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

Harry Mumford-Turner

School of Engineering
University of Bristol
SEPTEMBER 2017

Supervision from
Dr. Oussama Metatla &
Dr. Katarzyna Stawarz.

I hereby declare that this dissertation is all my own work,
except as indicated in the text:

Signature ______________________________

Date ______/______/______

University of BRISTOL

Acknowledgements

Thank you to my supervisors Dr. Oussama Metatla and Dr. Katarzyna Stawarz, for their patience and guidance through this project.

Executive Summary

Rewards motivate people to complete actions. Habit formation systems use rewards to motivate people to form habits. This thesis looks at the affect of three types of positive reinforcement rewards on habit formation delivered by a chatbot, from three modes: visual, auditory and visual-auditory combined. The findings are evaluated against two hypotheses: i) rewards effect on habit performance and automaticity, ii) multiple modalities effect on habit performance and automaticity. 60 people participated in a 4-week study followed by voluntary semi-structured interviews. The findings showed that participants receiving the bot-delivered rewards had higher habit performance than the control group without rewards. A correlation was found between the habit formation method and habit automaticity. However, all participants interviewed (N = 7) found a drop in habit performance after one week without the prototype. Further research for using different rewards with behaviour change technology is needed to validate how each modality affected habit automaticity and habit performance.

In summary this research produced four key deliverables.

1. Review of existing behaviour change techniques and habit formation rewards with research into how they are used in technology.
2. Design recommendations for building chatbots that deliver rewards in different modalities.
3. Constructing Harry's Habits, a chatbot to track habits and deliver rewards from three modalities, visual, auditory and visual-auditory.
4. Running and evaluating Harry's Habits in a 4-week trial with 60 participants to test how each reward impacted habit automaticity and habit performance.

Definitions

Human Computer Interaction (HCI) - Field of computer science that studies how people interact with computers.

Modality - 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.

Chatbot - A method of communicating with a computer system via a conversation.

Contents

1 Introduction

Understanding how to design systems that support behaviour change is important to human computer interaction (HCI) 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 [3]. 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 [4]. 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 [5] and habits are more effectively developed when specific and measurable goals are set [6]. Psychology defines habits as learned automatic cue-response actions. The action will perform automatically in response to a trigger that has been actioned repeatedly in the past [4]. 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, 4].

Technology can help people stick to a routine by sending repeated messages [7] and encouraging people [8]. However, these techniques do not always work and may build repetitive actions rather than habit automaticity [9]. 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 [10, 11]. Habit automaticity is a measure of habit strength [12] and is the key that removes this dependency [13]. Automaticity can be increased by building motivation to complete the action [14, 15]. Motivation can be encouraged by giving people positive reinforcement rewards after they complete an action. 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 [16] advised that delivery of interaction should span different modalities to increase retention and better suit the needs of users. Although interaction across modalities is important, this research does not have configurable feedback, but aims to compare how each type of feedback can affect motivation. Monetary (extrinsic) rewards can hinder motivation [15], whereas, satisfaction-based (intrinsic) rewards can be beneficial to motivation and should be preferred. This study uses a chatbot to deliver intrinsic positive reinforcement rewards from different modalities to see how participants habit automaticity and performance are affected.

This thesis is comprised of four main sections. First, existing literature into habit formation, behaviour change technologies and existing chatbots is reviewed to present three hypotheses. Second, a 4-week situated study is conducted to test the hypotheses with analysis of how each modality delivered by the chatbot, visual, auditory and visual-auditory, impacted participants habit automaticity and the number of habits participants marked as completed. Third, discussion of the results along with follow up interviews about how the bot-delivered rewards encourage people to stick to a routine and perform their habit. Finally, the affect of this habit tracking chatbot on participants and the success of the prototype for conducting research is evaluated against the hypotheses presented.

1.1 Aims and Objectives

The project aims to deliver insight into how rewards from different modalities affect habit automaticity and habit performance, opening up new research avenues for investigating the use of chatbots as vehicles for promoting behaviour change and aiding research trails.

To achieve the aim the following objectives are followed.

1. Present an overview of existing research into:
   (a) Habit formation and behaviour change techniques.
   (b) Different types of rewards 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 guidelines that focus on rewards for habit formation.
3. Create a prototype that tracks habits and delivers rewards from visual, auditory and visual-auditory modalities, using these guidelines to ensure the prototype is based on theory.
4. Conduct and analyse a 4-week study to evaluate the success of the prototype and the impact of each modality on habit automaticity and habit performance.
5. Evaluate habit automaticity and habit performance for each mode. Finally, determining the success of Harry's Habits.

1.2 Added Value

Evaluation from real-users reveals positive and negative aspects of the requirements. The follow-up trail in determines if the requirements were effective for building habit automaticity and tests the validity of the hypothesis. If user's habit automaticity does not increase, the project still presents a novel method of interacting with users and track habits. A system evaluation provides value on how to build a chatbot to deliver rewards from different modalities to support habit formation. Finally, the project opens up new research avenues for investigating the use of chatbots as vehicles for promoting behaviour change.

This thesis analyses literature around habit formation and rewards from different modalities, constructs a prototype to deliver rewards, conducts an evaluation trial and summarising with design guidelines and prototype analysis.

2 Background

2.1 Habit Formation

Habits are automatic actions that require little concious effort [4]. 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 [5] and habits are more effectively developed when specific and measurable goals are set [6]. The formation of new habits requires behaviour change. Three elements are needed to make this change permanent: positive reinforcement, repetition and contextual cues [17]. 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 [8]. Contextual cues act as triggers with constant repetition, as habits, on average, take up to 66 days to form [18]. However, people still fail at forming new positive habits and give up, often due to their lack of routine [4, 2].

Contextual Cues

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 stretching habit, you could attach it onto an existing habit like brushing your teeth. The contextual cue of brushing your teeth will trigger you to stretch. Behaviour change literature [19] shows that attaching habits onto existing event-based cues are easier to remember, when compared with time-based habits, e.g. stretch every 4 hours. These help connect the contextual information with the action and builds habit automaticity [20]. Further research into the design implications of contextual cues shows how multi-cue routines are more effective that a single cue [21].

Repetition

The process of creating a new habit takes on average 66 days of repetitive use [18]. 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). However, existing routines and cues are needed before the action develops into a habit [22, 23]. An existing routine acts as a trigger to motivate the desired action. Context from that routine serves as the cue for the trigger. For example, if you wanted to adopt a habit to weigh yourself every day, you could attach it onto an existing habit like brushing your teeth. The contextual cue of brushing your teeth will trigger you to weigh yourself. When designing behaviour change interventions, using different types of cues can be beneficial. Multi-cue routines have shown to be more effective than a single cue [21]. Attaching habits onto existing event-based cues are easier to remember [19] when compared with time-based habits, mainly due to change in time with change in environment, e.g. the weekend [9]. Event-based cues help connect the contextual information with the action and builds habit automaticity [20].

Rewards

Rewards grant motivation, fuelling the belief in success and self-efficacy, which plays a large part in forming habits. Some researchers [14] 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 [24]. This technique has proved to be an effect method of increasing repetition as shown in slot machines that vary their reward payout. But although variable rewards increase habit repetition, they hinder habit automaticity [24], which is the key to create permanent and long-lasting behaviour change.

Positive Reinforcement

This thesis discusses intrinsic positive reinforcement as the method of reward, rather than other types of rewards e.g. negative reinforcement. Rewarding a person with intrinsic positive reinforcement strengthens the habit by giving the feeling of satisfaction [2], particularly in relation to interesting tasks [15]. This study explores different types of positive reinforcement rewards and how they effect behaviour change.

Technology

Research [25, 26] into how mobile systems can support habit formation and behaviour change, shows a large number of habit forming systems are mobile apps. Studies into the effectiveness of these apps has been recently conducted [13, 10] revealing that although most of these apps are rated highly, they do not ground themselves in behaviour change theory. Further surveys of these apps [26, 25] 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 [27]. This is also the case with many behaviour change systems, when the system is removed any improved performance is lost [10, 11].

Building habit automaticity requires the desired action to be built around an existing routine [18, 19]. Technology should allow for this and help with routine creation [28]. Additional checks should guard against changes in routine to remind people about their habit if their situation changes. Technology can be used to send back-up notifications and post-completion notifications to see if the action has already happened. This is vital for sustaining performance and building habit automaticity.

Two strategies for designing habit formation systems are reviewed: Stawarz et al. [13] discuss guidelines for building habit forming systems and Weiser et al. [29] show that '*motivation is a key requirement for behaviour change*' presenting five design principles and six requirements about the implementation mechanics of habit forming systems that focus on rewards and motivational needs. These share three recommendations: i) make personalized, well-defined, structured multi-cue routines, with examples and support users choice of not setting remembering strategies, ii) reminders can effectively support prospective memory in the short-term, increasing the logging of health data [30] and educating them about how they should perform in the long-term, iii) 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. These 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. This project builds upon this set of requirements, combining them into a new set of design requirements for habit formation systems that focuses on rewards.

2.2 Guidelines

The combined guidelines based on methods from [31, 29] create a list grounded in habit formation theory and focused on rewards. Each requirement provides detailed breakdown about why it's used and what mechanics it relates to from theory.

1. Help users define a memorable strategy.
Make personalized, well defined, structured multi-cue routines and support users choice of not setting remembering strategies. In addition, provide examples of some strategies to users.

2. Give users small difficult tasks.
Turn the bigger habit into smaller assignments to make it more enjoyable, being careful to not make them forced. These should be user specified to support user autonomy, but should be specific and challenging to get better results.

3. Enable competition, comparison and cooperation.
Friends, teams, groups, leader-boards and collections.

4. Show insights for improvements and support changes.
Give users meaningful and accumulated instant feedback based on their system usage.

5. Remind them about cues and remembering strategies.
Reminders can effectively support prospective memory in the short-term, increase the logging of health data [30] and educate them about what they should perform in the long-term.

6. Reward users.
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. These 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. Use achievements and badges as means to identify methods that enable internalisation of externally motivated behaviour.

7. Disable cue reminders when behaviour is routine.
Relying on reminders in the long-term can hinder habit development, therefore ease off from reminders later.

8. Check if the action has already happened.
People find it easy to forget whether an automatic task was completed, check to see if the action has already happened.

Using these recommendations for building habit formation systems that focus on delivering rewards, we decide what type of rewards are needed and discuss what modality would best suit positive reinforcement.

Combined	**Stawarz. [31]**	**Weiser et al. [29]**
1. Help users define a memorable strategy.	REQ 1. Help users define a good remembering strategy. REQ 2. Provide examples of good remembering strategies	Design REQ 2. Support User Choice. Design REQ 4. Provide personalized experience.
2. Give users small difficult tasks	NA	Mechanic REQ 3. Offer challenges
3. Enable competition, comparison and cooperation.	NA	Mechanic REQ 5. Competition & Comparison. Mechanic REQ 6. Cooperation
4. Show insights for improvements and support changes.	REQ 3. Provide suggestions for strategy improvements and support changes.	Design REQ 1. Offer meaningful suggestions. Mechanic REQ 2. User Education.
5. Remind them about cues and remembering strategies.	REQ 4. Remind about cues and remembering strategies.	Design REQ 3. Provide User Guidance. Mechanic REQ 2. User Education.
6. Reward users.	NA	Mechanic REQ 4. Rewards.
7. Disable cue reminders when behaviour is routine.	REQ 5. Disable cue reminders when the behaviour becomes a part of a routine.	Design REQ 5. Design for every stage of behaviour change.
8. Check if the action has already happened.	REQ 6. Help users check whether the habit has already happened.	Mechanic REQ 1. Feedback.

Figure 1: Guidelines for building habit formation systems that focus on rewards.

2.3 Modalities

The three chosen modes, visual, auditory, and visual-auditory combined, are chosen for testing. Feedback from each of these different modalities has been shown to improve task performance for small tasks [32]. Combining modalities are essential for user interaction enabling systems to be highly flexible and accessible to varying ages [16]. Giving people a choice of mode to interact with gives people a highly configurable, non-disruptive system that provides alternative ways of interaction [33, 34]. Technology to support behaviour change should be non-disruptive to encourage usage. Habit rewards should not be annoying as positive reinforcement should be a satisfactory experience. Research into *visual*, *auditory* and *visual-auditory* feedback was identified to understand how they impact behaviour to see how habit formation is impacted.

2.3.1 Visual

Visual feedback can encourage task performance consistency as demonstrated in one study by Matthew Lee et al. [11] 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 sound when designing for behaviour change technology, especially with a varied target market [35]. For example, combing different sounds for different actions to suit different users [36]. 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 the completion of a simple task, can be successful e.g. [38, 39]. A meta-analysis of 43 multi-modal studies [40] revealed that it was the most effective to increase 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 [32]. 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]. In addition, 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 it will be useful to compare the success of the task with the other modes. Using a combination of modalities for interaction gives a means of communication to people with varying levels of sensory awareness. A choice of mode when designing for interaction [16] is important, but system consistency is necessary.

Other Modes

The majority of electronic activity monitors have behaviour change techniques and these monitors present a medium which behaviour change interventions could occur. By using tactile vibration integrated into a wearable device. A survey on activity monitors [42] ranked Fitbit (`www.fitbit.com`) devices as good vehicles for behaviour change techniques. Therefore, the Fibit, would be a good primary platform for integrating vibration for rewards. However, due to technical limitations, as discussed in Section 6, we are unable to implement them.

Mapping Modalities

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 the visual is inferred, therefore this mapping is *semi-congruent*. Next the technology used to delivery these rewards are reviewed.

3 Design

To verify our hypotheses, a prototype was constructed to track habits and deliver rewards. The platform for the tool needed to be highly available for participants, interactive and time effective to build. A system on a mobile device grants us access to a highly available, contextually aware and interactive platform [27, 43]. We will base the design on previous recommendations for building behaviour change systems.

3.1 Existing Technology

Interaction with current habit formation systems is often via a mobile app. This creates a notable difference in the person when the system is removed [27]. This is also the case with many mobile feedback systems that aid with behaviour change. When we remove the system any improved performance is lost [10, 11]. Surveys of current habit formation systems [25, 26, 43] revealed most of them fall into a low behavioural theory adherence scale. Research into how to build systems that are theory-based, suggest four main stages for designing health technology solutions. Conceptualisation, Formative Research and Pretesting, Pilot trials and Evaluation trials [43]. The first two stages use the Behaviour Change Wheel framework [44] (method for planning health behaviour interventions). This framework allows us to understand the behaviour, better define the characteristics and turn concept into prototype. Pilot trials test the prototype before it is production-ready without the commitment of a full trial. The final stage evaluates the finalised prototype with a wider range of participants. Next, conceptualisation and design phases are discussed for design steps to build the prototype.

Conceptualisation and Formative Research

This first stage explores different mediums to develop a prototype upon and different methods of user interaction. We conclude with a series of implementation options discussion about how the chatbot was chosen.

When it comes to mobile phones, users have plenty of options for interaction. A popular choice that has revisited the market are *chatbots*. Chatbots are applications that parse questions using Natural Language Processing (NLP) to provide a response. 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 [45], Eliza by Joesphs Weizenbaum, used simple expression matching to return a certain response for user trials. In the present day, these applications (commonly referred to as bots or chatbots), are found integrated into many different apps on the majority of users mobile phones. For example, Facebook Messenger, a popular messaging application (`www.messenger.com`) 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* [46] 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. 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. NLP 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. NLP will not be used to limit the scope of this project and avoid these potential problems. 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.

Another design option is building a *webapp*. However, the functionality is limited, due to the lack of notification features 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 hard to get users to respond. 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 web app 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 the recommendations (Section 2.2).

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

Figure 2: Comparing different prototype mediums.

3.2 Platform

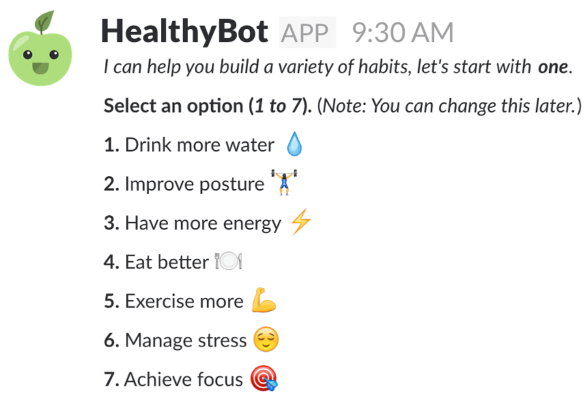


Figure 3: Healthy Bot (`https://healthybot.io`): A Slack bot for forming new positive habits.

There are lots of options about what platform to build the chatbot into. For example, *Slack* (`www.slack.com`) bots provide additional functionality and can complete complex tasks such as habit tracking (Figure 3, into the popular workplace communication service. *Whatsapp* (`www.whatsapp.com`) also provides interactive features with a large user base, however it lacks many additional features. *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 4: 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) [47]
- 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.

Facebook Messenger

Interaction with Facebook Messenger chatbots is varied. Existing bots use three methods of interaction that utilise different Facebook Messenger UI paradigms. First, NLP is fully utilised. 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.

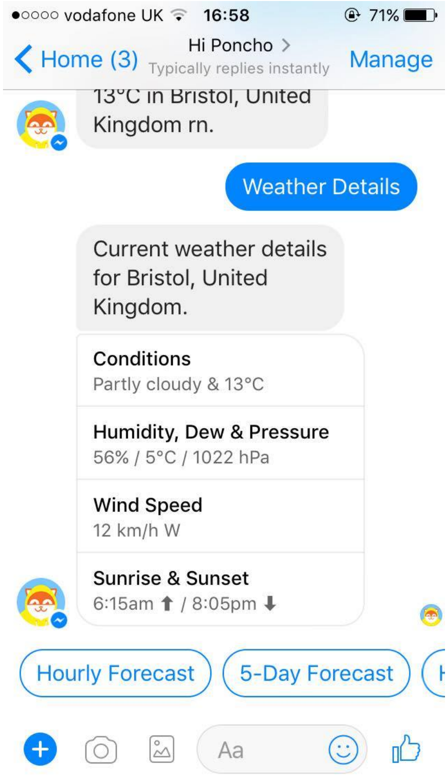


Figure 5: Poncho: An example of Facebook Messenger Weather Chatbot

Second, NLP is not used at all and Quick Replies are heavily utilised. Everist (`https://www.everist.ai/`) and Cleo (`https://meetcleo.com/`) bots use Quick Replies almost like a menu with a set of options. After users message the bot once, the set of quick replies returns 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 (`http://www.hellojoy.ai/`) 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.

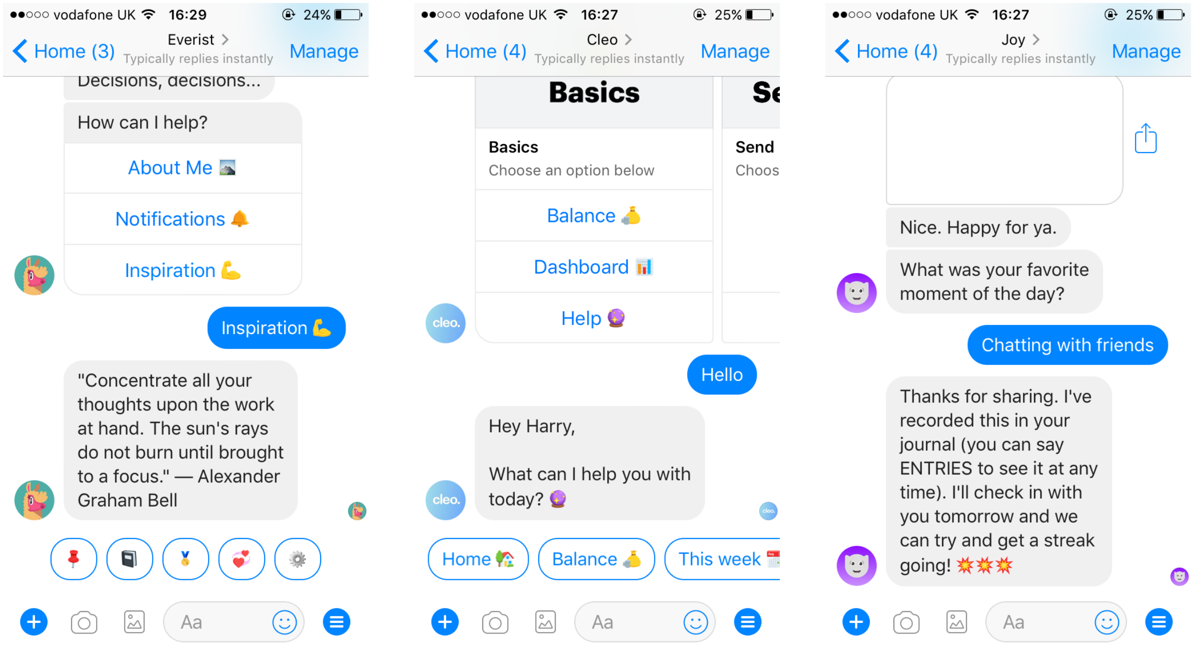


Figure 6: Examples of chatbots performing different actions

Supporting Habit Formation

Harry's Habits is built into an existing social network that people are used to and will track habits and deliver rewards. The social scaffolding around the interaction with the bot will help reinforce the response to notifications, as people are more likely to respond to the bot because they could perceive the bot as a real person as it is located in a list next to their other contacts. The bot will track simple actions of similar difficulty to reduce the amount of time needed before the action turns into a habit [18]. Other design methods to help people form new habits were looked at but will not be added to the prototype to limit the scope. Gamification, such as using gambling elements and engineering luck into games have been shown to encourage interaction [48]. Using monetary rewards to create a feedback loop where people keep coming back for the reward [49]. 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.

Delivering Rewards

The types of rewards are separated into three categories, visual, auditory and visual-auditory. The content of these rewards are experimented with to test if they provide user satisfaction. The rise of social media and increasing use of humorous memes was utilised to provide the motivating content of the rewards. This works well with the prototype integration into an existing social media messaging platform. These rewards will be displayed to the user within the chatbot after they complete their habit. For a playful prototype, visual rewards are light-hearted gifs and auditory rewards are selected to match the gifs.

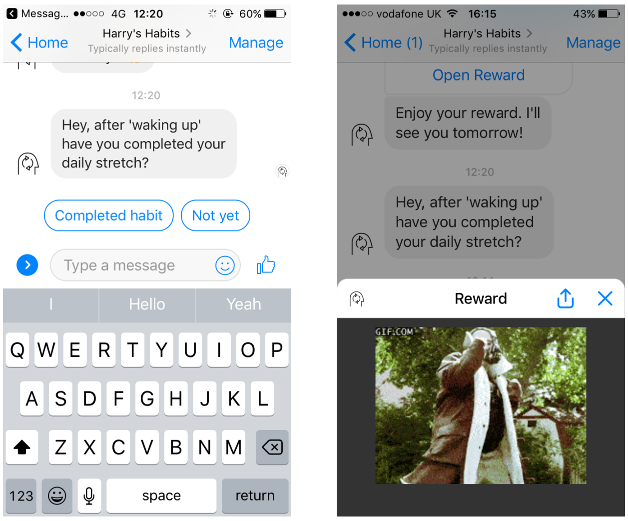


Figure 7: Harry's Habits delivering a back-up notification and a visual reward.

There are two different methods of delivering visual, auditory and visual-auditory rewards to users. First, gifs and audio sent as a message to the user (inline). This has the benefit of being native to Messenger, providing a better user experience and is faster to receive, requiring less buttons to be pressed. However, each reward was not displayed consistently. Visual rewards started as soon as they were delivered, but the auto reward had 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.

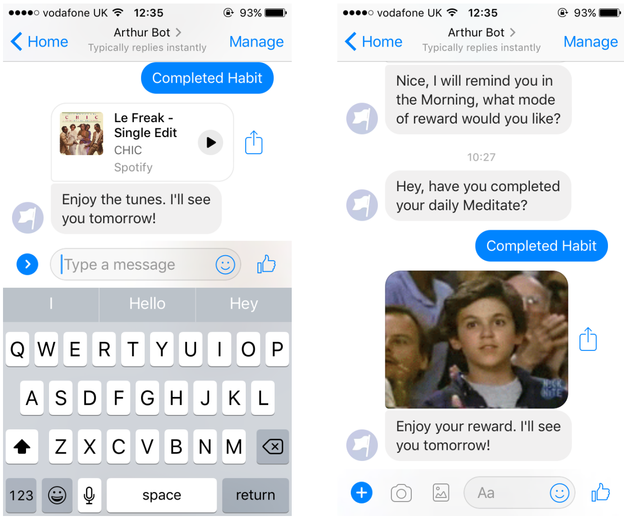


Figure 8: Auditory and visual inline rewards.

The second method allows for consistency across reward deliver. A standard method was deigned to deliver the reward to users, instead of sending the rewards inline a *webview* was used to display a website where users can open their reward. The website could use *HTML*, *CSS* and *JavaScript*. 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.

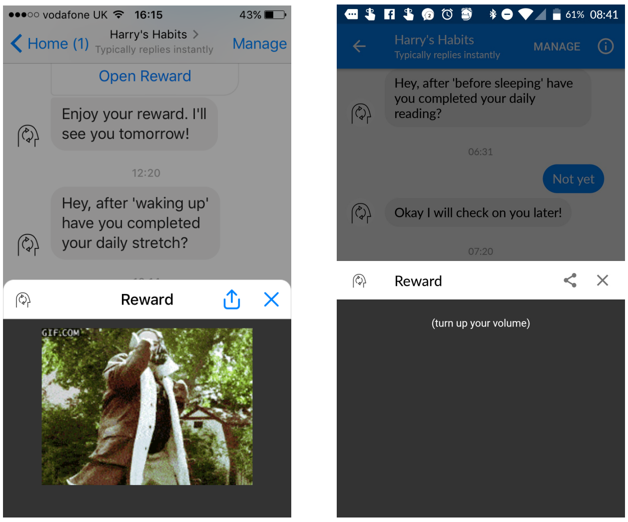


Figure 9: Consistency between every reward, even across devices. Left, visual reward, right, auditory reward.

Design Limitations

First, 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 [50]. 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. 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.

3.3 User Flow

Full user interaction from setup to completing a reward was designed. This setup used a combination of Quick Replies and free text to gather demographic information and complete the setup for each user. Harry's Habits logo was design by the Noun Project by Yu luck (`https://thenounproject.com/term/custom/402041/`).

- **Setup**
  - 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.

- **Trigger**
  - 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 message Quick Reply that leads to a reward.
- **Reward**
  - 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: video with no sound
    * Audio: soundtrack
    * Visual-Auditory: video with sound

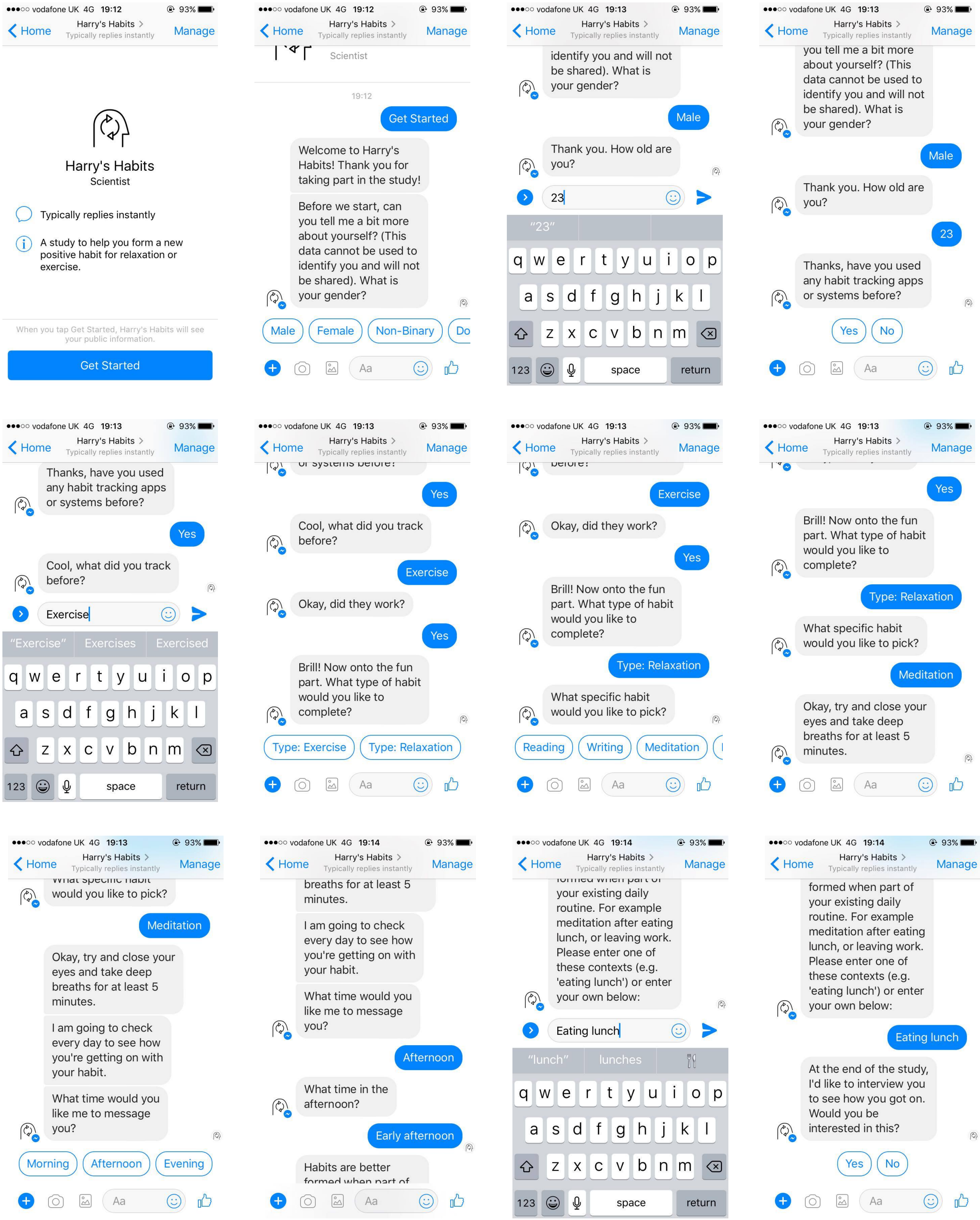


Figure 10: Setup flow for Harry's Habits.

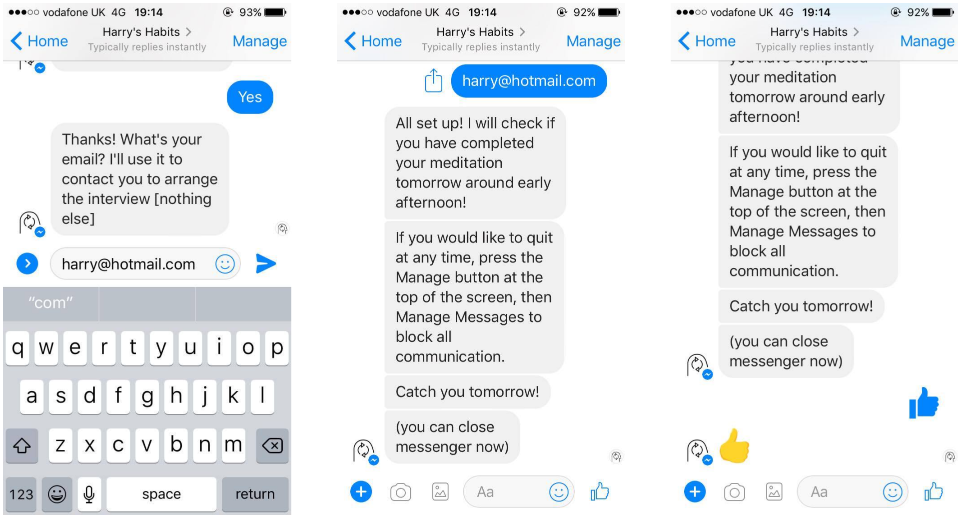


Figure 11: Continued: Setup flow for Harry's Habits.

4 Implementation

Figure 12 shows an abstract overview of communication between the chatbot and Facebook Messenger servers. The user interacts with the chatbot using the Facebook Messenger platform on either a mobile phone or website. When the user sends a message to the chatbot it is sent to Facebook servers that decide what chatbot application to forward the message onto. The message comes in a *json* format that is parsed by the chatbot application to understand how the user interacted and what message was sent. The chatbot reads and writes the necessary information into the PostgreSQL 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. All code is open-source and hosted on *GitHub* (`www.github.com/harrymt/harryshabits`).

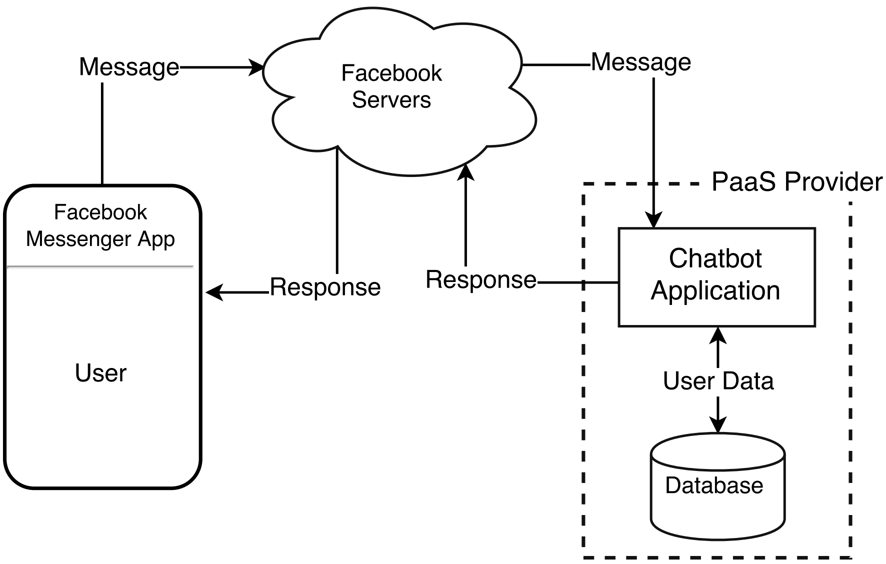


Figure 12: Prototype Component Overview

The prototype was developed in *node.js* (`www.nodejs.org`), 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* [46] 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 NLP to limit the scope of the project. The bot also tracked various 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* (`https://www.postgresql.org/`) database. Figure 13 outlines each database table with what information is stored. The *Facebook 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.

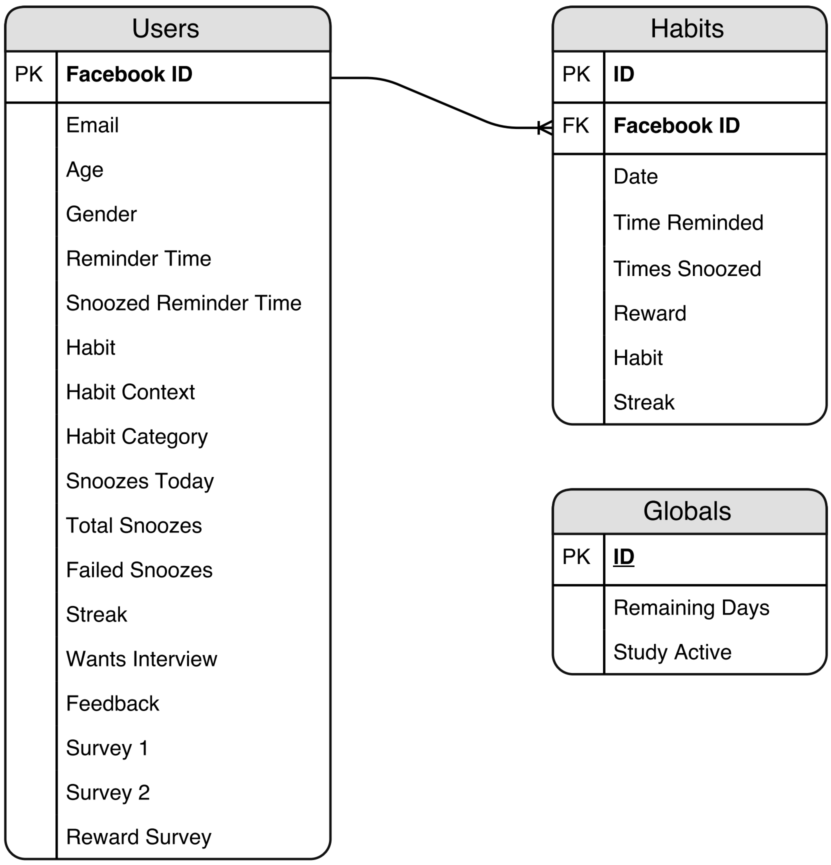


Figure 13: Database Table Entity Relationship

Different languages, server, hosting provider and database provider were considered. JavaScript with node.js on Heroku with PostgreSQL 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 integrated, however, it only provided 3,000 free rows and therefore was discarded in favour of Herokus PostgreSQL database.

Chatbot Overview

When Facebook servers forward a message from the user, it could come in two forms. A simple string of characters with a Facebook ID (*fbid*) or a Quick Reply key with a Facebook ID.

TODO THIS NEXT SECTION

`read()`—

`getUser()`—

`send()`—

`updateUser()`—

A scheduler was used to send backup notifications to each user after their existing routine. Heroku uses a scheduler () to handle this process.

However, the chatbot did not account for limitations with timezone...

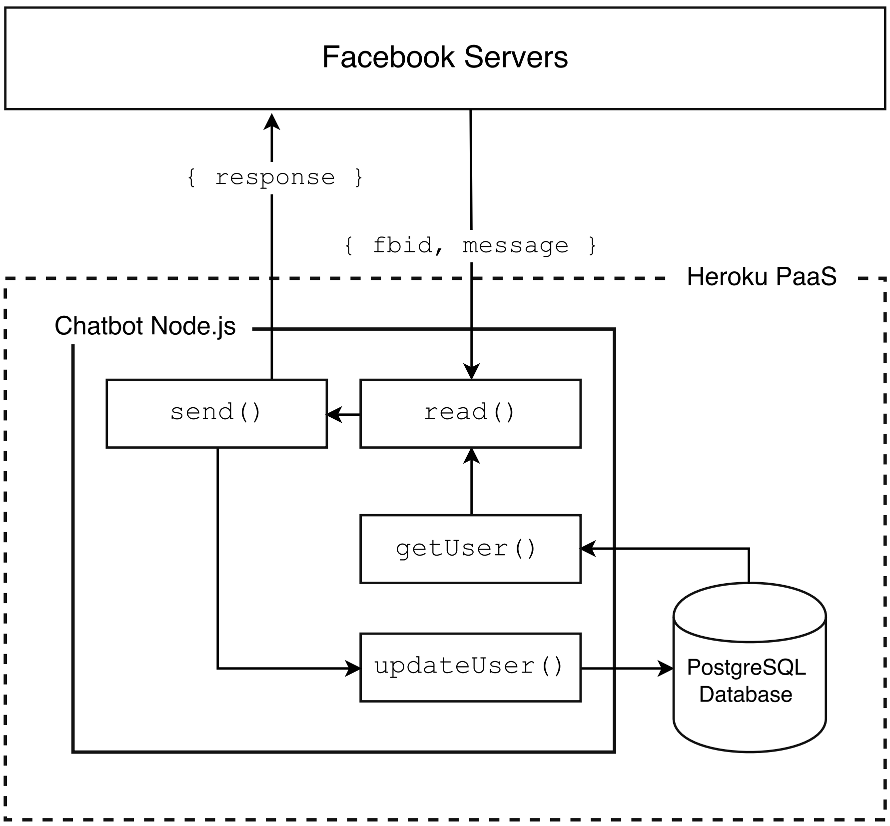


Figure 14: Detailed Overview

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.

Sometimes Facebook servers would send the same message to the chatbot application twice. This would mean the reply would also be sent twice back to users. This issue was very rare and only occurred in a handful of cases and therefore was consider minor. However, it did effect this similar issue when handling free text from users. When the bot asked users to enter in a free text, e.g. email, a flag would be set to wait for free text input of that type. However, if duplicate messages were sent at this time, the application would assume users did not enter anything for the free text and would continue.

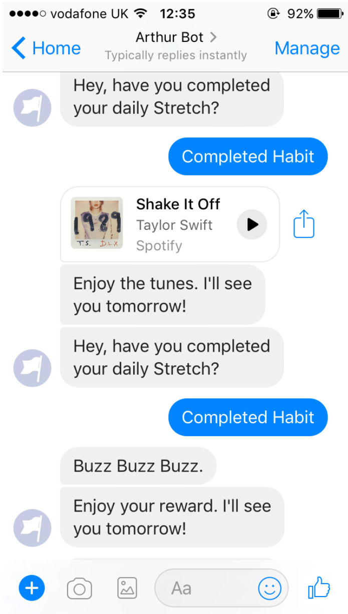


Figure 15: Using vibration as another modality was tested, but due to technical limitations was difficult to implement.

Using vibration as a modality would've been a great additional modality to use. Unfortunately the chatbot sandbox meant that the vibration ability in the phone could not be 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, with the pattern of vibration matching the frequency of the audio. However, the majority of these devices did not have an API that exposed the vibration element. The best method was found 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've reduced the wait time with some additional dialogue. 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 modality was abandoned. 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 backup 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.

TODO: Go to github issues and paste them into here!

4.1 Testing

A test harness was written to perform functional testing on the chatbot. An on-line continuous integration (CI) service was used to programatically run these tests when a *commit* in *version control* was performed on the *master branch*. *Travis CI* (`https://travis-ci.org/harrymt/harryshabits`) was used to conduct the tests to ensure the functionality worked throughout development. A pilot trial with five users also tested the basic chatbot functionality preparing for the full evaluation study (Section 5.2). Next we evaluate the effectiveness of the prototype against hypotheses in a 4-week study.

5 Evaluation

Below we present our evaluation process, that place the prototype in the hands of real people, as recommended by validated guidelines [43, 1].

Many people agree about the importance of designing systems for health and behaviour change [43, 51, 35]. But each have varying opinions about how to evaluate these systems. Klasnja et al. [3] focuses on system usability and does it meet the needs of users. Whereas, Stawarz and Cox [35] argue evaluating a system of this type requires information from other fields to properly consider the systems effectiveness. The validated Behaviour Change Wheel Framework [44] 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 evaluation trials. These will test the long-term effect and efficiency of the bot, with information from two fields of study, HCI and health psychology.

Evaluation trials are the final part of the Behaviour Change Wheel Framework [44]. HCI research that focuses on health interventions [43], demonstrates the importance of evaluation trials for evaluating mobile health systems. These trials have three goals to test: objective-quantitative efficacy, subjective-qualitative feedback measures and real-world feedback about how the system is utilised [3]. I will conduct an evaluation trial for this project.

The length of the trial will be based on two factors, the time needed to form a habit [18] and the results of a previous habit formation trial [13]. First, the number of repetitive days required for an action to be considered a habit varies based on the complexity of the action [18]. 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 [18] showed an increase in habit automaticity after 4 weeks. This project will mirror that timeframe.

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. Chatbot interaction will be removed during the follow up study 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 strength will be measured using a validated 12-question questionnaire that specifically looks at automaticity [52]. Automaticity will also be measured using a validated subset of the questionnaire from [52] to test users habit behavioural automaticity index [12]. This will show the impact each modality has on habit automaticity and test the hypothesis. Participants will fill out the questionnaires [52, 12] at three stages: half-way through the trial (at 2-weeks), after the trial has finished and after the follow up trial.

5.1 Hypotheses

The study tests two hypotheses, the affect of the rewards and the effect of multiple modes versus singular modes. They are measured by 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 [12] questionnaires. In addition to these two hypotheses, the success of the prototype is also evaluated during the interview stage.

H1: The rewards affect on habit performance and automaticity

- *M1: The rewards affect on habit performance*
- *M2: The rewards affect on habit automaticity*

H2: Multiple modes versus singular affect on habit performance and automaticity

- *M1: Multiple modes versus singular affect on habit performance*
- *M2: Multiple modes versus singular affect on habit automaticity*

5.2 Study: Comparing Different Rewards

To explore the influence of different modalities on habit formation, a situated study was conducted, followed by semi-structured interviews. For the purpose of the study a chatbot was developed to help users form healthy habits by providing different types of positive reinforcement. The study and the bot itself are described below.

Briefly talk about Ethics approval, Kathys PHD 5.3.

5.2.1 Method

The study aims to test the hypotheses comparing how well people form new habits based on the rewards and how the bot-delivered rewards effect participants habit formation.

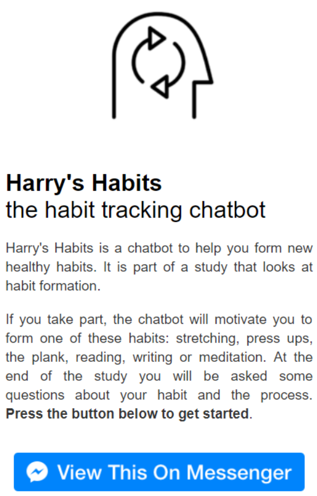


Figure 16: The landing page participants would see before they are taken to the Messenger bot: `www.harrymt.com/harryshabits`.

5.2.2 Participants

60 participants were recruited on social networks (using the landing page: `www.harrymt.com/harryshabits`), and were mostly University students and staff. The process participants undertook is documented by the recruitment adverts (Figure 17), the landing page (Figure 16) and the setup screens (Section 3.3). They 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 3.3).

- 'What is your gender?',
- 'How old are you?',
- 'Have you used habit tracking systems before? If did they work and what were they?',
- 'What habit do you want to track?',
- 'What existing routine would you like to use?',
- 'What time does that routine occur?',
- 'Would you like to be interviewed about your experience after the study?'.

Figure 17: Adverts and feedback received from Facebook.

Modality	*Reward*	*No. Participants*
visual	15s gif	15
auditory	15s audio	15
visual-auditory	15s gif and audio	15
none (control group)	no reward	15

Table 1: Participants were randomly assigned one of these four rewards.

5.2.3 Design

A brief description accompanied these questions, 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. The study used four conditions: visual rewards, auditory rewards, visual-auditory rewards and no rewards (control group). Participants were randomly assigned a condition by the bot after they completed the set-up, 15 were assigned to visual rewards, 15 assigned to auditory rewards, 15 assigned to visual-auditory rewards and 15 with no rewards (control group). 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.

5.2.4 Materials

The bot will collect the amount of habits participants complete versus their reward type. Habit completion will be verified by using the SRBAI questionnaire [12]—A validated set of questions to measure habit automaticity levels. This will be used to test if participants are forming a habit. Various questions about how participants found their rewards were also asked (Appendix 9.1. Finally, interviews with participants discuss the bot-delivered rewards for further verification.

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

5.2.5 Procedure

First, a internal pilot trial was conducted to develop the strength of the prototype. Second, a 4-week situated trial was conducted with real people who wanted to try and form a new positive habit. Finally, follow-up interviews with participants revealed if participants continued with their habit without the bot.

The 4-week evaluation trial was split into two sections. First a 3-week trial tests the success of the chatbot by evaluating the tool and the effectiveness of each modality on participants habit automaticity using a validated questionnaire [12]. During the fourth week, chatbot interaction is removed during a 1-week follow up trial to test if participants continue with the habit. Participants were split into four groups, all groups receiving reminders, three groups receiving rewards each from a different modality, and one group (control group) did not receive any rewards.

At the beginning of the 3-week period, participants 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 (unknowingly 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 (see figure 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 hadn't completed their habit yet, the bot would ask participants if they would like to change the time their routine occurred (Figure 18. This allowed the participants in the beginning to refine the time they would perform their habit.

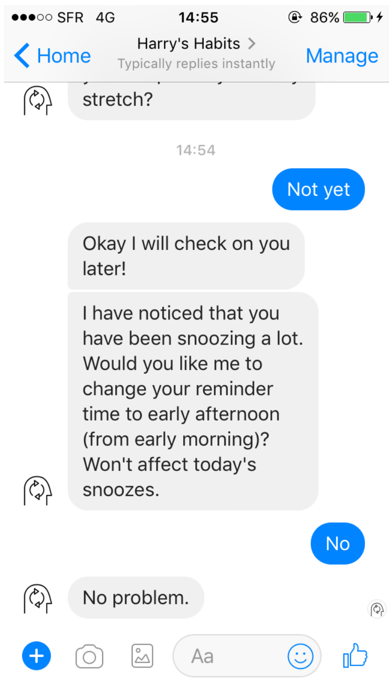


Figure 18: Participants would receive a message asking them if they would like to change their backup notification time.

After 3 weeks of participants interacting with the bot, all participants who completed the set up are 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. Both questionnaires presented 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.

5.3 Results

60 participants connected with the bot by pressing *'Get Started'* in Facebook Messenger on the following platforms: 25 browser, iOS 18 and Android 12. 14 participants (23%) dropped out of the study at various stages (see figure 19): 54 participants (90%) continued interacting with the bot and started the set-up. 39 participants (65%) completed the set-up and out of these 39 participants, 3 participants just ignored all messages from the bot during the trial. Leaving 36 participants (66%) that are considered active throughout and included in the final analysis. These 36 participants are 18-63 years old, (mean: 27 years old, SD: 12), 23 (64%) male, 11 (30%) female and 2 (6%) didn't say. These 36 remaining participants are split into the following modalities: 14 participants (93%) visual, 9 participants (60%) visual-auditory, 7 participants (46%) auditory and 6 participants (40%) had no rewards (control group). 28 participants volunteered for an interview and 7 were carried out.

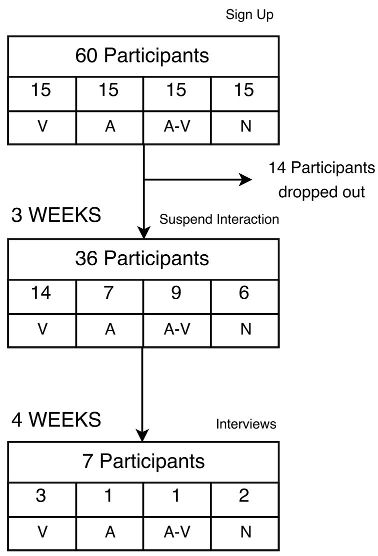


Figure 19: Participant drop-out during the 4-week study.

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 and were sent 0 rewards.

General

TODO: Security of results

TODO:'Does giving users too much personalisation effect the study?'

TODO: Add charts for the below results

Comparing the number of participants who dropped out of the study versus their reward modality, shows that 7 participants who blocked the bot 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 (19%) previously used habit tracking systems and 100% of 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 hadn't tracked before. Meditation was the most popular habit chosen (12 participants 33%), followed by Press ups (8 participants 22%), then Stretching (6 participants 15%). Reading and writing were the least, only selected by 4 and 2 participants respectively. Stretching (6 participants 15%) was the most completed habit based on selection (60 times), ranking 10.0 (where 10.0 = 60/6). Meditation ranked 6.25 (6.25 = 75/12) and the least were the plank and reading with 3.75 (15/4) and 1.75 (7/4) respectively.

Participants (14 participants 39%) with visual rewards 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 and sound had the most number of failed snoozes (10, split over 6 participants, 1 person had 5 failed snoozes). Then it was sound by 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 participants). High streaks (streak ¿ 10) had habits: meditation (total 135 streaks) and stretch (126 streaks). Visual-Auditory rewards had the most streaks (126 combined), control group had 75, then Visual had 60. Stretch had the highest peak streak (18), meditation (15). However, overall, for all habits completed, meditation was the most streaked (269 combined), stretching (231 combined), then press ups (39) then plank (24).

Compare the number of people dropped out verses Modality?

Visual: 7 people that had average 3.85 rewards and avg 2 snoozes
Auditory: 2 people that had avg 2 rewards
Visual and Auditory: 2 people that had an avg of 6.5 rewards.

What was users habit context?

31 participants (86%) 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 people chose the following context: 'Having a snack', 'Sitting in bed', 'Early morning', 'during breakfast' and 'Before sleeping'.

Suggested habit contexts (by the bot):(todo insert table: morning, afternoon, evening habit context)

Morning: - waking up - eating breakfast - arriving at work

Afternoon: - eating lunch - leaving work

Evening: - leaving work - eating dinner - getting ready for bed

Morning: 100: 7 (early) + 66 (mid) + 27 (late) Afternoon: 54: 7 + 4 + 43 Evening: 32: 6 + 15 + 11

How many Facebook API calls were made?

2,995 FB API calls, 319 errors

What was users habit context?

31 people 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 people chose the following context: 'Having a snack', 'Sitting in bed', 'Early morning', 'during breakfast' and 'Before sleeping'.

Suggested habit contexts (by the bot):(todo insert table: morning, afternoon, evening habit context)

When did people most say 'Not Yet'?

Most people said 'Not Yet' in the morning (100 times), specifically mid (66) and late (27) morning.

How many people previously used habit tracking systems before?

7 people (19%) used them before and 100% of those people said they worked. They tracked habits such as 'Diet' (3 people), 'Exercise' (3 people), 'Deadlines' (2 people), 'Audiobook reading' (1), 'Weight' (1). These people all chose new habits that they hadn't listed that they had tracked before.

What is the comparison between the habits people chose and the number of habits people completed?

Meditation was the most popular habit selected (12 people), followed by Press ups (8). Reading and writing were the least, only selected by 4 and 2 people respectively.

However, stretching (6 people) was the most completed habit (60 times) based on number of people who selected it: 10.0, (10.0 = 60/6). Followed by meditation (75 times), 6.25 (6.25 = 75/12). The least was the plank and reading with 3.75 (15/4) and 1.75 (7/4).

The number of habits completed v chosen habit breakdown

Stretching had the most habits completed v chosen habit, 60 times marked as completed, 6 people chose stretching (10 = 60/6).

What is the highest streaked habit?

Meditation had the most cumulative streak (17) followed by stretching (7 people).

What is the correlation between reward and streaks?

High streaks (streak over 10) had habits: meditation (total 135 streaks) and stretch (126 streaks). Rewards wise, VISUAL AND SOUND had the most streaks (126 combined), control group 75, then VISUAL (60)

What was the habit that lasted the longest?

Stretch had the highest streak (18), meditation (15). However, overall, for all habits completed, meditation was the most streaked (269 combined), stretching (231 combined), then press ups (39) then plank (24)

Failed snoozes

Visual and sound had the most number of failed snoozes (10, split over 6 people, 1 person had 5 failed snoozes). Then it was sound by 4.

5.3.1 M1: Habit Performance

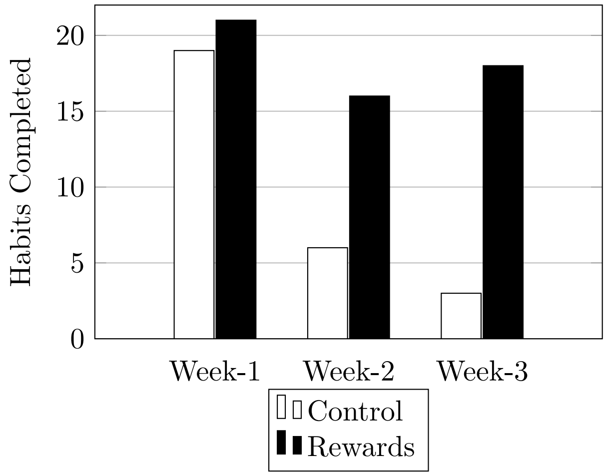


Figure 20: H1M1: The rewards effect on habit performance. The sum of habits completed by participants with rewards versus the control group during 3-week study period.

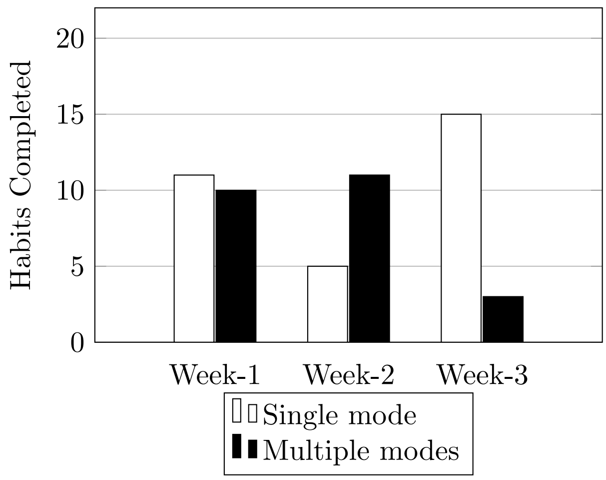


Figure 21: H2M1: Multiple modes versus singular on habit performance. Sum of completed habits for multiple modalities compared with singular modes.

H1M1: The rewards effect on habit performance

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 (see figure 20). Participants were divided into two groups according to their reward mode (Group 1: visual rewards, auditory rewards, visual-auditory rewards; Group 2: control group). There was a statistically significant difference at the $p < .005$ level in both groups for 23 of the Weeks: Week 1, $F(1, 23.20) = 9.48$, $p = .005$, Week 2, $F(1, 33.35) = 4.46$, $p = 0.42$ and Week 3, $F(1, 50) = 17.01$, $p < 0.005$. The effect size for Week 1, Week 2 and Week 3 are large, calculated using eta squared, were 0.25, 0.39 and 0.43 respectively, this shows a large difference in the mean scores.

H2M1: multiple modalities versus singular mode

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 (see figure 21). 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 $p < .005$ level in Week 2 and Week 3, and lots in all groups: Week 1, $F(1, 50) = 0.69$, $p = .410$, Week 2, $F(1, 50) = 23.04$, $p < .005$ and Week 3, $F(1, 50) = 8.85$, $p = .005$. The effect size for Week 1, Week 2 and Week 3 are large, calculated using eta squared, were 0.25, 0.39 and 0.43 respectively, this shows a large difference in the mean scores.

5.3.2 M2: 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), $t(10) = 2.469$, $p < .005$ (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 squared statistic (.37) indicated a large effect size.

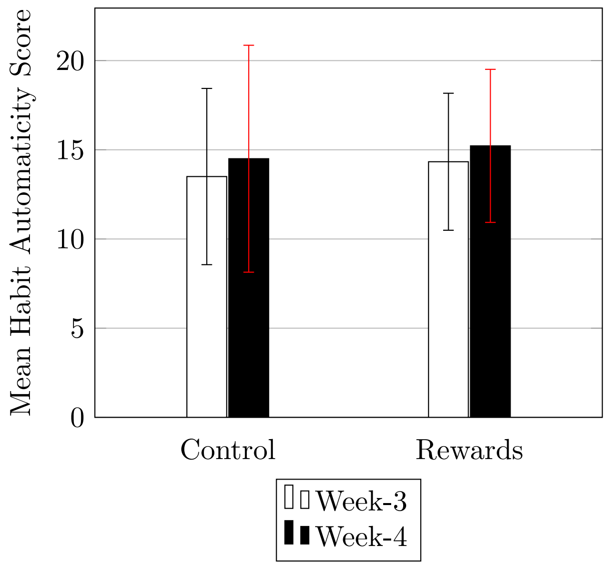


Figure 22: H1M2: The rewards effect on habit automaticity. Comparing mean habit automaticity for rewards versus control group.

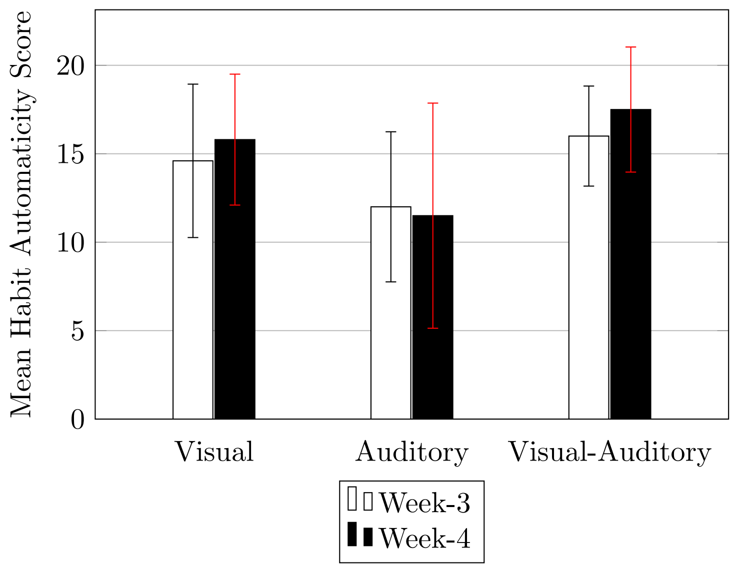


Figure 23: H2M2: Multiple modes versus singular on habit performance. Comparing mean habit automaticity for each group.

H1M2: the rewards effect 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 (see figure 22). 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 = .85, two-tailed). The magnitude of the differences in the means (mean difference = .83, 95% CI: 29.43 to 27.76) was very small (eta squared = .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 = .84, two-tailed). The magnitude of the differences in the means (mean difference = .72, 95% CI: 8.80 to 7.36) was very small (eta squared = .004).

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: visual rewards, auditory rewards, visual-auditory rewards combined; Group 2: control group. There was not a statistically significant difference for the two groups at SRBAI 1: F (1, 9) = 0.02, p = .88, and SRBAI 2: F (1, 9) = 0.07, p = .78. In addition, the difference in mean scores between the groups had, at SRBAI 1: a medium effect with an effect size of .11, and at SRBAI 2: a large effect, with an effect size of .17, both calculated using eta squared.

H2M2: multiple modes versus singular mode 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 (see figure 23). Participants were divided into two groups according to their reward mode (Group 1: visual rewards, auditory rewards; 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 = .33, and SRBAI 2: F (1, 9) = 0.64, p = .44. In addition, the difference in mean scores between the groups had, at SRBAI 1: a medium effect with an effect size of .11, and at SRBAI 2: a large effect, with an effect size of .17, both calculated using eta squared.

5.4 Discussion

This research aimed to understand more about rewards from different modalities and their role in habit formation. Participants receiving bot-delivered rewards completed more habits than the control group without rewards. There was a significant correlation between the habit formation method and habit automaticity. However, this was contradicted during participant interviews (N = 7) where all participants found a drop in habit performance after 1-week without the prototype. Detailed interview discussion and results compare the two hypotheses with each measurement below.

5.4.1 Interview Feedback

The interview transcripts were analysed and participants were anonymised following a thematic approach [53]. 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.

Questions

- General questions about rewards
- Habit formation insight
- Chabot interaction
- Modality interaction
- Are you still doing that habit?

- Why did you pick that habit?
- Do you want it back?

7 interviews with participants outlined their experience with their habit performance after the prototype bot was removed. Participants picked habits they wanted to perform, but when the bot stopped notifying them, they lacked motivation to completed their habit. Some participants enjoyed the rewards at the beginning, but most participants disliked them after the first week.

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, some participants would put the message off and eventually get their performance would get *'worse and worse'*, until they stopped all together. However, after the bot was removed all interviewed participants (N = 7) found 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.

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'*.

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.

M1: 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.

M2: Habit Automaticity

It is inconclusive whether the rewards or the combined modalities effected 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.

5.4.2 Prototype Success

The prototype was somewhat successful at running a research trail. There were several issues and various limitations with development. However, it was generally liked by participants and managed to easily 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. 7 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).

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 a auditory reward.

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. The results found, 31 (84%) 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'*. A participant also pointed out that changing the time their existing routine occurred was prompted to them, but the description about it, was not.

Feedback

Two reviews were publicly left on the bot. They echoed interviewees feelings towards the bot.

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.'*

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.'*

Done these already?

- Add a 'before' or 'after' option to avoid sythetic awkwardness (erasmo) - Liked music rewards. - If you tell the bot they haven't done their habit after their context, when they snooze it isnt going to be their context anymore!!! e.g. lunch at 5pm. The solution is to remove habit context from snoozed reminders.

Development issues:

From the change to different database it was hard to test all aspects of the bot. For example the logic to send end of day messages was altered but not fixed. Changed name of bot but did not update link on landing page Removed FB messenger extensions but didn't remove Boolean value on reward button. Enter negative age Sometimes the reminder time is set to false. Maybe when people don't enter anything? Or if people abandon the setup, check a users conversation The end of day reminders to decide if someone needed to be sent a 'unlucky' message was lost during the migration. People reported multiple thanks messages / and messages in general. If users interrupted the setup they restarted it. This annoyed users. Double messages, j and Mum Algo to auto assign modality has bad defaults " False in reminder times

Discussion about the issues I had w the bot

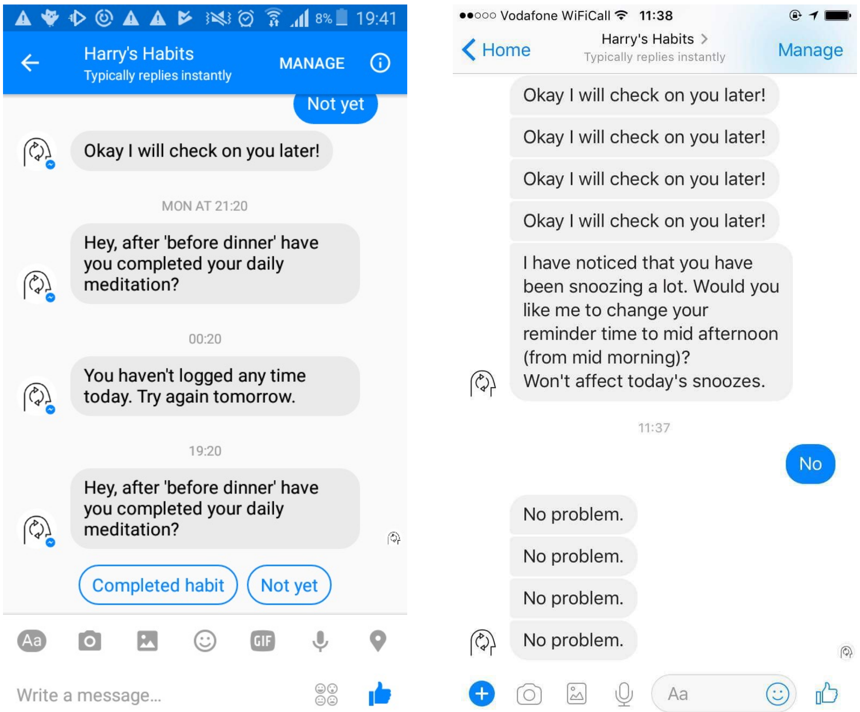


Figure 24: Example of 2 issues that occured, habit context using the word 'after' and the bot sending multiple of the same messages.

Dependence

Streaks could have been better used to give insight to participants progress and challenged them to maintain it, using loss aversion [54] 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.

6 Limitations and Future Work

There were several limitations to these findings. First, 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. Second, this only applies to intrinsic positive reinforcement rewards, further research into how different types of rewards from different modalities is needed. Third, we outline the dependence between our prototype and participants habit performance during the 3-week period and how this has disadvantages for habit formation. Fourth, TODO: Vibration limitations. Fifth, TODO: using wit.ai for NLP. Finally, the content and method of delivery is another variable that effects these results, additional studies into bot-delivered rewards 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 recommendations for building reward-based chatbots that support habit formation. A prototype is designed and constructed from these recommendations, 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) How is habit performance effected. ii) How is habit automaticity effected. 36 participants completed the study interacting with the bot for 3 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 7 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 effected habit performance and habit automaticity. More specifically, the results found that the bot-delivered rewards improved habit performance. Participants were more likely to complete their habit if given the reward. Habit performance was also effected by different modalities, although not in the way our hypotheses assumed, as singular modalities had higher habit performance than visual-auditory rewards. Finally, the limitations of the study do not show any clear 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. We encourage the use of these findings with further comparisons against different types of rewards for behaviour change technology. Finally, we hope this opens up new research avenues for investigating the use of bots as vehicles for promoting behaviour change and forming new habits.

8 References

[1] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. How to evaluate technologies for health behavior change in hci research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3063–3072, New York, NY, USA, 2011. ACM.

[2] Phillippa Lally and Benjamin Gardner. Promoting habit formation. *Health Psychology Review*, 7(sup1):S137–S158, 2013.

[3] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. How to evaluate technologies for health behavior change in hci research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3063–3072, New York, NY, USA. ACM.

[4] Wendy Wood and David T. Neal. The habitual consumer. *Journal of Consumer Psychology*, 19(4):579–592, 2009.

[5] Benjamin Gardner, Kate Sheals, Jane Wardle, and Laura 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*, 11(1):135, 2014.

[6] 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*, 21(2):117–122, 1995.

[7] Roelof A.J. de Vries, Khiet P. Truong, Sigrid Kwint, Constance H.C. Drossaert, and Vanessa 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*, CHI '16, pages 297–308, New York, NY, USA, 2016. ACM.

[8] Robert J. Vallerand and Greg Reid. On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory. *Journal of Sport Psychology*, 6(1):94–102, 1984.

[9] Catalina Danis, Katherine Vogt, Justin D. Weisz, Yu Ma, and Russell 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*, CHI EA '17, pages 2480–2486, New York, NY, USA, 2017. ACM.

[10] Ian Renfree, Daniel Harrison, Paul Marshall, Katarzyna Stawarz, and Anna Cox. Don't kick the habit: The role of dependency in habit formation apps. pages 2932–2939, 2016.

[11] Matthew L. Lee and Anind K. Dey. Real-time feedback for improving medication taking. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2259–2268, New York, NY, USA, 2014. ACM.

[12] Benjamin Gardner, Charles Abraham, Phillippa Lally, and Gert-Jan 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*, 9(1):102, 2012.

[13] Katarzyna Stawarz, Anna L. Cox, and Ann 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*, CHI '15, pages 2653–2662, New York, NY, USA, 2015. ACM.

[14] A. Bandura. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev*, 84(2):191–215, Mar 1977.

[15] Edward L. Deci, Richard Koestner, and Richard M. Ryan. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 1999.

[16] Marilyn Rose McGee-Lennon, Maria Klara Wolters, and Stephen Brewster. User-centred multimodal reminders for assistive living. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2105–2114, New York, NY, USA, 2011. ACM.

[17] Phillippa Lally, Jane Wardle, and Benjamin Gardner. Experiences of habit formation: A qualitative study. *Psychology, Health & Medicine*, 16(4):484–489, 2011. PMID: 21749245.

[18] Phillippa Lally, Cornelia H. M. van Jaarsveld, Henry W. W. Potts, and Jane Wardle. How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, 40(6):998–1009, 2010.

[19] B Verplanken. Habits and implementation intentions. In *The ABC of behavioural change.*, pages 99–109. Elsevier, 2005.

[20] Peter M Gollwitzer. Implementation intentions: strong effects of simple plans. *American Psychologist*, 54(7):493, 1999.

[21] Katarzyna Stawarz, Marcela D Rodreguez, Anna L Cox, and Ann Blandford. Understanding the use of contextual cues: design implications for medication adherence technologies that support remembering. *Digital Health*, 2:2055207616678707, 2016.

[22] Marieke A Adriaanse, Peter M Gollwitzer, Denise TD De Ridder, John BF De Wit, and Floor M Kroese. Breaking habits with implementation intentions: A test of underlying processes, 2011.

[23] Mark A McDaniel and Gilles O Einstein. Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied cognitive psychology*, 14(7), 2000.

[24] LeiPhillip Wang, Fei Li, Dong Wang, Kun Xie, Deheng Wang, Xiaoming Shen, and JoeZ. Tsien. Receptors in dopaminergic neurons are crucial for habit learning. *Neuron*, 72(6):1055 – 1066, 2011.

[25] H. Joshua West, Cougar P. Hall, L. Carl Hanson, D. Michael Barnes, Christophe Giraud-Carrier, and James Barrett. There's an app for that: Content analysis of paid health and fitness apps. *J Med Internet Res*, 14(3), 2012.

[26] Logan T Cowan, Sarah Van Wagenen, Brittany A Brown, Riley Hedin, Yukiko Seino-Stephan, P Hall, and Joshua 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. 40, 09 2012.

[27] Leena Venta, Minna Isomursu, Aino Ahtinen, and Shruti Ramiah. "my phone is a part of my soul” - how people bond with their mobile phones. pages 311–317, 2008.

[28] Katarzyna Stawarz, Anna L. Cox, and Ann 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*, CHI '14, pages 2269–2278, New York, NY, USA, 2014. ACM.

[29] Paul Weiser, Dominik Bucher, Francesca Cellina, and Vanessa De Luca. A Taxonomy of Motivational Affordances for Meaningful Gamified and Persuasive Technologies. *3rd International Conference on ICT for Sustainability (ICT4S)*, 22:271–280, 2015.

[30] Frank Bentley and Konrad Tollmar. The power of mobile notifications to increase wellbeing logging behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1095–1098, New York, NY, USA, 2013. ACM.

[31] Katarzyna Stawarz. *Towards better medication adherence apps: Preventing forgetfulness by facilitating the formation of routine-based remembering strategies.* Doctoral thesis, UCL (University College London), 2017.

[32] Oussama Metatla, Nuno N. Correia, Fiore Martin, Nick Bryan-Kinns, and Tony 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*, CHI '16, pages 1055–1066, New York, NY, USA, 2016. ACM.

[33] Julie R. Williamson, Marilyn McGee-Lennon, and Stephen Brewster. Designing multimodal reminders for the home: Pairing content with presentation. ICMI '12, pages 445–448. ACM, 2012.

[34] Alexandre Stouffs. Interruptions as multimodal outputs: Which are the less disruptive. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, ICMI '02, pages 479–, Washington, DC, USA, 2002. IEEE Computer Society.

[35] 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.

[36] Rodrigo de Oliveira, Mauro Cherubini, and Nuria Oliver. Movipill: Improving medication compliance for elders using a mobile persuasive social game. *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, pages 251–260, 2010.

[37] Marilyn R. McGee-Lennon. Reminders that make sense: designing multisensory notifications for the home. *Journal of Assistive Technologies*, 6(2):93–104, 2012.

[38] Jan B.F. Van Erp and Hendrik A.H.C. Van Veen. Vibrotactile in-vehicle navigation system. *Transportation Research Part F: Traffic Psychology and Behaviour*, 7(4):247 – 256, 2004.

[39] Christopher D. Wickens, Juliana Goh, John Helleberg, William J. Horrey, and Donald A. Talleur. Attentional models of multitask pilot performance using advanced display technology. *Human Factors*, 45(3):360–380, 2003. PMID: 14702989.

[40] Jennifer L. Burke, Matthew S. Prewett, Ashley A. Gray, Liuquin Yang, Frederick R. B. Stilson, Michael D. Coovert, Linda R. Elliot, and Elizabeth 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*, ICMI '06, pages 108–117, New York, NY, USA, 2006. ACM.

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

[42] Mayrsohn BG Rowland JL Lyons EJ, Lewis ZH. Behavior change techniques implemented in electronic lifestyle activity monitors: A systematic content analysis. *J Med Internet Res*, (16(8)), 2014.

[43] Anthony Smith, Kristy de Salas, Benjamin Schüz, Stuart G Ferguson, and Ian Lewis. mhealth intervention design: Creating mhealth interventions for behaviour change. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, OzCHI '16, pages 531–536, New York, NY, USA, 2016. ACM.

[44] Susan Michie, Maartje M. van Stralen, and Robert West. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1):42, 2011.

[45] Joseph Weizenbaum. Eliza - a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45, 1966.

[46] Facebook. Quick replies. *Facebook for developers*, 2017.

[47] statista.

[48] Will Luton. *Free-to-Play: Making Money From Games You Give Away.* Pearson Education, 2013.

[49] Author. How to design outstanding feedback loops. *Smashing Magazine.*

[50] Mozilla. Audio html5 element, mdn web docs.

[51] Sunny Consolvo, Predrag Klasnja, David W. McDonald, and James A. Landay. Designing for healthy lifestyles: Design considerations for mobile technologies to encourage consumer health and wellness. *Foundations and Trends in HumanComputer Interaction*, 6(34):167–315.

[52] Bas Verplanken and Sheina Orbell. Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology*, 33(6):1313–1330, 2003.

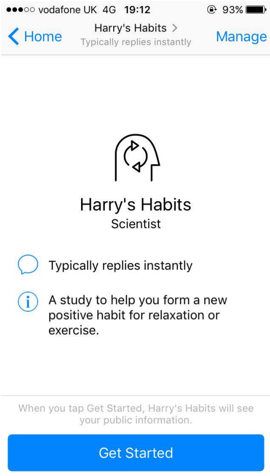
[53] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

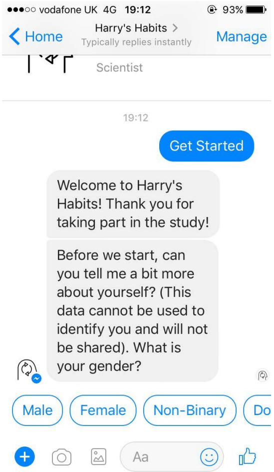
[54] Amos Tversky and Daniel Kahneman. Loss aversion in riskless choice: A reference-dependent model*. *The Quarterly Journal of Economics*, 106(4):1039–1061, 1991.

9 Appendix

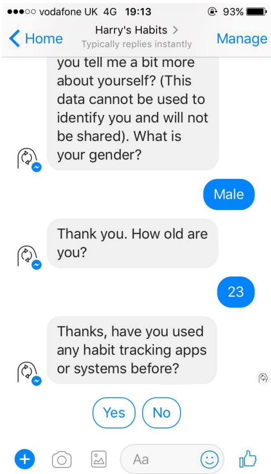
9.1 End of Study Questionnaire

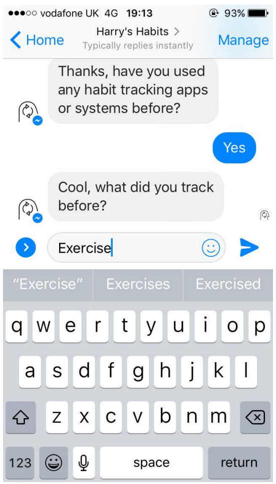
TODO CHANGE THIS

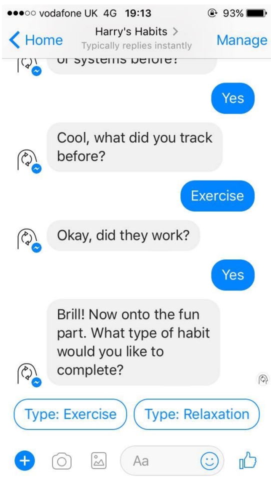

●●●○○ vodafone UK 4G 19:12 93%
Home
Harry's Habits
Typically replies instantly
Manage
Harry's Habits
Scientist
Typically replies instantly
A study to help you form a new positive habit for relaxation or exercise.
When you tap Get Started, Harry's Habits will see your public information.
Get Started

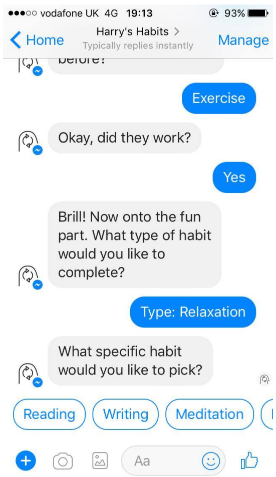

●●●○○ vodafone UK 4G 19:12 93%
Home
Harry's Habits
Typically replies instantly
Manage
Scientist
19:12
Get Started
Welcome to Harry's Habits! Thank you for taking part in the study!
Before we start, can you tell me a bit more about yourself? (This data cannot be used to identify you and will not be shared). What is your gender?
Male
Female
Non-Binary
Aa

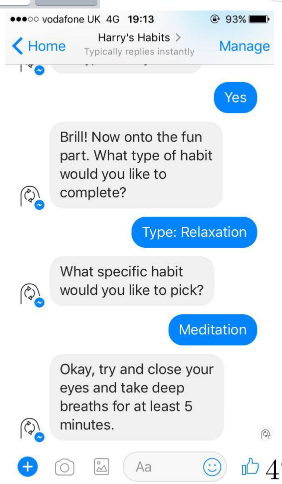

●●●○○ vodafone UK 4G 19:13 93%
Home
Harry's Habits
Typically replies instantly
Manage
identify you and will not be shared). What is your gender?
Male
Thank you. How old are you?
23
"23"
space
return

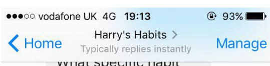

●●●○○ vodafone UK 4G 19:13 93%
Home
Harry's Habits
Typically replies instantly
Manage
you tell me a bit more about yourself? (This data cannot be used to identify you and will not be shared). What is your gender?
Male
Thank you. How old are you?
23
Thanks, have you used any habit tracking apps or systems before?
Yes
No
Aa

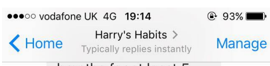

●●●○○ vodafone UK 4G 19:13 93%
Home
Harry's Habits
Typically replies instantly
Manage
Thanks, have you used any habit tracking apps or systems before?
Yes
Cool, what did you track before?
Exercise
"Exercise"
Exercises
Exercised
space
return

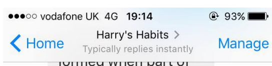

●●●○○ vodafone UK 4G 19:13 93%
Home
Harry's Habits
Typically replies instantly
Manage
Yes
Cool, what did you track before?
Exercise
Okay, did they work?
Yes
Brill! Now onto the fun part. What type of habit would you like to complete?
Type: Exercise
Type: Relaxation
Aa


●●●○○ vodafone UK 4G 19:13 93%
Home
Harry's Habits
Typically replies instantly
Manage
Exercise
Okay, did they work?
Yes
Brill! Now onto the fun part. What type of habit would you like to complete?
Type: Relaxation
What specific habit would you like to pick?
Reading
Writing
Meditation
Aa


●●●○○ vodafone UK 4G 19:13 93%
Home
Harry's Habits
Typically replies instantly
Manage
Yes
Brill! Now onto the fun part. What type of habit would you like to complete?
Type: Relaxation
What specific habit would you like to pick?
Meditation
Okay, try and close your eyes and take deep breaths for at least 5 minutes.
Aa


●●●○○ vodafone UK 4G 19:13 93%
Home
Harry's Habits
Typically replies instantly
Manage


●●●○○ vodafone UK 4G 19:14 93%
Home
Harry's Habits
Typically replies instantly
Manage


●●●○○ vodafone UK 4G 19:14 93%
Home
Harry's Habits
Typically replies instantly
Manage