

2.3.1 Visual

Fast visual feedback is preferred over audio when the relative importance of the task is high [33]. Visual feedback can encourage task performance consistency as demonstrated in one study by Lee et al. [10] where they constructed a device that gave constant visual feedback for patients taking medication. They found that visual feedback improved consistency of the habit and increased the rate of self-efficacy. But when the device was removed, their performance dropped (as measured after a 2-month period). This suggests that users did integrate the visual feedback display cue with their routines, but they also became dependant on the technology and the visual feedback did not build habit strength. Users should instead build these cues outside of the system, with another routine to build habit automaticity and allowing for permanent behaviour change after the system is removed.

2.3.2 Auditory

There is a need to use audio when designing behaviour change technology, especially with a varied target market [34]. For example, combining different sounds for different actions to suit different users [35]. Auditory feedback from levels under 94dB increases task performance [36] and using auditory as a method of delivery, has shown to improve task success [37], but little work has shown how it impacts habit performance.

2.3.3 Visual-Auditory

Several studies show that combining audio with visual as feedback after performing a simple task can be successful [38, 39]. A meta-analysis of 43 multi-modal studies [40] revealed that visual-auditory feedback was the most effective at increasing performance, when a single task is being performed and when compared with visual or auditory feedback alone. Additional research shows this combination of visual and auditory sensory channels has been shown to increase performance with complex tasks [30]. However, care needs to be taken when adding an extra modality as this study showed that visual feedback improved task performance, but sacrificed task quality [40]. While all modalities contribute to perceptual experience, one sense can override another if the sensory channel mediates less ambiguous information than the other [41]. Therefore, using a combination of modalities for interaction gives a means of communication to people with varying levels of sensory awareness.

2.3.4 Tactile Vibration

The majority of electronic activity monitors already use behaviour change techniques and therefore present a medium where tactile vibration behaviour change interventions could occur [34, 38]. However, low levels of vibration levels should be used, as high levels decrease task performance [36]. A survey on activity monitors [42] ranked *Fitbit* (`www.fitbit.com`) devices as good vehicles for behaviour change techniques. Therefore, the Fitbit, would be a good primary platform for integrating vibration for rewards. However, due to technical limitations, as discussed in Section 6, they were unable to be implemented.

2.4 Existing Technology

A survey into how technology can support habit formation and behaviour change [43], reports a large number of habit forming systems are mobile apps. Studies into the effectiveness of these apps has been recently conducted [9, 12] revealing that although most of these apps are rated highly, they do not ground themselves in behaviour change theory. Further surveys of these apps [44] suggest that habit performance is not sustained when the app is removed, due to the lack of habit automaticity built during the habit tracking process. Using apps consistently to manage behaviour can create a notable difference in the person when the system is removed [45]. This is also the case with many behaviour change systems, when the system is removed any improved performance is lost [9, 10]. Building habit automaticity requires the desired action to be built around an existing routine [20, 21]. Technology should help with routine creation by having additional checks to guard against changes in routine that remind people about their habit if their situation changes and issue post-completion notifications to check if the action has already happened [46]. This is vital for sustaining performance and building habit automaticity.

Choice of System

Although there has been lots of research using mobile apps for behaviour change interventions [43] (Figure 2). There are plenty of other options available when designing be behaviour change interventions. One option that has been recently used for behaviour change interventions are *chatbots*. They have been integrated into messaging apps, such as the *Telegram* app to track food habits [47] and are seen as the next big thing in HCI [48]. Chatbots are applications that parse questions using Natural Language Processing (NLP) to provide a response and are platform independent. This means the bot can be accessed on mobile phones and desktops. Bots act as a user interface to expose data and would use online services to parse the response, such as Amazon lex (`https://aws.amazon.com/lex/`). These programs have conversations with users to achieve a goal and are not new inventions. Since 1966 [49], Eliza by Weizenbaum, used simple expression matching to return a certain response for user trials. In the present day, these applications (referred to as both bots and chatbots), are found integrated into many different apps on the majority of users mobile phones. For example, *Facebook Messenger* (`www.messenger.com`), a popular messaging application encourages developers to create bots to interact with their users. These bots act as a real person with similar interaction flow, plus a few additional features, such as *Quick Replies* [50] for revealing a list of options to a user. Quick Replies provide a way to present buttons to the user in response to a message (Figure 4). However, these bots would not reply like a real person, but rather would only reply if that question was pre-trained using machine learning algorithms. This technology requires the bot to be trained on a large set of data and the majority of use cases would have to be accounted for. Natural language processing would enable users to chat to the bot and get a friendly understandable reply. However, this interaction may build a dependence on the user-chatbot interaction and could lead to losing automaticity.

	Mobile App	**Cross-Platform App**	**Webapp**	**Chatbot**
Notifications	✓	✓	✗	✓
Development Time	Long	Medium	Short	Short
High Availability	✗	✗	✓	✓
Simplicity	✗	✗	✓	✓

Figure 2: Comparing different options for behaviour change interventions.

Another design option is building a *webapp*. However, the functionality is limited, due to the lack of the notification feature in current webapps. To allow notifications the webapp could be paired with another native app or use SMS (Short Message Service) notifications. However, because these technologies are separated it is more difficult to get users to respond and install a separate app. Another design option is a native mobile app as has the ability to supply notifications, but for each platform a completely separate app would need to be built and users would need to download the app before it would be available to them. A single cross-platform app could be constructed to reduce development time and complexity, but still users would need to download the app to start using it. A webapp has the advantage of being available to all users with a web browser (with users being able to save the site to home screen), but without notifications on all platforms (iOS only), it won't meet our requirements. Finally, a chatbot integrated into a popular messaging platform is easily available (if you have the messaging app already installed), simple (the user interface is already supplied), works on any platform the messaging app is available on and has notifications built in. The table (Figure 2) summarises the available development choices for a system that meets our aim.

2.5 Platform

The platform for the tool needed to be highly available for participants, interactive and time effective to build. A chatbot provides us access to both the mobile platform and the desktop platform, granting us access to a highly available, contextually aware and interactive platform [45, 51].

There are lots of options about what platform to build the chatbot into (Figure 3). For example, *Slack* (`www.slack.com`) bots (Figure 4) provide additional functionality to the popular workplace communication service and can complete complex tasks such as habit tracking (`https://healthybot.io`). *Whatsapp* (`www.whatsapp.com`) also provides interactive features with a large user base, however it lacks many additional features, such as quick replies. *Telegram* (`www.telegram.org`) and *Facebook Messenger* (`www.messenger.com`) have these additional features, but only Facebook Messenger has the large user base. Therefore, because our main aim is to interact with lots of people easily, we need to target existing platforms that have high participant availability.

	Facebook Messenger	**WhatsApp**	**SMS**	**Telegram**	**Slack**
High Availability	✓	✓	✓	✗	✗
Interactive	✓	✓	✗	✓	✓
Additional Features	✓	✗	✗	✓	✓

Figure 3: Comparing different chatbot platforms.

Facebook Messenger looks like the attractive option for user interaction with the ease of additional features, such as quick replies and with the benefit of:

- 1,200 Million active users per month (as of April 2017) [52]
- Embedded into a service users already use
- Quick replies allows for easy interaction

The success of the chatbot will depend on how people differentiate between the bot and another contact and if people prefer the interaction.

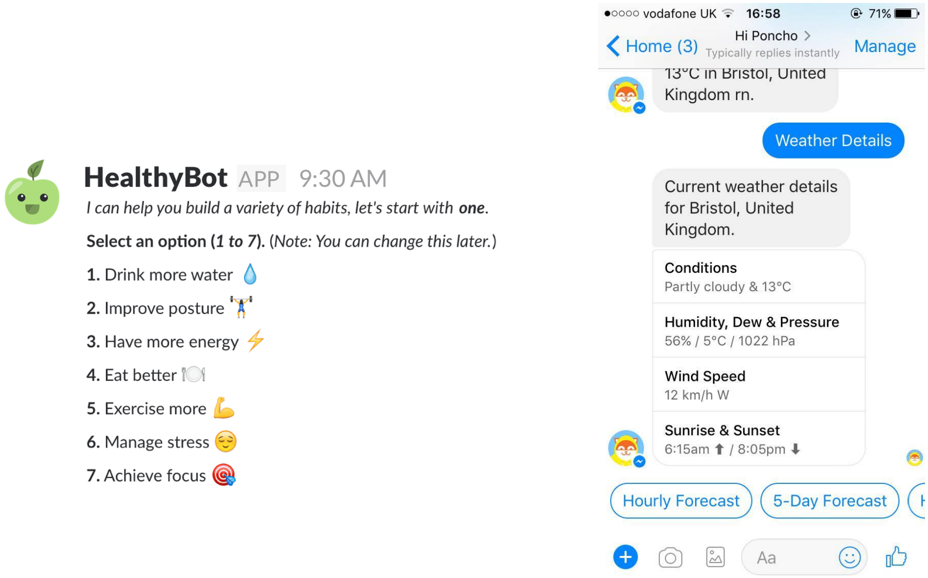

Figure 4: Left: *Healthy Bot*: a Slack chatbot for forming new positive habits. Right: *Poncho*: an example of Facebook Messenger Weather Chatbot.

Facebook Messenger

Interaction with existing Facebook Messenger chatbots is varied. Existing bots use three methods of interaction that utilise different areas of the Facebook Messenger interface. The first method fully utilises natural language processing. The bot sits patiently until it receives a message, then it sends the message off to a service that performs processing and returns the message breakdown. The bot then chooses an action based on the message. For example, the Poncho weather bot (`https://poncho.is/`) displays the forecast for Bristol if you message *bristol forecast*. As previously discussed this method of interaction does not suit our needs because we do not require users to ask questions to the bot.

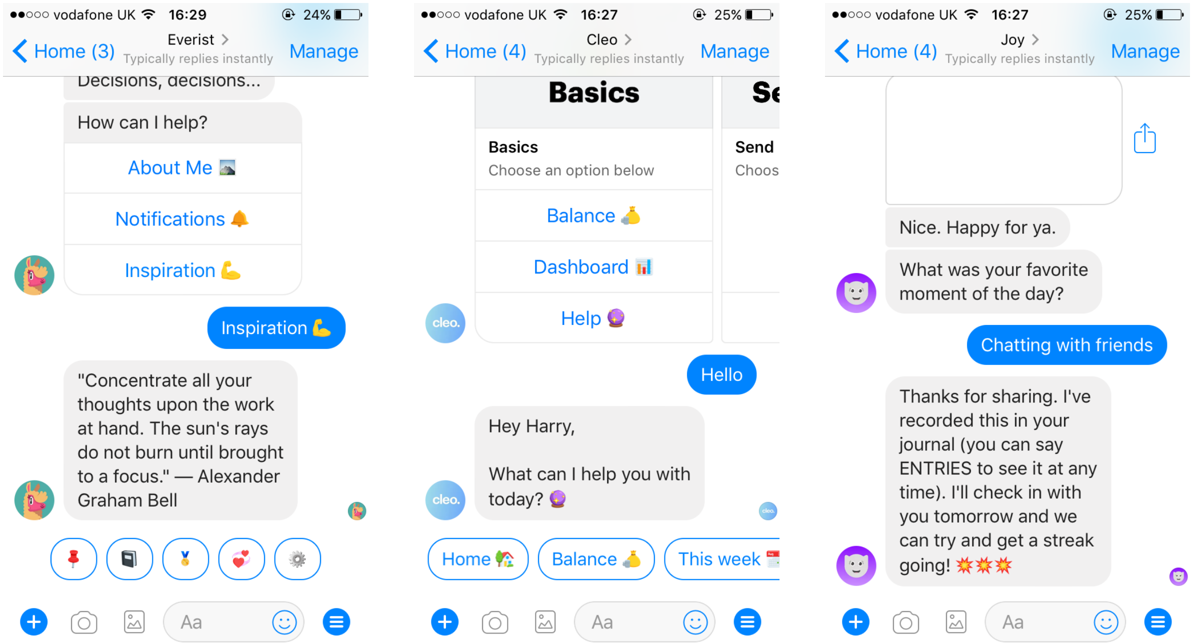

Figure 5: Examples of existing chatbots that help track and form new habits with different methods of user interaction. Left Everist (`https://www.everist.ai/`), middle Cleo (`https://meetcleo.com/`), right Joy (`http://www.hellojoy.ai/`).

The second method does not use natural language processing and quick replies are heavily utilised. Everist and Cleo bots (Figure 5) use quick replies almost like a menu with a set of options. After users message the bot once, the set of quick replies are returned to encourage users to only use these as the method of communication. Although this makes coding the prototype easier, it is difficult to get *free text* from users. Finally, Joy (Figure 5) uses a combination of free text and quick replies to set-up users and interact with them every day. This is the combination that Harry's Habits will use to get free text from users and enable quick choice from a menu of actions.

Supporting Habit Formation

Harry's Habits will be built Facebook Messenger — an existing social network that people are used to — to track habits and deliver rewards. The social scaffolding around the interaction and the integration with the platform will mean that interactions with the bot will be visible alongside participant's other conversations, making it easier to report the completion of the habit [53]. The bot will track simple actions of similar difficulty to reduce the amount of time needed before the action turns into a habit [20]. Other design methods to help people form new habits were looked at but will not be added to the prototype to limit the scope of the project. For example, gamification, such as using gambling elements and engineering luck into games have been shown to encourage interaction [54] and using monetary rewards to create a feedback loop where people keep coming back for the reward [55]. Both of these elements can help with interaction but do not build habit automaticity, therefore leading to a drop in performance when the feedback is removed.

2.6 Design Requirements

After reviewing literature from habit formation, modality task performance and existing habit formation technology. A new set of design requirements is created, for building a chatbot that is grounded in habit formation theory and focused on rewards.

REQ 1. Help users define a memorable strategy.
Help users make personalised routines and provide examples of existing strategies based on REQ 1, 2 from [12] and REQ 2, 4 from [28].

REQ 2. Remind them about cues and remembering strategies.
Reminders can effectively support prospective memory in the short-term, increase the logging of health data [53] and educate them about what they should perform in the long-term. Based on REQ 4 from [12] and REQ 3 from [28].

REQ 3. Reward users.
Positive reinforcement rewards are a good form of external motivation. Based on REQ 4 from [28].

REQ 4. Send positive reinforcement from different modes.
Task performance is impacted when sending positive reinforcement from visual, auditory and visual-auditory modalities.

REQ 5. Check if the action has already happened.
People find it easy to forget whether an automatic task was completed. Therefore a post completion notification to check to see if the action has already happened should be issued, based on REQ 6 from [12] and REQ 1 from [28].

REQ 6. Disable cue reminders when behaviour is routine.
Relying on reminders in the long-term can hinder habit development, therefore ease off from reminders later. Based on REQ 5 from [12] and REQ 5 from [28].

3 Design

A prototype was constructed to track habits and deliver rewards based on the requirements from Section 2.6.

Track Habits

To help users define a memorable strategy (REQ 1) users are allowed to personalise the experience slightly with their own routine and choice of habit with examples given during the setup phase (Figure 10). Post completion checks are delivered after users existing routine to check if the action has already happened (REQ 5) and remind users about their cue (REQ 2). Finally, positive reinforcement is delivered after the habit has been performed from a modality to improve habit performance (REQ 4).

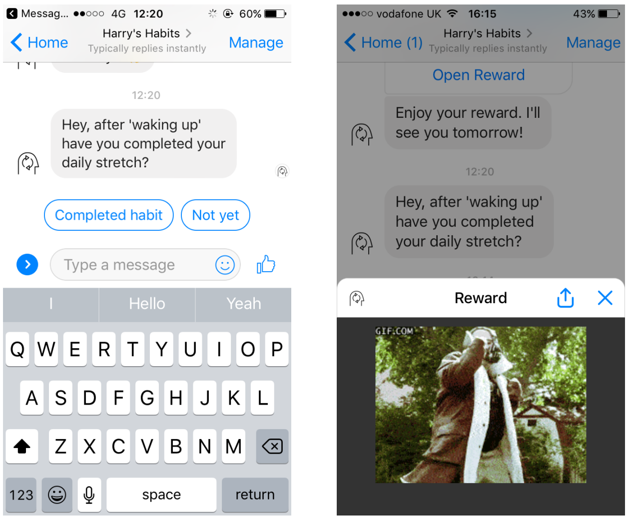

Figure 6: Harry's Habits delivering a post completion message (left) and an example of a visual reward (right).

Delivering Rewards

To reward users (REQ 3) and to answer all research questions. Rewards from three different modalities are delivered by the bot: visual, auditory and visual-auditory. The content of these rewards are experimented with. For the visual rewards, popular internet memes are used, i.e. content that is passed along from person to person via social media posts [56]. Humorous GIF memes make people laugh and are popular on some social media sites where they usually are the most engaging content [57]. Given that the bot is integrated with Facebook, and that Facebook introduced built-in support for animated GIFs [58], memes will integrate well with the bot environment. These rewards are displayed to the user within the chatbot after they complete their habit (Figure 6). The modalities are mapped to identify a pattern across the modes before they are implemented and adapted for delivering rewards. The rewards were aimed at motivating participants to keep coming back every day and completing their habits. Several GIFs and audio files were identified that categorised themselves as motivational and each GIF file was vaguely matched and tweaked to match the audio frequency. The relationship between the audio and visuals is inferred, therefore this mapping is *semi-congruent*.

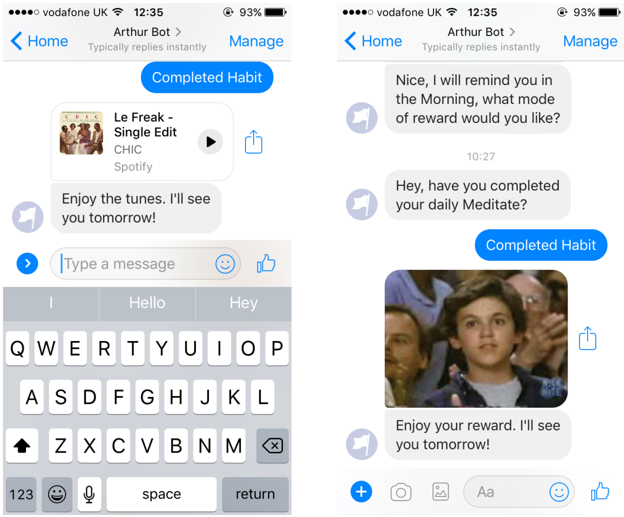

Figure 7: Inconsistent design with auditory and visual rewards when sent as a inline message.

Design Interface Consistency

During the design phase two different interface methods are discovered for delivering visual, auditory and visual-auditory rewards to users. First, GIFs and audio sent as a inline message to the user (Figure 7). Second, a consistent method displays the reward from a website and links it to users (Figure 8). The first method (inline message) has the benefit of being native to Messenger, providing a better user experience, is faster to receive and requires less buttons to be pressed. However, each reward is not displayed consistently. Visual rewards started as soon as they were delivered, but the auditory reward had to have a *play* button that had to be pressed before the audio started and the visual-auditory rewards also had a button that needed to be pressed. Inconsistency with delivery would make it difficult to evaluate the effectiveness of each reward. Therefore the second method is chosen (Figure 8) to allow for consistency when delivering rewards.

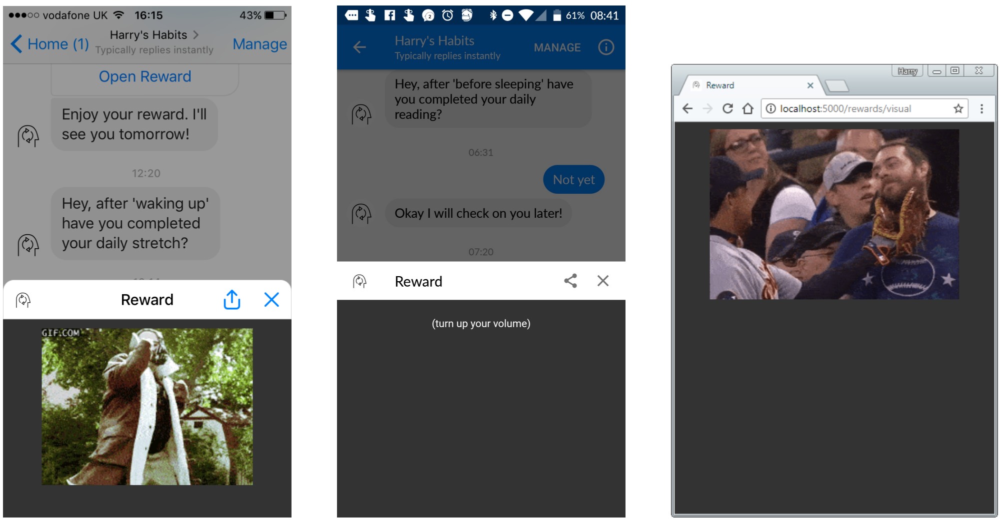

Figure 8: Consistency between every type of reward modality on all platforms, displayed by a website. Left, iOS, visual reward. Middle, Android, auditory reward. Right, browser, visual-auditory reward.

3.1 User Flows

Full user interaction from setup to completing a reward was designed. The setup used a combination of quick replies and free text to gather demographic information and complete the setup for each user.

Figure 9: Harry's Habits logo was design by the Noun Project by Yu luck (`https://thenounproject.com/term/custom/402041/`).

3.1.1 Setup Flow

- Press button that opens the bot in the Facebook Messenger app.
- Answer a series of setup questions to the bot via the Facebook Messenger app.
- User chooses an existing habit they would like to develop from a list.
- User supplies an existing routine to integrate their new habit into.
- User chooses a time the existing routine normally happens.
- User finishes the setup and closes the Facebook Messenger app.

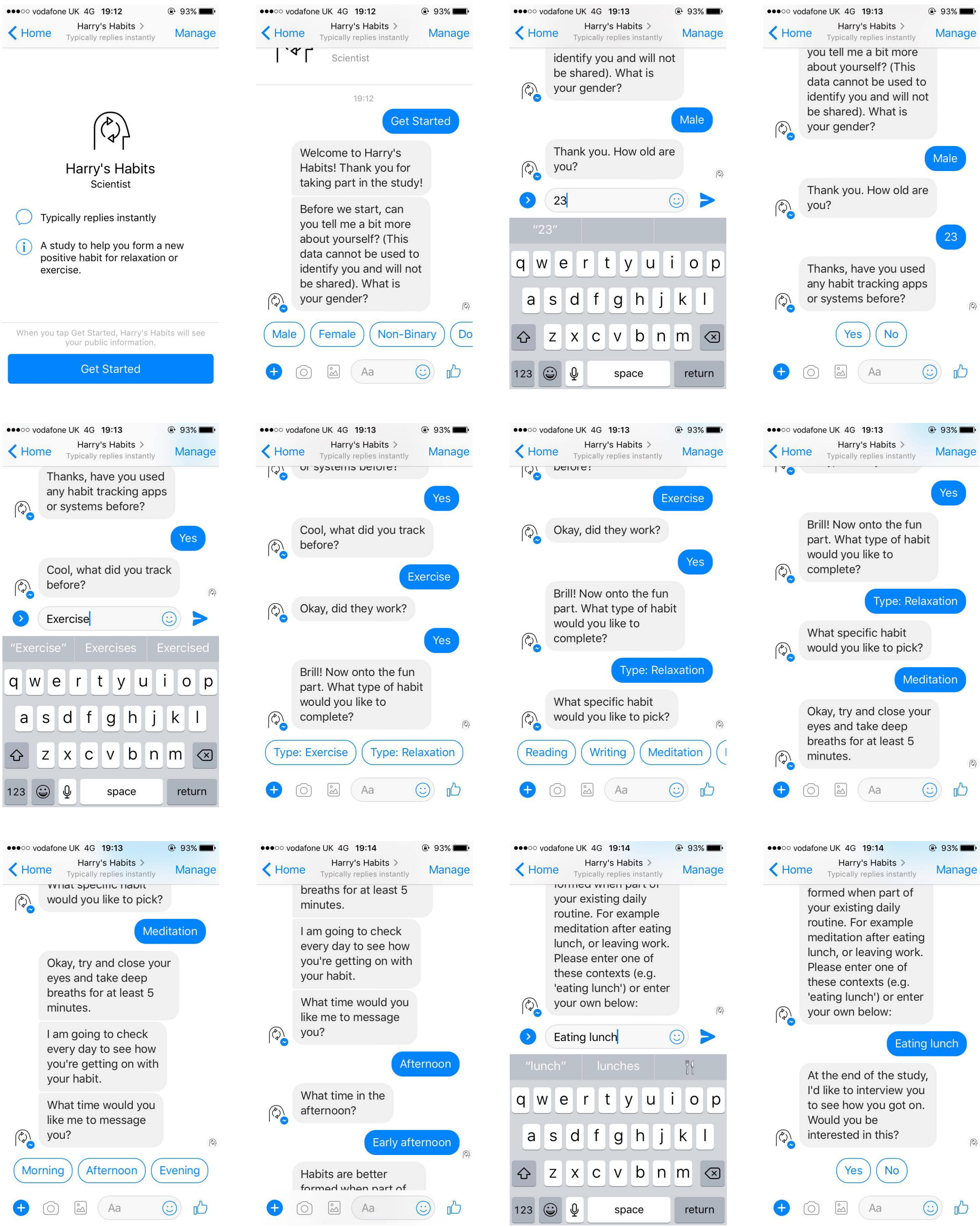

Figure 10: Setup flow for Harry's Habits.

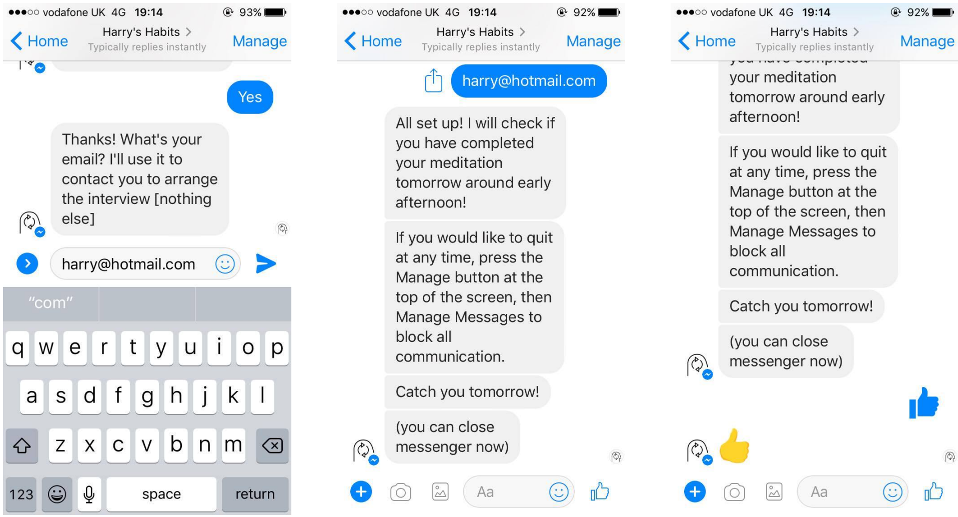

Figure 11: Continued: Setup flow for Harry's Habits.

3.1.2 Trigger Flow

- At the time of the existing routine the user performs their chosen habit.
- The user receives a notification after the routine, asking if they managed to complete their habit or if they need more time.
- If they need more time, the notification will *snooze* for about an hour and be sent again.
- If users regularly snooze they will be asked if the time of their existing routine has changed.
- If users say they have completed their habit, they will be sent a quick reply message that links to a reward.

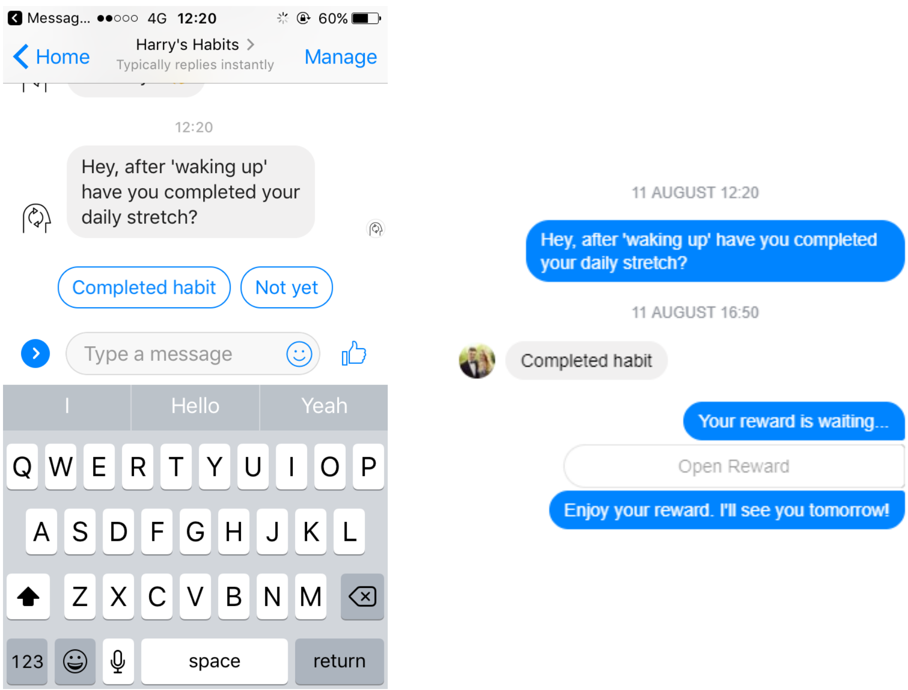

Figure 12: Trigger flow for Harry's Habits.

3.1.3 Reward Flow

- Users will press the reward button that will take them to a website that contains a link to open a reward. This allows the experience to be consistent for each reward modality.
- User will receive a reward from one of the following modalities:
  - Visual: GIF with no audio
  - Audio: soundtrack audio
  - Visual-Auditory: GIF with soundtrack audio

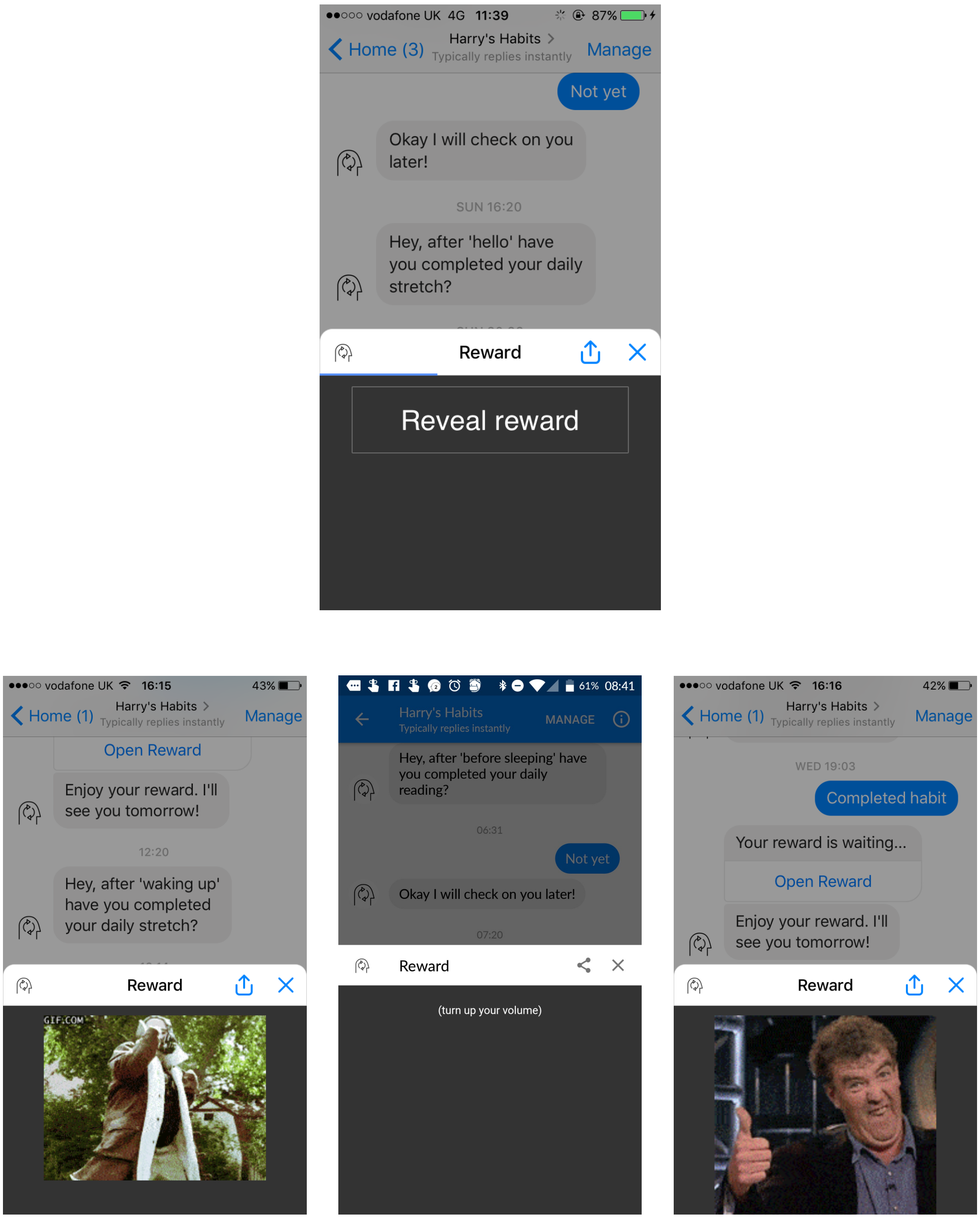

Figure 13: Reward flow for visual, auditory and visual-auditory rewards for Harry's Habits.

4 Implementation

Figure 14 describes interaction between the chatbot and the Facebook Messenger servers. The user interacts with the chatbot using the Facebook Messenger platform on either a mobile phone or the desktop via the website (`www.messenger.com`). When the user sends a message to the chatbot, it is sent via the Facebook servers who decide where to forward the message onto. The message comes in a *json* format that is parsed by the chatbot application to understand the content of the message and how the user interacted with the bot. For example, if the response is from a quick reply, then additional parameters are sent. The chatbot reads and writes the necessary information into a database to decide a response. When it is ready, it sends a request to the Facebook servers that forward it onto the user's Facebook Messenger account that can be read from the Messenger platform (Figure 14). All code is open-source and hosted on *GitHub* (`www.github.com/harrymt/harryshabits`).

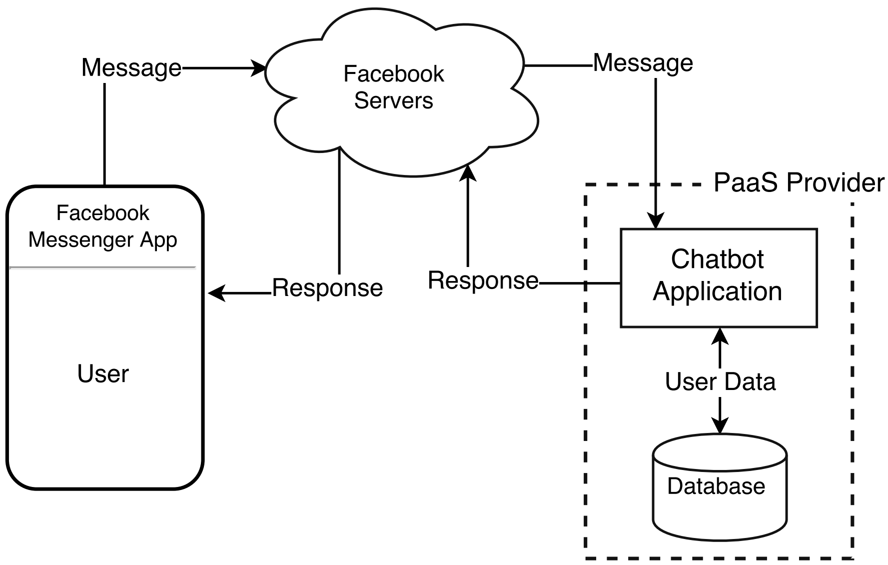

Figure 14: Interaction between a user and the chatbot application via Facebook servers.

The prototype is written in JavaScript running on a *node.js* (`www.nodejs.org`) server, built on the Facebook Messenger chatbot platform and hosted on *Heroku* (`www.heroku.com`) — a free PaaS (Platform-as-a-Service) option. Facebook Messenger encourages developers to create bots to interact with their users as these bots act as a real person with similar interaction flow, plus a few additional features, such as *Quick Replies* [50] for revealing a list of options to a user. Simple call and responses were used to interact with users and track their data instead of natural language processing to limit the scope of the project. Instead of natural language processing, the location of the bot (inside an existing messaging app) ease of interaction and the additional features (quick replies) help easily communicate with users.

Different languages, server, hosting provider and database provider were considered. JavaScript with node.js on Heroku with *PostgreSQL* (`https://www.postgresql.org/`) were chosen because of the following reasons: node.js enables us to share the same language for the server and the client; large amount of packages to handle client and server side functionality; suitable for prototyping and rapid product iteration; Heroku has a free-tier which allows full deployment with a PostgreSQL database of up to 10,000 rows; Heroku hosts the application in the cloud, giving benefits for scalable deployments that benefit any potential future application growth. Finally, *Airtable* (`www.airtable.com`) — a database integration was tested, however, it only provided 3,000 free rows and therefore was discarded in favour of the PostgreSQL database hosted on Heroku.

4.1 Database

The bot tracked data about how people logged their habits, such as what day they tracked their habit and how many times they delayed their checking messages. Heroku provides 10,000 rows as their free option in a *PostgreSQL* database. Figure 15 outlines each table in the database with what information is stored. The *Facebook user ID* is used to identify each user and what habits they tracked. When a user has told the bot they completed their habit, a new row in the *Habits* table would be added and linked to the user in the *Users* table. Two global variables are used to maintain the state across local and live versions of the system to display the length of study and if it is active.

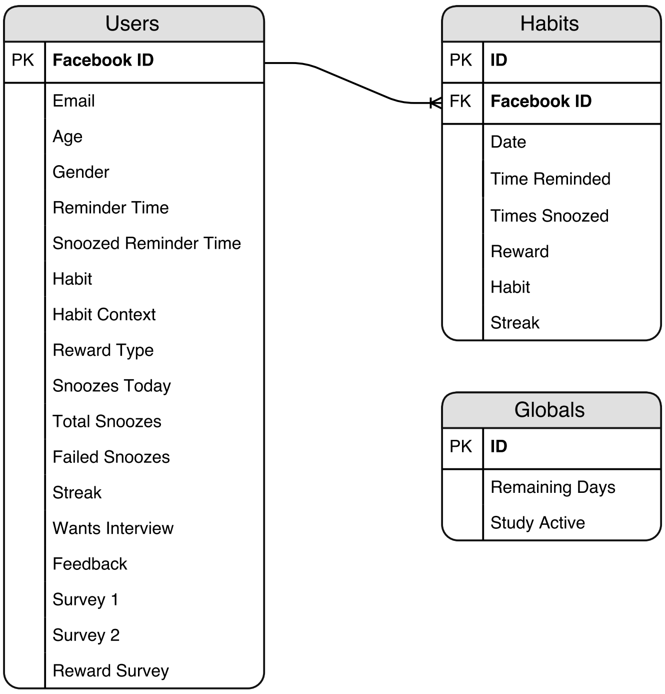

Figure 15: PostgreSQL Database entity table relationship diagram showing all information stored for Harry's Habits.

4.2 Custom Application

Figure 16 shows a detailed overview of the custom chatbot application and how it interacts with the database and the Facebook servers.

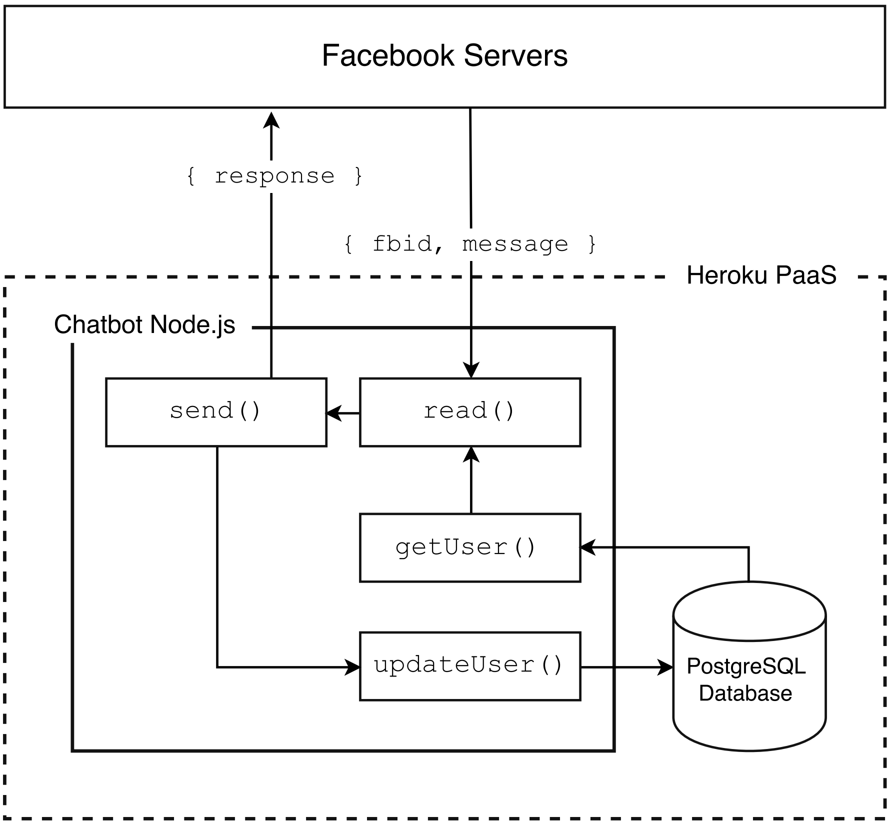

Figure 16: Overview of Harry's Habits interacting with Facebook servers to send and receive messages from the Messenger platform.

When a user messages the chatbot on Facebook Messenger, the message is first sent to Facebook servers who make a HTTP request to a specified *webhook URL* with a secret parameter for security. The chatbot handles this request then sends another with the same secret. The incoming HTTP request contains a payload with either a simple string of characters with a Facebook user ID (*fbid*) or a quick reply key with a fbid. The chatbot application handles the incoming request with four components.

First, the `read()` function checks if the incoming object is valid, then decides if it is a quick reply or just a regular message and strips the object of the unimportant parts until it is left with the fbid and either the message as a `string` or the quick reply key.

Second, the `getUser()` function uses the extracted fbid to collect information about the Facebook user from the database. If the fbid already exists in the database, it means the user has previously interacted with the chatbot, so details about that user e.g. the type of reward, are extracted from the *Users* Table (Figure 15). Or if the user does not exist, it means the user has never interacted with the chatbot before.

Third, the **send()** function decides what message to send back to the user. If this is the first time the user has messaged the chatbot, then a setup quick reply message is created and sent to start the setup flow (Section 3.1.1). Otherwise, based on the quick reply key or the individual message, the chatbot will respond differently. For example, if a user sends 'help' as a message, the application will prepare to display a list of quick reply options that users can choose to show more information about the bot. After the message is prepared, the application sends a request back to the Facebook servers containing a regular message or a quick reply object with options and keys.

Finally after the message is sent, the `updateUser()` will create a new row in the database if there is a new fbid, or it will update information about the user if they responded with a quick reply. For example, during the Trigger flow (Section 3.1.2) the chatbot asks users using a quick reply, if they completed their habit. If a user marked a habit as completed, the message will have a quick reply key that the application can understand and a new row in the Habits Table will be created with that habit information.

4.3 Delivering Rewards

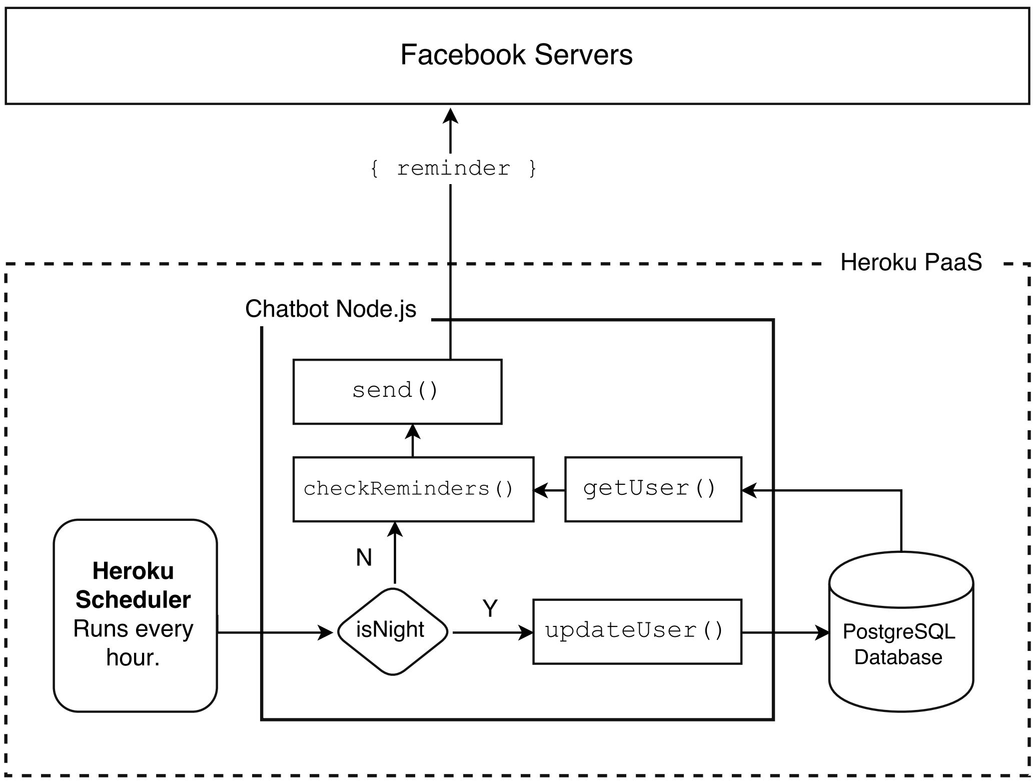

Figure 17: Overview of how notifications are sent by the application using a scheduler to check if it is time to send the notifications every hour.

A post completion notification is sent to a user just after the time of their habit context. Figure 17 shows how the Heroku scheduler (`https://elements.heroku.com/addons/scheduler`) handles this process by running a job on the chatbot application at scheduled intervals, similar to a *cron job*. The scheduler runs every hour to run a node.js JavaScript file. This script first checks if the current time matches any of the 11 pre-defined times (Figure 18) and if it does, it reads the database to get information about all users that want to be sent a notification at that time. Then sends a quick reply message asking the user if they completed their habit or would like to be asked later (unless it is the night). If the latter is chosen, the next reminder time later in the day is set to that user. Otherwise a reward is delivered.

```
const reminderTimes = {
    earlyMorning: 7,
    midMorning: 9,
    lateMorning: 11,
    earlyAfternoon: 12,
    midAfternoon: 14,
    lateAfternoon: 16,
    earlyEvening: 18,
    midEvening: 20,
    lateEvening: 21,
    night: 22,
    newDay: 23
  };
```

Figure 18: The JavaScript object that defines the scheduler times that users can set to receive post-completion notification.

The type of reward delivered is chosen when a user sends the 'completed habit' quick reply response. The reward modality is based on the fbid of the user and is read from the database, then another message is delivered to that user containing the reward. A consistent method was built to deliver the reward to users, instead of sending the rewards in-line a *webview* was used to display a website where users can open their reward. The website uses *HTML*, *CSS*, *JavaScript* and also used server-side template rendering with *Pug* (`www.pugjs.org`) and CSS preprocessing with *SASS* (`www.sass-lang.com/`). The full source code for the webview is available on GitHub (`www.github.com/harrymt/harryshabits/tree/master/docs`). This also allowed us to break free of the Messenger chatbot sandbox and use *HTML* elements to display the content in the same way, ensuring consistency across devices. The website could start the GIF or play the music or both for each reward type when a user pressed a button. Although this was not without limitations. The auditory reward would not stop playing when users closed the webview. This could be performed programatically, however during testing, would not always work. This was worked around by only playing the music for 15 seconds — an appropriate time to view the reward. Auto-playing the auditory reward was not available when sending the audio in-line and the webview could use the *HTML5* `<audio>` element to enable auto-play. But, for auto-play elements, the HTML5 standard needs a button press before it starts [59]. This required another button to create a *JavaScript* hook to auto-play. However, during tests on low mobile data speeds, users found that they would have to press the button multiple times before the audio played. This was because the audio would only play after it had loaded and created a lengthy delay, along with seemingly broken display. Figure 19 describes the process of loading a reward on fast internet speeds and slow internet speeds. To create a better user experience, the button disappeared when pressed, using *CSS* that would execute even if the audio hadn't loaded, and even on a poor connection. Then JavaScript would execute after the page had fully loaded and play the audio if the button had already been pressed from the CSS, but if it hadn't then it would create a hook to play the audio after it had been pressed. This ensured a seemly experience when using rewards for all levels of connection.

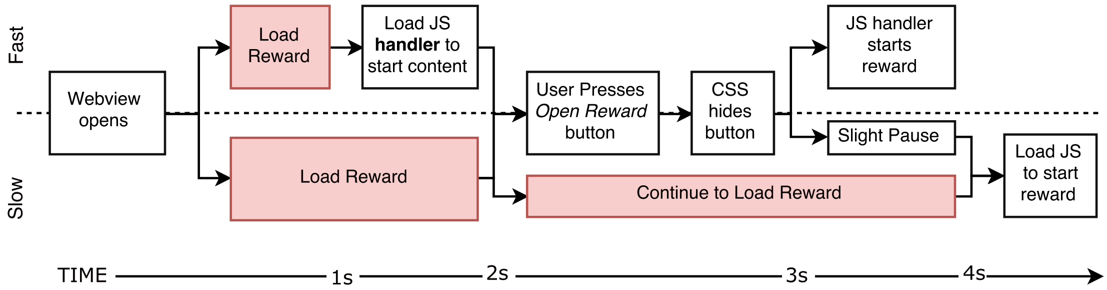

Figure 19: How rewards are displayed to the user when they are on fast internet (top lane) and slow internet (bottom lane) over time.

4.4 Testing and Continuous Integration

To get a measure of code quality, three online services: codacy (`www.codacy.com`), codebeat (`www.codebeat.co`) and code climate (`www.codeclimate.com`), were used to track quality over time (see more on the GitHub repo `www.github.com/harrymt/harryshabits`). A test harness was also written to perform functional testing on the chatbot. An on-line continuous integration (CI) service, Travis CI (`https://travis-ci.org/harrymt/harryshabits`) was used to programatically run these tests when a *commit* in *git* (the *version control* used) was performed on the *master branch*. *Travis CI* ran the defined tests to ensure the functionality worked throughout development. A pilot trial with three users also tested the basic chatbot functionality preparing for the full evaluation study (Section 5.1).

4.5 Technical Issues

Throughout the implementation process different techniques were explored to implement the design. Some of the research areas were not used in the final prototype due to technical issues and limitations with the approach.

Habit Context

During the pilot trails one participant added the word 'before' inside of their habit context (Figure 20). This resulted in post completion messages being awkwardly worded, e.g. "After 'before dinner' have you completed your daily press ups?". This would be solved by adding another option to ask if they would like to perform it before or after, when asking for their habit context. Users also reported that after they snoozed the notifications several times, their habit context no longer became relevant, e.g. "After 'eating lunch' have you completed your daily press ups?" at 7pm. The context should be removed or altered in snoozed notifications. In addition, users should be asked if they want to change their context when being asked to change the time they receive their post-completion checks.

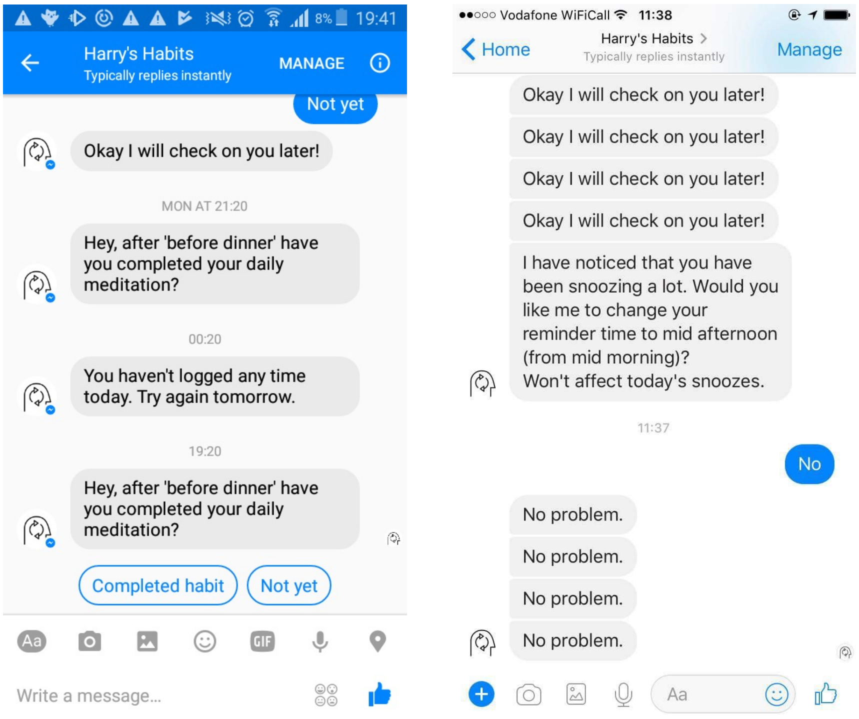

Figure 20: Example of 2 issues that occurred. First, the problem with habit context using the word 'after'. Second, the bot sending the same message multiple times.

Message Delivery

The chatbot would sometimes send multiple of the same message (Figure 20). This occurred when Facebook sent the same HTTP request multiple times. This could perhaps be improved by moving to a specific Facebook Messenger node JavaScript library module (`https://github.com/rickydunlop/fbmessenger-node`) to better suit interactions between Facebook servers and the application. This issue also effected another area during the process of handling free text input from users. When the bot asked users to enter in a free text, e.g. an email, a flag would be set, to signal a wait for free text input. However, if duplicate messages were sent at this time, the application would assume users did not enter anything for the free text and the flow would continue. Finally, if users accidentally sent the bot a message instead of using the quick replies, they would be unable to return to that quick reply menu and therefore break the flow of interaction. To fix this the question should be re-asked if it didn't fit the expected criteria.

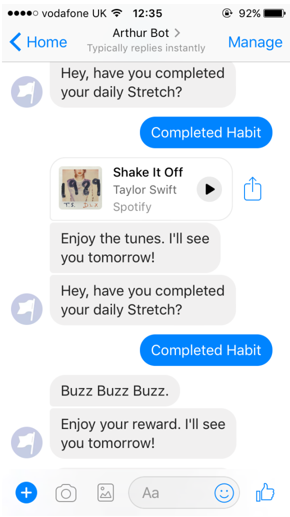

Figure 21: Vibration was tested, but due to technical limitations was difficult to implement.

Vibration

Using vibration as a reward modality would have been useful for evaluating the effect of tactile vibration on habit performance and automaticity. Unfortunately the chatbot sandbox prevented the vibration feature on the device from being used, so another device would be used in combination with the bot. Smart watches and fitness trackers were researched to test if they could programatically vibrate using the pattern of vibration to match the frequency of the audio. However, the majority of these devices did not have an API that exposed the vibration aspect. The best method found, was to programatically set an alarm 1-minute into the future using a Fitbit fitness tracker. This would trigger the vibration when the alarm sounded. Although this would mean a 1-minute delay after completing a habit, a good user flow could have reduced the wait time with some additional dialogue (Figure 21). But, this approach relied on the fitness tracker to sync with the phone after the alarm was programatically set. Unfortunately forcing the tracker to sync wasn't available, so this reward modality was abandoned.

Scheduler Time Zone

The scheduler had several limitations. First, the documentation stated it runs on a best effort basis and cannot guarantee to run every time, however, this did not impact the chatbot as it missed very few schedules. Second, the time zone of the scheduler had to be converted to Coordinated Universal Time (UTC) time before checking against the defined times. This issue was revealed in a pilot trail and fixed, however, for this application to work across different time-zones (ones that were not UTC), Figure 18 would have to be custom to each user.

Another issue occurred with stopping the audio after it had been played during a reward. If a user closed the reward box, there was no way to stop the audio, unless a user waited until it had finished. This limitation was very minor, but also showed how difficult it is to seemly connect a website and a chatbot. Finally, edge cases throughout interaction were revealed during development and were coded for, for example deciding what would happen if users snoozed their post completion notification when it reached the end of the day led to additional logic to stop users from snoozing when it reached the end of the day.

5 Evaluation

Below we present our evaluation process, that places the prototype in the hands of real people, as recommended by vaildated guidelines [1, 51].

Many people agree about the importance of designing systems for health and behaviour change [34, 51, 60]. But each have varying opinions about how to evaluate these systems. Klasnja et al. [1] focuses on system usability and does it meet the needs of users. Whereas, Stawarz and Cox [34] argue evaluating a system of this type requires information from other fields to properly consider the systems effectiveness. The validated Behaviour Change Wheel Framework [61] does just this: evaluating the system with validated behaviour change techniques from multiple domains. This project will use this framework to evaluate the chatbot with an evaluation trial. HCI research that focuses on health interventions [51], demonstrates the importance of evaluation trials for evaluating behaviour change systems. The length of the trial will be based on two factors, the time needed to form a habit [20] and the results of a previous habit formation trial [12]. First, the number of repetitive days required for an action to be considered a habit varies based on the complexity of the action [20]. Simple actions, such as drinking 2 glasses of water a day, can take a minimum of 18 days to form. The suggested actions used for this project will be considered as simple, e.g. stretching for 30 seconds. Second, a previous evaluation trial on habit-formation systems [20] showed an increase in habit automaticity after 4 weeks. This project will use that timeframe.

5.1 Study: Comparing Different Rewards

To explore the influence of different modalities on habit formation and answer the research questions, a situated study was conducted, followed by semi-structured interviews. A 4-week evaluation trial will test the success of the chatbot by evaluating the tool and the effectiveness of each modality on users habit strength. Cue reminders from the chatbot will be disabled at week 3 (REQ 6) to test if users continue with the habit. Participants will split into four groups, all groups will receive reminders, three groups will receive rewards each from a different modality, and one group (control group) will not recieve any rewards. Habit automaticity will be measured using a validated 4-question questionnaire to test users habit behavioural automaticity index [11]. This will show the impact each reward has on habit automaticity and test the following two hypotheses. Participants will fill out these questionnaires at week three and week four.

5.1.1 Hypotheses

To see what the effect positive reinforcement rewards from modalities have on habit formation (RQ1) and how effective chatbots are at supporting habit formation (RQ3) two hypotheses are presented.

Hypothesis 1: positive reinforcement is effective at supporting habit formation by increasing automaticity and regular habit performance.

Rewards from the following modalities will be tested: visual, auditory and visual-auditory. They will be delivered using a prototype chatbot to different groups of participants. They will be compared against two measurements: habit automaticity and habit performance. Previous studies [11, 62] demonstrate the strength of a habit is based on how automatic the action is and how regular the habit is performed [2, 19]. These two measurements are:

- *H1M1: The effect of positive reinforcement rewards on habit performance*
- *H1M2: The effect of positive reinforcement rewards on habit automaticity*

Hypothesis 2: multiple modalities rewards are more effective then singular mode rewards.

Multiple modalities have been shown to improve task performance in several studies [30, 38–40]. This hypothesis will test how each modality impacts habit performance and habit automaticity.

- *H2M1: The effect of multiple modalities versus singular on habit performance*
- *H2M2: The effect of multiple modalities versus singular on habit automaticity*

The measurements are habit performance (measurement 1, M1), calculated by the number of habits a participant marked as completed, incremented when a participant presses '*completed habit*' on the bot and habit automaticity (measurement 2, M2), calculated from two Self-Report Behavioural Automaticity Index (SRBAI) [11] questionnaires (Figure 22). In addition to these two hypotheses, the success of the prototype will also be evaluated with user interviews to understand what people's attitudes are towards the chatbot (RQ2) and how effective they are for supporting habit formation (RQ3).

1. "[Habit] after [context] is something I do automatically."
2. "[Habit] after [context] is something I do without having to consciously remember."
3. "[Habit] after [context] is something I do without thinking."
4. "[Habit] after [context] is something I start doing before I realise I am doing it."

Figure 22: SRBAI questionnaire [11] to measure habit automaticity, presented at the end of week 3 and week 4 during the study for Harry's Habits.

Figure 23: The landing page (`www.harrymt.com/harryshabits`) participants would see before they are taken to the Messenger bot.

5.1.2 Method

The 4-week evaluation trial was split into two sections. First a 3-week trial tested the success of the chatbot by evaluating the tool and the effectiveness of each modality on participants habit automaticity using the SRBAI questionnaire [11]. During the fourth week, chatbot interaction was removed during a 1-week follow up trial to test if participants continue with the habit. Participants were split into four groups (Table 1), all groups receiving reminders, three groups receiving rewards each from a different modality and one group (control group) did not receive any rewards.

5.1.3 Participants

Sixty participants were recruited with public posts to social networks and were mostly University students and staff. The process participants undertook is documented by the recruitment adverts (Appendix 2), the landing page (Figure 23) and the setup screens (Section 3.1). Participants were instructed to connect with the bot via Facebook Messenger and pick a series of options to set-up their habit tracking. Participants answered general demographic information as set up questions from the bot when they connected with it (Section 3.1), this was seen as their consent to participate in the study. These questions were:

- "What is your gender?"
- "How old are you?"
- "Have you used habit tracking systems before? If did they work and what were they?"
- "Would you like to be interviewed about your experience after the study?"

5.1.4 Study Design

After participants answered the demographic information, they were asked to pick a series of options to set-up their habit tracking. A brief description accompanied these options to guide the participants on how to answer. The habits participants could choose were split into two categories, physical and relaxation. They were: stretching, press ups, the plank, reading, writing or meditation. Participants were aware of these habits before they consented to the study and if they did not want to form any of these habits they would not continue with the study. These simple actions were chosen to match the length of the study, as simple tasks become automatic quicker than complex actions [20], for example a drinking water habit only takes, on average, 18 days of repetitive use.

The study used four conditions (Table 1): visual rewards, auditory rewards, visual-auditory rewards and a control without a reward. Participants were randomly assigned a condition by the bot after they completed the set-up, 15 participants were assigned to each condition. The rewards were aimed at motivating participants to keep coming back every day and completing their habits. Each GIF and audio track were selected based on popular combinations of videos and soundtracks (Figure 24). A search for "motivation" on the website `https://www.reddit.com/r/GifSound/` to find suitable pairings of video and audio. The video was converted into a GIF with the speed adjusted to match the speed of the audio. The relationship between the audio and the visual is inferred, therefore this mapping is *semi-congruent*. 5 GIFs and 5 matching audio tracks were selected as the rewards. Visual rewards had only the GIFs, auditory had only the audio and visual-auditory had them both at the same time. They can be found in the GitHub repo of the chatbot (`https://github.com/harrymt/harryshabits/tree/master/public`).

Condition	*Modality*	*Reward*	*No. Participants*
V	visual	15s GIF	15
A	auditory	15s audio	15
V-A	visual-auditory	15s GIF and audio	15
N	none (control group)	confirmation message	15

Table 1: Study conditions and their corresponding types of positive reinforcement that were randomly assigned to participants.

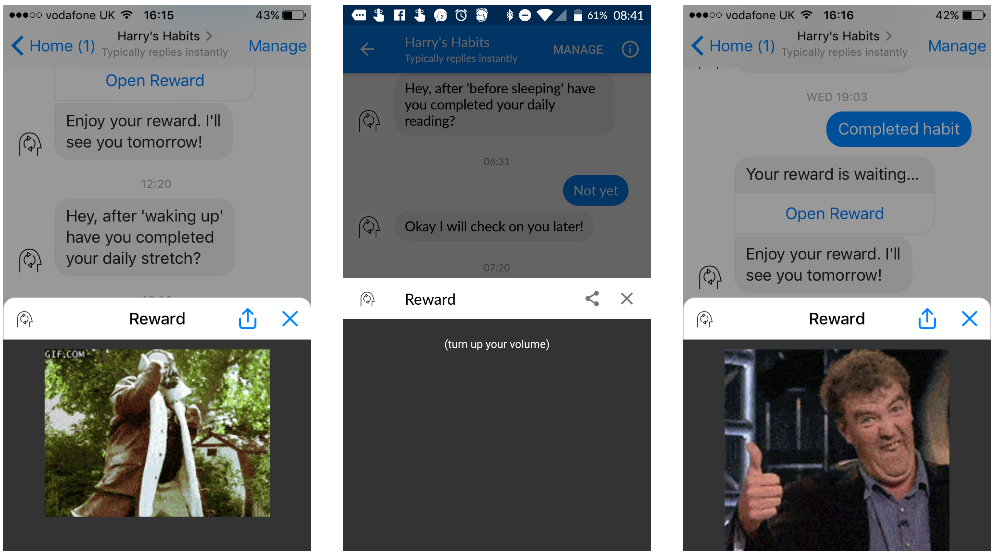

Figure 24: Example rewards for visual (left), auditory (middle) and visual-auditory (right) rewards for Harry's Habits.

- *visual rewards*: participants would receive a message that when tapped, would reveal a GIF.
- *auditory rewards*: participants would receive a message that when tapped, would play a song.
- *visual-auditory rewards*: participants would receive a message that when tapped, would play a song and reveal a GIF.
- *no rewards (control group)*: participants would only receive a confirmation message.

Participants would receive a confirmation message, followed by a positive reinforcement reward after they marked a habit as complete, unless they were in the control group, then they would only receive a confirmation message. Information about how well a participant was performing was not revealed, to separate the rewards from other types of motivation, as streaks can provide motivation [9].

5.1.5 Materials

The bot collected the amount of habits participants completed (M1) and their reward type. Habit automaticity (M2) was verified by using the Self-Report Behavioural Automaticity Index (SRBAI) [11] (Figure 22) — A validated set of questions to measure habit automaticity levels. This was used to test if participants were forming a habit. Additional questions were also asked about how participants found their rewards and how they found interacting with the chatbot. Finally, interviews with participants were conducted for further verification to discuss how they found the bot-delivered rewards.

5.1.6 Procedure

First, an internal pilot trial with three participants was conducted to develop the strength of the prototype. Minor language changes were made as a result of this, to better explain how participants should proceed. Second, a 4-week situated study was conducted with real people who wanted to try and form a new positive habit. The length of the study was appropriate for the simplicity of each habit and was based on the lengths of a previous habit formation study [12]. Finally, follow-up interviews with participants revealed if participants continued with their habit without the bot.

At the beginning of the 3-week period, participants gave their consent to take part in the study as required by the ethics committee (reference id: 54701) and were asked to answer basic demographic information and choose a new habit they would like to develop from a list of habits of similar difficulty, divided into two categories: physical and relaxation. Then they were asked to state an existing routine they could build their new habit around, and choose a time that routine normally occurred (morning, afternoon, evening). After they had answered these questions, they would be randomly assigned a modality (unknown to users) for rewards. Participants would complete their chosen habit every day after their existing routine, then wait for the bot to send them a message asking them for one of two choices.

Option one (completed habit): participants would receive a message thanking them, then participants not in the control group would receive a message linking them to a reward. This reward would be from the modality auto-assigned to that participant during the set-up phase (Table 1). Option two (not yet): the bot would check on the participant an hour later. This allowed for the checks to be snoozed, to ensure the new habit fit in with participants routine. If participants constantly told the bot they had not completed their habit yet, the bot would ask participants if they would like to change the time their routine occurred (Figure 25). This allowed the participants in the beginning to refine the time they would perform their habit.

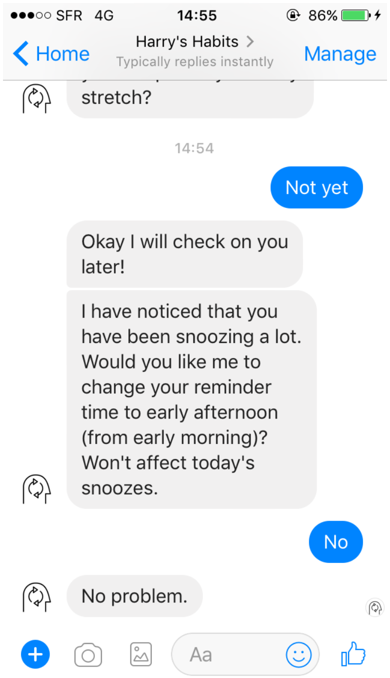

Figure 25: Participants would receive a message asking them if they would like to change their post completion notification time.

After 3 weeks of participants interacting with the bot, all participants who completed the set up were asked to complete the SRBAI questionnaire. Then the bot interaction was suspended for 1 week. After the full 4-week period, participants were asked to complete the questionnaire again. The SRBAI presents the questions on a 5-point Likert scale with answers from 'Strongly Disagree' (5) to 'Strongly Agree' (1). Higher scores indicates higher self-reported levels of automaticity. In addition, participants had an option to opt-in for an interview about their experience after the 4-week period.

5.2 Results

Sixty participants connected with the bot by pressing *'Get Started'* in Facebook Messenger on the following platforms: 25 participants used a web browser, 18 used iOS and 12 used Android. 14 participants, or 23% dropped out of the study at various stages (Figure 26): 54 participants, or 90% continued interacting with the bot and started the set-up. 39 participants, or 65% completed the set-up and out of these 39 participants, 3 participants ignored all messages from the bot during the trial. Leaving 36 participants, or 66% that are considered active throughout and are included in the final analysis.

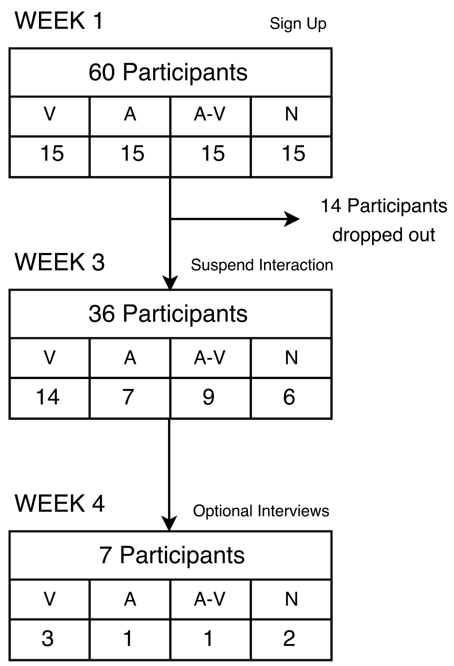

Figure 26: Participant drop-out and participation in optional interviews.

The thirty-six participants who completed the study are 18-63 years old, (mean: 27 years old, SD: 12), 23 or, 64% male, 11 or, 30% female and 2 or, 6% didn't say. These 36 remaining participants are split into the following modalities: 14 participants or, 93% visual, 9 participants or, 60% visual-auditory, 7 participants or, 46% auditory and 6 participants or, 40% had no rewards (control group). 36 participants sent a total of 1.1k messages to the bot (mean = 65 messages per participant) and the bot sent 2.7k total messages back. 184 total habits were marked as completed, with the bot issuing 69 visual rewards, 58 visual-auditory rewards and 17 auditory rewards to those participants. The control group completed 40 habits.

Habit Performance

Comparing the number of participants who dropped out of the study versus their reward modality, shows that 7 participants who dropped out of the study had 27 visual rewards in total (mean = 3.85). 2 participants had 4 total auditory rewards (mean = 2) and 2 participants had 2 visual-auditory combined (mean = 6.5). These visual-auditory participants that dropped out had the highest amount of snoozes (24 total snoozes, 6 and 18 individually, mean = 12), compared with visual 10 total (mean = 2), and auditory with 0 snoozes.

7 participants, or 19% previously used habit tracking systems and 36 participants, or 100% stated they worked well for tracking habits. These habits were: 'Diet' (3 mentions), 'Exercise' (3 mentions), 'Deadlines' (2 mentions), 'Audiobook reading' (1 mention), 'Weight' (1 mention).

All of these participants chose new habits that they had not tracked before. Meditation was the most popular habit chosen 12 participants, or 33%, followed by Press ups 8 participants, or 22%, then Stretching 6 participants, or 15%. Reading and writing were the least, only selected by 4 and 2 participants respectively. Stretching 6 participants, or 15% was the most completed habit based on selection, 60 times, ranking 10.0. Meditation ranked 6.25 and the least were the plank and reading with 3.75 and 1.75 respectively.

Participants with visual rewards, 14 participants, or 39% had the highest total number of snoozes (72 total presses to 'Not Yet'), auditory had the smallest (14 total). The control group had 55 snoozes and visual-auditory had 45 snoozes. Most participants snoozed (answered 'Not Yet') in the morning (100 times), specifically mid (66) and late (27) morning. Visual-auditory had the most number of failed snoozes 10, split over 6 participants, 1 person had 5 failed snoozes, then it was auditory with 4.

Habit performance was also tracked in the form of a streak. If a participant did not track a habit for a day, their streak would be reset to 0. Meditation had the most cumulative streak (17) followed by stretching (7). High streaks (streak > 10) had habits: meditation (135 streaks) and stretch (126 streaks). Visual-auditory rewards had the most streaks (126), control group (75) and visual (60). Stretch had the highest peak streak (18), meditation (15). However, overall, for all completed habits, meditation was the most streaked (269), stretching (231), press ups (39) and plank (24).

Mostly participants chose existing routines that were suggested to them. For example, during the pilot trials when asking for an existing routine, users were confused, so examples of habit contexts were provided (Figure 27). The results found, 31, or 86% participants chose one of the contexts that were listed as examples, some with a slight change in before and after wording, e.g. *before* getting home from work, rather than *after* getting home from work. The remaining 5 participants chose the following context: *'Having a snack'*, *'Sitting in bed'*, *'Early morning'*, *'During breakfast'* and *'Before sleeping'*.

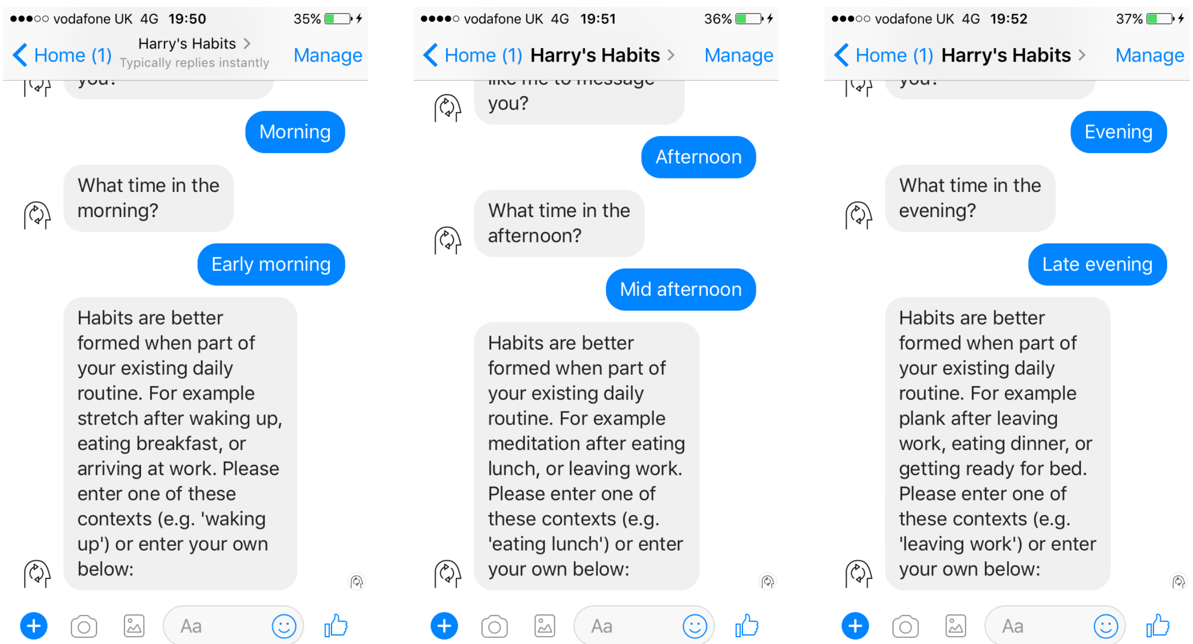

Figure 27: The examples given to participants for habit contexts based on the time they chose. An item from a list is randomly suggested to them.

H1M1: The effect of positive reinforcement rewards on habit performance

Analysis of each individual reward on habit performance (Figure 28) revealed that there was a statistically significant difference between each reward as determined by one-way ANOVA ($F(2, 624) = 27.007$, $p < 0.05$). A Tukey post hoc test revealed that the difference in completed habits was statistically significantly lower for the auditory (0.081 + 0.274, $p < 0.005$) and visual-auditory (0.267 + 0.484, $p = 0.025$) feedback compared to the visual feedback (0.373 + 0.484).

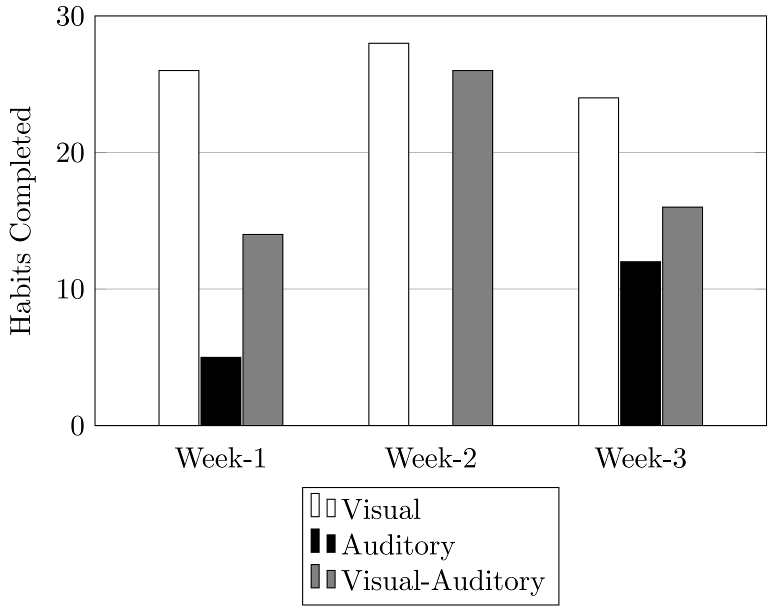

Figure 28: The number of habits completed for each reward over the 3-week period.

A one-way between-groups analysis of variance with planned comparisons was conducted to explore the effect of rewards on the number of habits completed, compared with the control group (Figure 29). Participants were divided into two groups: group 1 with rewards, group 2 without rewards. There was a statistically significant difference at the $p < 0.005$ level in both groups for 2/3 of the weeks: week 1, $F(1, 23.20) = 9.48$, $p = 0.005$, week 2, $F(1, 33.35) = 4.46$, p = 0.42 and week 3, $F(1, 50) = 17.01$, $p \leq 0.005$. The effect size for week 1, week 2 and week 3 are large, calculated using η^2, were 0.25, 0.39 and 0.43 respectively, this shows a large difference in the mean scores.

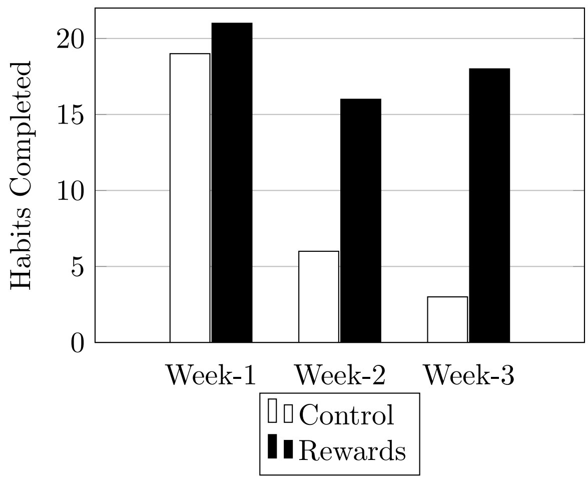

Figure 29: H1M1: The effect of positive reinforcement rewards on habit performance. The sum of habits completed by participants with rewards versus the control group during 3-week study period.

H2M1: The effect of multiple modalities versus singular on habit performance

A one-way between-groups analysis of variance with planned comparisons was conducted to explore the effect of multiple modalities and singular modalities on the number of habits completed (Figure 30). Participants were divided into two groups according to their mode: group 1: visual rewards, auditory rewards, group 2: visual-auditory rewards. There was a statistically significant difference at the $p < 0.005$ level in week 2 and week 3, and lots in all groups: week 1, $F(1, 50) = 0.69$, $p = 0.410$, week 2, $F(1, 50) = 23.04$, $p < 0.005$ and week 3, $F(1, 50) = 8.85$, $p \leq 0.005$. The effect size for week 1, week 2 and week 3 are large, calculated using η^2, were 0.25, 0.39 and 0.43 respectively, this shows a large difference in the mean scores.

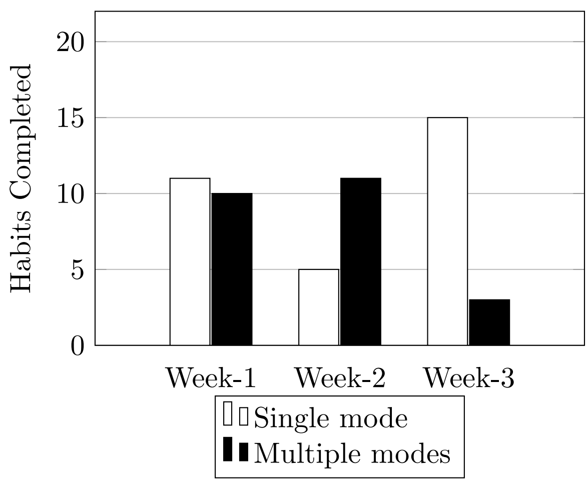

Figure 30: H2M1: The effect of multiple modalities versus singular on habit performance Sum of completed habits for multiple modalities compared with singular modes.

Habit Automaticity

11 participants completed both SRBAI questionnaires, 2 control group, 2 auditory, 5 visual and 2 visual-auditory. A paired-samples t-test was conducted to evaluate the change of habit automaticity between the first SRBAI questionnaire (after week 3) and the second (after week 4). There was a statistically significant increase in automaticity scores from SRBAI 1 (mean = 14.18, SD = 3.78) to SRBAI 2 (mean = 15.09, SD = 4.34), $t(10) = 2.469$, $p < 0.005$ (two-tailed). The mean increase in SRBAI scores was 0.90 with a 95% confidence interval ranging from 0.08 to 1.72. The $\eta^2 = 0.37$ indicates a large effect size.

H1M2: The effect of positive reinforcement rewards on habit automaticity

An independent-samples t-test was conducted to compare the habit automaticity scores for rewards and control at both SRBAI 1 and SRBAI 2 (Figure 31). For SRBAI 1, there was also no significant differences in scores for rewards (mean = 14.33, SD = 3.84) and control (mean = 13.50, SD = 4.94; $t(9) = 0.224$, $p = 0.85$, two-tailed). The magnitude of the differences in the means (mean difference = 0.83, 95% CI: 29.43 to 27.76) was very small ($\eta^2 = 0.005$). For SRBAI 2, there was no significant difference in scores for rewards (mean = 15.22, SD = 4.29) and control (mean = 14.50, SD = 6.36; $t(9) = 0.202$, $p = 0.84$, two-tailed). The magnitude of the differences in the means (mean difference = 0.72, 95% CI: 8.80 to 7.36) was very small ($\eta^2 = 0.004$).

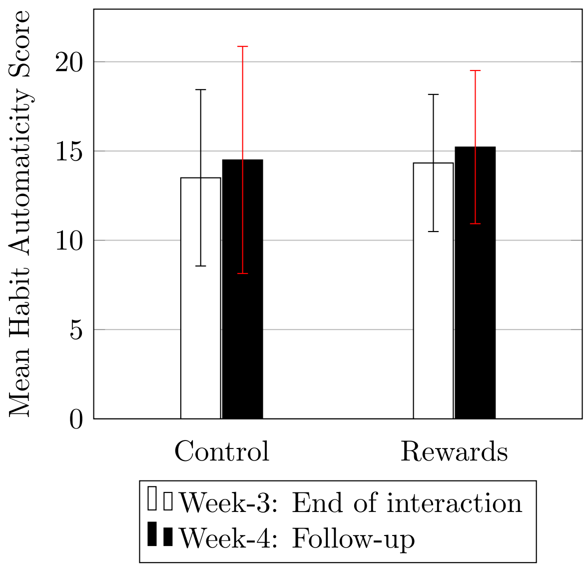

Figure 31: H1M2: The effect of positive reinforcement rewards on habit automaticity Comparing mean habit automaticity for rewards versus control group.

A one-way between-groups analysis of variance with planned comparisons was also conducted to explore the impact of rewards on habit automaticity, as measured by the SRBAI 1 and 2. Participants were divided into two groups: group 1 with rewards, group 2 without rewards. There was not a statistically significant difference for the two groups at SRBAI 1: $F(1, 9) = 0.02$, $p = 0.88$, and SRBAI 2: $F(1, 9) = 0.07$, $p = 0.78$. In addition, the difference in mean scores between the groups had, at SRBAI 1: a medium effect with an effect size $\eta^2 = 0.11$, and at SRBAI 2: a large effect, with an effect size of $\eta^2 = 0.17$.

H2M2: The effect of multiple modalities versus singular on habit automaticity

A one-way between-groups analysis of variance with planned comparisons was conducted to explore the impact of multiple modalities on habit automaticity, compared with singular modes as measured by the SRBAI 1 and 2 (Figure 32). Participants were divided into two groups according to their reward mode: group 1: visual rewards, auditory rewards and group 2: visual-auditory combined rewards. There was not a statistically significant difference for the two groups at SRBAI 1:$F(1,9) = 1.04$, $p = 0.33$, and SRBAI 2: $F(1,9) = 0.64$, $p = 0.44$. In addition, the difference in mean scores between the groups had, at SRBAI 1: a medium effect with an effect size of $\eta^2 = 0.11$, and at SRBAI 2: a large effect, with an effect size of $\eta^2 = 0.17$.

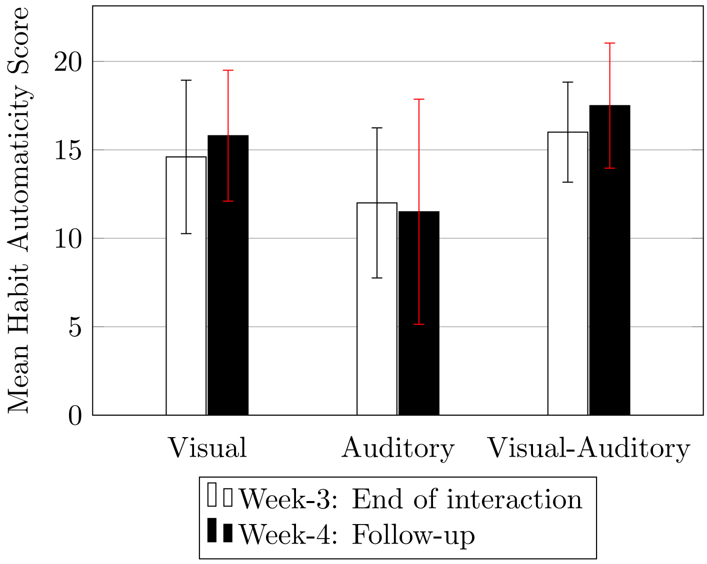

Figure 32: H2M2: The effect of multiple modalities versus singular on habit automaticity Comparing mean habit automaticity for each group.

Interview Transcripts

Seven interviews with participants were conducted to gather their experience with their habit performance after the prototype bot was removed. The breakdown of participants are detailed in Figure 33 and full interview transcripts are in Appendix 1. In addition to the above interview transcripts, one participant left feedback directly with the bot, and two people left public reviews about the bot.

> Participant with visual-auditory rewards and the meditation habit: *I felt the questions [SRBAI] afterwards about meditation being something I do without thinking were just repeating the same question with a different phrasing. So I don't know how useful they are. The chatbot made me aware I wasn't keeping up with something and I feel it could have been a little more adaptive to try and remind me in different ways or at different times. I would say I felt more like it was a reminder that I wasn't doing well at forming the habit.*

> Review 1: (Rated 2 out of 5) *'I felt a bit like I was being nagged, and I don't like being nagged. The rewards were a bit odd. Would be nice to actually set a time to be asked about the habit.'*

Review 2: (Rated 4 out of 5) *'Interesting project but I did not ultimately feel motivated, the reminders became more of a chore than something I wanted to do.'*

Participant	**Reward**	**Habit**
1	Visual	Meditation
2	Auditory	Stretch
3	Visual	Writing
4	No reward	Meditation
5	Visual	Meditation
6	No reward	Meditation
7	Visual-Auditory	Stretch

Figure 33: Participant interview breakdown. Full transcripts in Appendix 1.

5.3 Discussion

This research aimed to understand more about positive reinforcement rewards from different modalities and their role in habit formation. The results report that participants receiving bot-delivered rewards completed more habits than the control group without rewards. Participants with visual-auditory rewards had the highest habit automaticity score and during participant interviews (N = 7) and all participants reported a drop in habit performance after 1-week without the prototype.

5.3.1 Interview Analysis

The interview transcripts (Appendix 1) were analysed and participants were anonymised following a thematic approach [63]. Below, we consider the role of each participants during the process of interacting with the chatbot and reflect how their interaction may of had a meaningful impact to their formation of a new habit.

Habit Repetition

Participants discussed how they picked their habit, they chose because they had wanted to start for the particular habit for a long time, it was *'not too much effort'* and *'something successful people do'*. They wanted *'to be more active'*, *'relieve stress'* and wanted a habit that was *'less time consuming'*. Throughout, participants mostly completed their habits, but, one participant always put the message off and eventually their performance got *'worse and worse'* until they stopped all together. This is confirmed after the bot was removed, with all interviewed participants (N = 7) finding it difficult to continue with their habit. They *'kept forgetting'*, found it *'harder to remember'* and lacked motivation, not performing the action if it had *'been a long day'*. Some tried to do it *'every now and again'*, but usually they would only complete it if *'they remembered'*. This reveals the dependency between technology and habits, suggesting that the bot did not increase habit automaticity, or that the existing routine participants chose was not suitable for new habits, or that they were not given enough time to develop automaticity.

Attitudes Towards Rewards

Participants had mixed feelings about the rewards. Some *'did not like the [visual] rewards'*, skipping over them after the first few, they *'just wanted to get rid of the notification dot'*. Another participant said *'some of them [visual-auditory rewards] were funny'*, but they did not like them overall and mentioned the auditory rewards were *'too random'*. One participant thought they did not give them an incentive towards their habit, just a *'nice little extra'*. They also discussed including time-sensitive rewards, as they did not want to listen to music before going to bed. This shows the importance of using an appropriate modality at particular times, e.g. not having auditory rewards at certain times of the day. Finally, an upbeat participant talked about *'always wanting to open them'* and *'the combination was perfect'*. However, they said they also found them *'repetitive'* and one participant stated that they simply wanted to get rid of the notification. This implies that the bot was not tracking how many habits that participant completed, but instead tracking how many times that participant wanted the bot to stop messaging them.

Attitudes Towards the Chatbot

Participants were asked about how they found the chatbot as the method of interaction. They found the method *'pretty good'*, they *'liked it'* and *'would have liked more interaction'*. Suggesting additional features, such as *'help and support throughout'*, *'ideas on how to improve your habit'* and *'advice on how to set aside time for your habit'*. Others were neutral, some expecting *'different messages, such as Hey [name], a bit more care about the person, a bit less like a robot'*. Lots of participants (N = 4) enjoyed the reminder aspect, but a few found it *'repetitive'* and *'got annoying if I pressed Not Yet'*. Participants wanted to see their progress as they tracked their habits, they talked about wanting to reflect on their data. They mentioned that they would feel *'more encouraged to keep doing it, rather than random music [auditory rewards]'*. Mostly participants wanted the prototype to come back with a few modifications: *'enclosed with Fitbit so it is all in a single place'*, *'fine without rewards'* (2 participants mentioned this), *'more interaction'* and *'with statistics about my progress'*. Participants wanted the bot as more of a *'constant persistent reminder'* with additional tracking elements to remind them to perform their habit to fit into their busy schedule. Participants (N = 5) mentioned the *headspace* app (`www.headspace.com`), mentioning that they wanted a combination of the bot and headspace. It prompted another participant to download the headspace app. They wanted the bot to keep on track of their habit and they would use the headspace app to help them perform their mindfulness.

5.3.2 Habit Performance

The results found that participants are more likely to complete their habit if given one of the rewards and participants completed more habits with visual or auditory rewards than visual-auditory rewards.

M1H1: The Rewards effect on Habit Performance

There was a statistically significant drop in habit performance for the control group without rewards. This reveals the effect the rewards had on participant interaction with the bot and habit performance. Although this is contradicted from the interviews where participants discussed negative feelings towards rewards later in the 3-week period. The findings show that rewards did improve habit performance.

M1H2: Multiple Modalities effect on Habit Performance

There is a significant difference between the visual-auditory modes on the number of completed habits. But the result appears to be different to the initial hypotheses, with more participants completing habits with singular mode rewards than multiple. This contradicts the belief that multiple modalities benefits task completion, however these results only impact the particular rewards delivered by this bot. Therefore, it is inconclusive whether multiple modalities rewards in the general sense impact habit performance.

5.3.3 Habit Automaticity

It is inconclusive whether the rewards or the combined modalities affected habit automaticity.

M2H1: The Rewards effect on Habit Automaticity

Each individual reward did not have a significant effect on habit automaticity. Although participants with rewards had slightly higher automaticity scores, follow up interviews suggest that automaticity did not develop.

M2H2: Multiple Modalities effect on Habit Automaticity

Participants with visual-auditory combined reward had higher habit automaticity scores compared with visual or auditory rewards. However, these were not statistically significant.

5.3.4 Chatbot Success

The chatbot was partially successful at running a research trial. There were several issues and various limitations with development. However, it was generally liked by participants and it was easy to gather a lot of useful data.

Participants had mixed feelings towards the bot. Their performance shows that the number of snoozes over time decreased, but the number of total habits completed per day also decreased for all reward types (including the control group). However, participant streaks over time increased and 36 participants manage to use it for 3-weeks.

Participants had various issues with bot interaction. Seven participants tried to message the bot during setup, instead of using the built in *quick reply* buttons. This broke the setup flow and they had to start again. Other participants tried to send multiple messages when asked for free input, they went around an endless loop when asking for a habit type and participants tried to mark their habit as completed using the Facebook Messenger thumb emotion (which the bot was not coded for). A participant also pointed out that they could change their existing routine time but were unable to change the description.

Participants gave additional feedback by simply messaging the bot. They asked inquisitive questions, such as *'what kind of thing are you looking to find out'*, *'this is not working for me'* and *'stop'* — to try and stop the daily messages (the participant then blocked the bot). Negative feedback towards the rewards and bot were also expressed. When asked about being messaged every day, a participant sent this reply and then blocked the bot: *'Do not do that, it will be annoying'*, and another said *'never message me'*. Another stated that this was the *'lame same band'* after receiving an auditory reward.

Participants were able to personalise the chatbot with several different variable configurations. This personalisation could have effected the findings by not narrowing the condition variables. Replicating this research with tighter a configuration is needed to validate this line of reasoning. Finally, streaks could have been better used to give insight to participants progress and challenged them to maintain it, using loss aversion [64] to compare the impact of their broken streak with the gain of keeping it. However, all participants interviewed (N = 7) struggled with maintaining habit performance after the bot was removed. This suggests a dependence between the technology and the habit as participants depended on bot notifications to continue repeating the desired action.

Summary

The chatbot rewards from different modalities effected habit formation (RQ1) with the singular modes improving habit performance, but not habit automaticity. Feedback towards interaction with the chatbot was mostly positive (RQ2), but all interviewed participants suggested features they would like, likely because it did not fulfil their needs. The rewards delivered by the chatbot got a mixed reception from participants and did not significantly increase habit automaticity, but they did improve habit performance. Therefore, the chatbot is partially effective for supporting habit formation (RQ3).

6 Limitations and Future Work

There were several limitations to these findings. Firstly, the evaluation of the second hypotheses relied on participants self-reporting. Participants could have simply lied to remove the alert and get the reward. This is particularly true with the snooze function, as some participants found it annoying and stopped using the chatbot over time. Participants may have been getting rid of the reminder rather than completing the habit, therefore the measurement could have been how participants reacted to notifications instead of if they competed their habit. Therefore, it is difficult to draw any valid conclusion on actual habit performance. Future work into how the platform participants used effected how quickly they responded to the alerts, e.g. browser, iOS or Android. Secondly, only seven participants responded to the follow up interviews for the evaluation of the first hypotheses although this might have been caused by the fact that these interviews were optional and not an integral part of the study. In addition, the study relied on participants recall, which could be inaccurate. Therefore, a larger sample size for the SRBAI questionnaire is needed to validate the hypotheses and the findings. Thirdly, additions to the chatbot, such as using natural language processing, could improve the functionality of the chatbot, proving a more conversational user interface and perhaps could increase participant interaction and perhaps not create a dependence between our prototype and participants habit performance during the 3-week period as this was disadvantage for habit formation. Fourthly, the small number of participants and the small sample of rewards used in the study make it unclear how these findings would generalise to other types of rewards with the same modality. Finally, these findings only apply to the intrinsic positive reinforcement rewards used in the study, as the content and method of delivery is another variable that effects these results. Additional studies using different types of bot-delivered rewards from different modalities would validate these findings.

7 Conclusion

We have surveyed three areas of literature: habit formation rewards, different modes of feedback and chatbot interaction. We found that habit performance and habit automaticity are keys to building lasting behaviour change. This thesis builds these three areas into a set of design requirements for building reward-based chatbots that support habit formation. A prototype is designed and constructed from these requirements, that aims to track habits and deliver rewards from three modalities: visual, auditory and visual-auditory combined. We evaluated our prototype during a 4-week study against two hypotheses, comparing these against a control group and each reward type: i) positive reinforcement is effective at supporting habit formation by increasing automaticity and regular habit performance, ii) multiple modalities rewards are more effective then singular mode rewards. Thirty-six participants completed the study interacting with the bot for three weeks to try and form a new positive habit, then for a further 1-week interaction with the prototype was suspended. Validated habit automaticity questionnaires and seven participant interviews were performed for further validation. This allowed us to evaluate how habit performance and habit automaticity was effected without technology. The results show how rewards delivered by Harry's Habits affected habit performance and habit automaticity. More specifically, the results found that the bot-delivered rewards improved habit performance; thus participants were more likely to complete their habit if given the reward. Habit performance was also affected by different modalities, although not in the way our hypotheses assumed, as singular modalities had higher habit performance than visual-auditory rewards. Finally, the findings of the study show there is no statistical significance whether the rewards or the combined modalities effected habit automaticity. More conclusive evidence is needed to show that rewards from combined modalities effect habit automaticity. Nevertheless, although there were limitations to this project, there is still a significant contribution with comparisons of visual, auditory and visual-auditory positive reinforcement rewards and we hope this may open up new research avenues for investigating the use of chatbots as tools to help form new habits.

References

[1] P. Klasnja, S. Consolvo, and W. Pratt, "How to evaluate technologies for health behavior change in hci research," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11, Vancouver, BC, Canada: ACM, 2011, pp. 3063–3072, ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1979396.

[2] P. Lally and B. Gardner, "Promoting habit formation," *Health Psychology Review*, vol. 7, no. sup1, S137–S158, 2013. DOI: 10.1080/17437199.2011.603640. eprint: http://dx.doi.org/10.1080/17437199.2011.603640.

[3] W. Wood and D. T. Neal, "The habitual consumer," *Journal of Consumer Psychology*, vol. 19, no. 4, pp. 579–592, 2009, ISSN: 1057-7408. DOI: 10.1016/j.jcps.2009.08.003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1057740809001065.

[4] B. Gardner, K. Sheals, J. Wardle, and L. McGowan, "Putting habit into practice, and practice into habit: A process evaluation and exploration of the acceptability of a habit-based dietary behaviour change intervention," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 11, no. 1, p. 135, 2014, ISSN: 1479-5868. DOI: 10.1186/s12966-014-0135-7. [Online]. Available: https://doi.org/10.1186/s12966-014-0135-7.

[5] G. Loas, D. Fremaux, and M. P. Marchand, "The 20-item toronto alexithymia scale: Structural validity, internal consistency and prevalence of alexithymia in a swiss adolescent sample," *L'Encéphale: Revue de psychiatrie clinique biologique et thérapeutique*, vol. 21, no. 2, pp. 117–122, 1995, ISSN: 0013-7006(Print).

[6] R. A. de Vries, K. P. Truong, S. Kwint, C. H. Drossaert, and V. Evers, "Crowd-designed motivation: Motivational messages for exercise adherence based on behavior change theory," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16, Santa Clara, California, USA: ACM, 2016, pp. 297–308, ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858229. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858229.

[7] R. J. Vallerand and G. Reid, "On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory," *Journal of Sport Psychology*, vol. 6, no. 1, pp. 94–102, 1984. DOI: 10.1123/jsp.6.1.94. eprint: https://doi.org/10.1123/jsp.6.1.94. [Online]. Available: https://doi.org/10.1123/jsp.6.1.94.

[8] C. Danis, K. Vogt, J. D. Weisz, Y. Ma, and R. Olsen, "Factors that help and hinder a daily weighing and reporting behavior," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '17, Denver, Colorado, USA: ACM, 2017, pp. 2480–2486, ISBN: 978-1-4503-4656-6. DOI: 10.1145/3027063.3053206. [Online]. Available: http://doi.acm.org/10.1145/3027063.3053206.

[9] I. Renfree, D. Harrison, P. Marshall, K. Stawarz, and A. Cox, "Don't kick the habit: The role of dependency in habit formation apps," CHI EA '16, pp. 2932–2939, 2016. DOI: 10.1145/2851581.2892495.

[10] M. L. Lee and A. K. Dey, "Real-time feedback for improving medication taking," in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing*

Systems, ser. CHI '14, Toronto, Ontario, Canada: ACM, 2014, pp. 2259–2268, ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557210.

[11] B. Gardner, C. Abraham, P. Lally, and G.-J. de Bruijn, "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*, vol. 9, no. 1, p. 102, 2012, ISSN: 1479-5868. DOI: 10.1186/1479-5868-9-102.

[12] K. Stawarz, A. L. Cox, and A. Blandford, "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*, ser. CHI '15, Seoul, Republic of Korea: ACM, 2015, pp. 2653–2662, ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702230.

[13] A. Bandura, "Self-efficacy: toward a unifying theory of behavioral change," *Psychol Rev*, vol. 84, no. 2, pp. 191–215, 1977.

[14] E. L. Deci, R. Koestner, and R. M. Ryan, "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation.," *Psychological Bulletin*, vol. 125, no. 6, 1999. DOI: 10.1037/0033-2909.125.6.627.

[15] M. R. McGee-Lennon, M. K. Wolters, and S. Brewster, "User-centred multimodal reminders for assistive living," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11, Vancouver, BC, Canada: ACM, 2011, pp. 2105–2114, ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1979248.

[16] F. Karray, M. Alemzadeh, J A Saleh, and M. N. Arab, "Human-computer interaction: Overview on state of the art," vol. 1, pp. 137–159, 2008.

[17] K. van Turnhout, A. Bennis, S. Craenmehr, R. Holwerda, M. Jacobs, R. Niels, L. Zaad, S. Hoppenbrouwers, D. Lenior, and R. Bakker, "Design patterns for mixed-method research in hci," in *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, ser. NordiCHI '14, Helsinki, Finland: ACM, 2014, pp. 361–370, ISBN: 978-1-4503-2542-4. DOI: 10.1145/2639189.2639220. [Online]. Available: http://doi.acm.org/10.1145/2639189.2639220.

[18] J. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* SAGE Publications, 2013, ISBN: 9781452226101. [Online]. Available: https://books.google.co.uk/books?id=4uB76IC_pOQC.

[19] P. Lally, J. Wardle, and B. Gardner, "Experiences of habit formation: A qualitative study," *Psychology, Health & Medicine*, vol. 16, no. 4, pp. 484–489, 2011, PMID: 21749245. DOI: 10.1080/13548506.2011.555774. eprint: http://dx.doi.org/10.1080/13548506.2011.555774.

[20] P. Lally, C. H. M. van Jaarsveld, H. W. W. Potts, and J. Wardle, "How are habits formed: Modelling habit formation in the real world," *European Journal of Social Psychology*, vol. 40, no. 6, pp. 998–1009, 2010, ISSN: 1099-0992. DOI: 10.1002/ejsp.674.

[21] B Verplanken, "Habits and implementation intentions," in *The ABC of behavioural change.* Oxford, UK: Elsevier, 2005, pp. 99–109. [Online]. Available: http://opus.bath.ac.uk/9438/.

[22] P. M. Gollwitzer, "Implementation intentions: Strong effects of simple plans.," *American Psychologist*, vol. 54, no. 7, p. 493, 1999.

[23] K. Stawarz, M. D. Rodreguez, A. L. Cox, and A. Blandford, "Understanding the use of contextual cues: Design implications for medication adherence technologies that support remembering," *Digital Health*, vol. 2, p. 2 055 207 616 678 707, 2016. DOI: 10.1177/2055207616678707. eprint: http://dx.doi.org/10.1177/2055207616678707.

[24] M. A. Adriaanse, P. M. Gollwitzer, D. T. De Ridder, J. B. De Wit, and F. M. Kroese, *Breaking habits with implementation intentions: A test of underlying processes*, 2011.

[25] M. A. McDaniel and G. O. Einstein, "Strategic and automatic processes in prospective memory retrieval: A multiprocess framework," *Applied cognitive psychology*, vol. 14, no. 7, 2000.

[26] L. Wang, F. Li, D. Wang, K. Xie, D. Wang, X. Shen, and J. Tsien, "Receptors in dopaminergic neurons are crucial for habit learning," *Neuron*, vol. 72, no. 6, pp. 1055 –1066, 2011, ISSN: 0896-6273. DOI: 10.1016/j.neuron.2011.10.019.

[27] K. W. Wendy B Michael S, *Why are gambling machines addictive*, 2016. [Online]. Available: http://www.bbc.co.uk/programmes/b07w11kg.

[28] P. Weiser, D. Bucher, F. Cellina, and V. De Luca, "A Taxonomy of Motivational Affordances for Meaningful Gamified and Persuasive Technologies," *3rd International Conference on ICT for Sustainability (ICT4S)*, Advances in Computer Science Research, vol. 22, pp. 271–280, 2015. DOI: 10.2991/ict4s-env-15.2015.31.

[29] F. Chen, *Designing Human Interface in Speech Technology.* Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005, ISBN: 0387241558.

[30] O. Metatla, N. N. Correia, F. Martin, N. Bryan-Kinns, and T. Stockman, "Tap the shapetones: Exploring the effects of crossmodal congruence in an audio-visual interface," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16, Santa Clara, California, USA: ACM, 2016, pp. 1055–1066, ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858456. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858456.

[31] A. Stouffs, "Interruptions as multimodal outputs: Which are the less disruptive," in *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, ser. ICMI '02, Washington, DC, USA: IEEE Computer Society, 2002, pp. 479–, ISBN: 0-7695-1834-6. DOI: 10.1109/ICMI.2002.1167043. [Online]. Available: http://dx.doi.org/10.1109/ICMI.2002.1167043.

[32] J. R. Williamson, M. McGee-Lennon, and S. Brewster, "Designing multimodal reminders for the home: Pairing content with presentation," ser. ICMI '12, Santa Monica, California, USA: ACM, 2012, pp. 445–448, ISBN: 978-1-4503-1467-1. DOI: 10.1145/2388676.2388774.

[33] D. P. Brumby, S. C. Davies, C. P. Janssen, and J. J. Grace, "Fast or safe?: How performance objectives determine modality output choices while interacting on the move," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11, Vancouver, BC, Canada: ACM, 2011, pp. 473–482, ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1979009. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979009.

[34] K. Stawarz and A. L. Cox, "Designing for health behavior change: Hci research alone is not enough.," *Crossing HCI and Health: Advancing Health and Wellness Technology Research in Home and Community Settings, CHI 2015 Workshop*, 2015.

[35] R. de Oliveira, M. Cherubini, and N. Oliver, "Movipill: Improving medication compliance for elders using a mobile persuasive social game," *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, UbiComp '10, pp. 251–260, 2010. DOI: 10.1145/1864349.1864371.

[36] E. Hoggan, A. Crossan, S. A. Brewster, and T. Kaaresoja, "Audio or tactile feedback: Which modality when?" In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '09, Boston, MA, USA: ACM, 2009, pp. 2253–2256, ISBN: 978-1-60558-246-7. DOI: 10.1145/1518701.1519045. [Online]. Available: http://doi.acm.org/10.1145/1518701.1519045.

[37] M. R. McGee-Lennon, "Reminders that make sense: Designing multisensory notifications for the home," *Journal of Assistive Technologies*, vol. 6, no. 2, pp. 93–104, 2012. DOI: 10.1108/17549451211234957. eprint: https://doi.org/10.1108/17549451211234957. [Online]. Available: https://doi.org/10.1108/17549451211234957.

[38] J. B. V. Erp and H. A. V. Veen, "Vibrotactile in-vehicle navigation system," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 7, no. 4, pp. 247 –256, 2004, ISSN: 1369-8478. DOI: http://dx.doi.org/10.1016/j.trf.2004.09.003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1369847804000385.

[39] C. D. Wickens, J. Goh, J. Helleberg, W. J. Horrey, and D. A. Talleur, "Attentional models of multitask pilot performance using advanced display technology," *Human Factors*, vol. 45, no. 3, pp. 360–380, 2003, PMID: 14702989. DOI: 10.1518/hfes.45.3.360.27250. eprint: http://dx.doi.org/10.1518/hfes.45.3.360.27250. [Online]. Available: http://dx.doi.org/10.1518/hfes.45.3.360.27250.

[40] J. L. Burke, M. S. Prewett, A. A. Gray, L. Yang, F. R. B. Stilson, M. D. Coovert, L. R. Elliot, and E. Redden, "Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis," in *Proceedings of the 8th International Conference on Multimodal Interfaces*, ser. ICMI '06, Banff, Alberta, Canada: ACM, 2006, pp. 108–117, ISBN: 1-59593-541-X. DOI: 10.1145/1180995.1181017. [Online]. Available: http://doi.acm.org/10.1145/1180995.1181017.

[41] D. W. Massaro, "Illusions and issues in bimodal speech perception," in *AVSP*, 1998.

[42] M. B.R. J. Lyons EJ Lewis ZH, "Behavior change techniques implemented in electronic lifestyle activity monitors: A systematic content analysis," *J Med Internet Res*, no. 16(8), 2014. DOI: 10.2196/jmir.3469. [Online]. Available: http://www.jmir.org/2014/8/e192.

[43] L. T Cowan, S. Van Wagenen, B. A Brown, R. Hedin, Y. Seino-Stephan, P Hall, and J. H West, "Apps of steel: Are exercise apps providing consumers with realistic expectations. a content analysis of exercise apps for presence of behavior change theory," vol. 40, Sep. 2012.

[44] H. J. West, C. P. Hall, L. C. Hanson, D. M. Barnes, C. Giraud-Carrier, and J. Barrett, "There's an app for that: Content analysis of paid health and fitness apps," *J Med Internet Res*, vol. 14, no. 3, 2012. DOI: 10.2196/jmir.1977. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22584372.

[45] L. Venta, M. Isomursu, A. Ahtinen, and S. Ramiah, ""my phone is a part of my soul” - how people bond with their mobile phones," UBICOMM '08, pp. 311–317, 2008. DOI: 10.1109/UBICOMM.2008.48.

[46] K. Stawarz, A. L. Cox, and A. Blandford, "Don't forget your pill!: Designing effective medication reminder apps that support users' daily routines," in *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '14, Toronto, Ontario, Canada: ACM, 2014, pp. 2269–2278, ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557079.

[47] B. Graf, M. Krüger, F. Müller, A. Ruhland, and A. Zech, "Nombot: Simplify food tracking," in *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '15, Linz, Austria: ACM, 2015, pp. 360–363, ISBN: 978-1-4503-3605-5. DOI: 10.1145/2836041.2841208. [Online]. Available: http://doi.acm.org/10.1145/2836041.2841208.

[48] A. Følstad and P. B. Brandtzæg, "Chatbots and the new world of hci," *interactions*, vol. 24, no. 4, pp. 38–42, 2017, ISSN: 1072-5520. DOI: 10.1145/3085558. [Online]. Available: http://doi.acm.org/10.1145/3085558.

[49] J. Weizenbaum, "Eliza - a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1966, ISSN: 0001-0782. DOI: 10.1145/365153.365168. [Online]. Available: http://doi.acm.org/10.1145/365153.365168.

[50] Facebook, "Quick replies," *Facebook for developers*, 2017. [Online]. Available: https://developers.facebook.com/docs/messenger-platform/send-api-reference/quick-replies.

[51] A. Smith, K. de Salas, B. Schüz, S. G. Ferguson, and I. Lewis, "Mhealth intervention design: Creating mhealth interventions for behaviour change," in *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, ser. OzCHI '16, Launceston, Tasmania, Australia: ACM, 2016, pp. 531–536, ISBN: 978-1-4503-4618-4. DOI: 10.1145/3010915.3010986.

[52] Statista, *Number of monthly active facebook messenger users from april 2014 to april 2017 (in millions)*. [Online]. Available: https://www.statista.com/statistics/417295/facebook-messenger-monthly-active-users/.

[53] F. Bentley and K. Tollmar, "The power of mobile notifications to increase wellbeing logging behavior," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13, Paris, France: ACM, 2013, pp. 1095–1098, ISBN: 978-1-4503-1899-0. DOI: 10.1145/2470654.2466140. [Online]. Available: http://doi.acm.org/10.1145/2470654.2466140.

[54] W. Luton, *Free-to-Play: Making Money From Games You Give Away*, W. Luton, Ed. Pearson Education, 2013, ISBN: 9780133411249. [Online]. Available: https://books.google.co.uk/books?id=QIXNquWvB2oC.

[55] Author, "How to design outstanding feedback loops," *Smashing Magazine*, 2013. [Online]. Available: https://www.smashingmagazine.com/2013/02/designing-great-feedback-loops/.

[56] R. Dawkins, *The Selfish Gene: 30th Anniversary Edition*, ser. ISSR library. OUP Oxford, 2006, ISBN: 9780199291151. [Online]. Available: https://books.google.co.uk/books?id=go0e5sBRznYC.

[57] S. Bakhshi, D. A. Shamma, L. Kennedy, Y. Song, P. de Juan, and J. J. Kaye, "Fast, cheap, and good: Why animated gifs engage us," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16, San Jose, California, USA: ACM, 2016, pp. 575–586, ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858532. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858532.

[58] *Celebrating 30 years of the gif at facebook newsroom*, https://newsroom.fb.com/news/2017/06/celebrating-30-years-of-the-gif/, Accessed: 2017-09-12, 2017.

[59] Mozilla, *Audio HTML5 Element, MDN Web docs*. [Online]. Available: https://developer.mozilla.org/en/docs/Web/HTML/Element/audio#Attributes.

[60] S. Consolvo, P. Klasnja, D. W. McDonald, and J. A. Landay, "Designing for healthy lifestyles: Design considerations for mobile technologies to encourage consumer health and wellness," *Foundations and Trends in HumanComputer Interaction*, vol. 6, no. 34, pp. 167–315, 2014, ISSN: 1551-3955. DOI: 10.1561/1100000040.

[61] S. Michie, M. M. van Stralen, and R. West, "The behaviour change wheel: A new method for characterising and designing behaviour change interventions," *Implementation Science*, vol. 6, no. 1, p. 42, 2011, ISSN: 1748-5908. DOI: 10.1186/1748-5908-6-42.

[62] B. Verplanken and S. Orbell, "Reflections on past behavior: A self-report index of habit strength," *Journal of Applied Social Psychology*, vol. 33, no. 6, pp. 1313–1330, 2003, ISSN: 1559-1816. DOI: 10.1111/j.1559-1816.2003.tb01951.x.

[63] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006. DOI: 10.1191/1478088706qp063oa.

[64] A. Tversky and D. Kahneman, "Loss aversion in riskless choice: A reference-dependent model," *The Quarterly Journal of Economics*, vol. 106, no. 4, pp. 1039–1061, 1991. DOI: 10.2307/2937956. eprint: /oup/backfile/content_public/journal/qje/106/4/10.2307/2937956/2/106-4-1039.pdf. [Online]. Available: +http://dx.doi.org/10.2307/2937956.

Appendices

1. Full Interview Transcripts

Why did you pick that habit?

Participant 1: *I was already doing meditation, but not the parts of meditation that were beneficial afterwards, I wasn't making it a habit. I wanted to not be stressed out of nut.*

Participant 2: *Want to be more flexible, easier habit, I am not going to do the plank every morning, it is easier with not too much effort, less time consuming.*

Participant 3: *I wanted to write in the morning, but half way through I changed to the later because I wasn't doing it.*

Participant 4: *It's something I wanted to do for a long time, so I thought this would help me make it more regular, I downloaded headspace meditation app, but it was hard to remember to actually do it. Using the bot in combination with headspace was good. Because headspace doesn't send me iphone reminders, chatbot is better with reminders because it is in messenger. So I was doing the habit with headspace while using chatbot.*

Participant 5: *It is something I wanted to do for a while, stress relief habit. Chance to give it a go for hope of it staying in the longer term.*

Participant 6: *Picked it because it is something that successful people do and I get stressed really easy.*

Participant 7: *I wanted to be more active and haven't had any exercise at all.*

Are you still doing that habit?

Participant 1: *Meditation has been harder, but the first few days I was like, I now know how to do this. But if I got distracted it was harder to remember.*

Participant 2: *No not stretching.*

Participant 3: *No, just keep forgetting to do it or can't be bothered.*

Participant 4: *Yes, but, its not as regular, I do miss it.*

Participant 5: *I did it every now and again, usually I think about it, it depends, if I had a long day, sometimes I want to go to bed. Done it a couple of times, done it more when I was using the chatbot.*

Participant 6: *I do it when I remember, I tend to do it now before bed, rather than after lunch, only if I remember.*

Participant 7: *No, been out of town recently.*

Did you like the chatbot?

Participant 1: *Neutral on it. Good that it let me know, but I am not sure about it.*

Participant 2: *Pretty good.*

Participant 3: *Yes I liked it. It has been great for reminding me to write.*

Participant 4: *Would've liked to of interacted with it a bit more, and to of been able to ask it for help and support throughout. Maybe it would tell you tips, like how to set aside some time to do things and ideas on how to meditate, where to go, different apps to use, or things like that. Liked to of had a bit more interaction with the bot.*

Participant 5: *Depending on my mood. If I was feeling okay it would remind me. Otherwise would get annoying, I ignored the message. If he said 'Not Yet', it would remind me sometimes when I already got to bed, (10.30pm), so needed a more precise time, maybe time boundaries.*

Participant 6: *Liked it because it reminded me 2 hours later! This other app that I used, didn't.*

Participant 7: *Yes, but was just always the same, repetitive. I was expecting different messages, like 'Hey man' or 'Hey [name]'. It needed to be a bit more caring about the person and a bit less like a robot.*

How were the rewards?

Participant 1: *I didn't like the rewards, getting the GIF I was like, oh. I skipped over the GIF. I just wanted to get rid of the red dot on messenger. Because it was a bot, I just wanted to get rid of it. After the fourth time I saw same GIF it was bad. If it recognised my progress it would've been better. Based on my trend, my progress and if I done it on time. They didn't encourage me to move forward with my habit or set me back.*

Participant 2: *A bit weird, a bit funny, weird just a black screen. I liked the buttons that you could pick. But I didn't need a reward, it was unnecessary, the message was enough. Maybe use stats as they are cool, because you can reflect on data. Better to have info, like 'you are almost there'.*

Participant 3: *Some of them were funny but maybe it works with other people and not me. I wanted different types of rewards and I didn't really like them. Maybe have, a measurement of something, if a person starts doing good, writing then give them their measurement, giving them constant well done's and stats. When they feel like they are making progress, they are more encouraged to keep doing it, rather than random music.*

Participant 5: *Cute, but I didn't think it gave me an intensive towards meditation, nice little extra. Not the sort of thing you want just before you go to* *bed, as I try to avoid the computer and phone before bed, not good for your sleep. But liked them.*

Participant 7: *I was always waiting to open them, the combination was perfect, it was funny. They were repetitive but not annoying.*

Do you want it back?

Participant 1: *I am interested in having it back and it was good while it was there.*

Participant 2: *Yes, if it was enclosed in a Fitbit, because it would be all in one place.*

Participant 3: *Yes, but maybe, but only thing that I didn't like, was my reward.*

Participant 4: *Yes, if it had a bit more interaction with it, I would. In its current form, I wouldn't. It made me think I should download the headspace app on my phone. But the chatbot reminders were useful.*

Participant 5: *Would be more willing to give it a go if it had a more precise time, then yes.*

Participant 6: *Yes, definition. But I also downloaded an app that sends you a reminder, because they sent you a mindful message. But Harry's Habits had the snoozing 'Not Yet' thing, that was better.*

Participant 7: *Yes. The bot was a helpful tool to remind me to do things.*

Additional Comments

Participant 1: *Would be fine without rewards.*

Participant 3: *I participated in it to change my habits. Problem was with being consistent with a habit. I was very busy for a week and so didn't really write.*

Participant 4: *I felt like I needed the rewards. I wanted to put my own habit in and wanted drawing.*

Participant 6: *First few days were great, stuck to it. Bot was easy to setup. After a few days, I read the message and I would just do it later. Moved to new house and ran out of data, so got no data. Worked well at beginning, got worse and worse, still doing it, but it was at midnight.*

Participant 7: *I used the stretching as reminder for that and then to be more active, and go swimming. Sometimes I remember it, but don't always do it still. I want a chatbot to continue to help with my depression, my therapist said: every hour each day set alarm, saying how are you feeling right now, what are you feeling, is it based on reality or based on expectations? The chatbot for me, is very helpful to break my cycle, in my case, sometimes I* *get a thought and it cycles in my head, these things break this. After a while I stopped setting reminders, but only after 6 months, these repetitive things helped make a habit and helped me stick to a routine. But for this project, I wanted some different messages because otherwise, it is just an alarm that I had before. I want something more to replicate some sort of person.*

2. Recruitment Adverts