

# Impact of Audio and Visual Display of Positive Reinforcement on Supporting Habit Formation

Leave Authors Anonymous  
for Submission  
City, Country  
e-mail address

Leave Authors Anonymous  
for Submission  
City, Country  
e-mail address

Leave Authors Anonymous  
for Submission  
City, Country  
e-mail address

## ABSTRACT

Habit formation technologies use rewards and points as means for providing positive reinforcement, often in visual or audio forms such as jingles, badges or animations. Providing appropriate rewards increases the chances for developing a new habit; yet, research on how these rewards should be delivered and the impact that this has on the process of habit formation is scarce. In this paper, we investigate how three types of positive reinforcement (audio, visual, audio-visual) influence habit completion and automaticity. We describe a 4-week study in which participants used a chatbot that delivered different types of rewards for completing a new daily habit. The results reported higher habit completion rates when a reward was present, especially for the audio-visual condition, without necessarily increasing behaviour automaticity. This has implications for the design of habit formation technologies that rely on audio and visual rewards as means of positive reinforcement.

## ACM Classification Keywords

H.5.m Information interfaces and presentation (e.g. HCI): Miscellaneous; H.5.2 User Interfaces: User-Centered Design

## Author Keywords

habit formation; behaviour change; positive reinforcement

## INTRODUCTION

Designing technology to support habit formation has become an important theme in HCI research. This is not surprising, as a good understanding of how to design systems that support habit development helps to ensure that they lead to long lasting changes in one's behaviour [21]. Rewards play an important part in supplementing the positive reinforcement necessary for habit forming processes [35]. An example of this in habit formation technologies is the use of coins and points as rewards to increase task repetition. However, studies have shown that this form of extrinsic rewards can also reduce motivation and therefore hinder habit formation in the long run [14].

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

CHI'18, April 21–26, 2018, Montreal, Canada

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 123-4567-24-567/08/06.

DOI: [http://dx.doi.org/10.475/123\\_4](http://dx.doi.org/10.475/123_4)

Other types of positive reinforcement include sending text messages to affirm behaviour [13] and coaching a person throughout the habit formation process [11]. But these approaches have been shown to encourage repetitive action and dependency, rather than automaticity of behaviour that is characteristic of habits [17]. An alternative approach is to provide intrinsic, rather than extrinsic, motivation after a habit is completed [5, 14]. Intrinsic motivation refers to behaviour that is driven by internal rewards that focus on increasing personal motivation to complete a task, such as health benefits, as opposed to, for example, monetary extrinsic rewards.

When delivered through technology, positive reinforcement aimed at facilitating intrinsic motivation is predominantly presented in a form of visual feedback, e.g. [3, 7, 25]. However, with the advent of ubiquitous and mobile technologies, visual feedback may not always be an appropriate form of presenting rewards, e.g. when a display is not visible or even not available. Alternative non-visual displays have been used to overcome these issues in a number of other applications (e.g. [8, 36, 28]), but the benefits of non-visual feedback as a form of positive reinforcement in the context of habit formation have received limited attention.

To explore the potential of non-visual feedback in supporting new habits, we conducted a 4-week situated study that explored the impact of different types of positive reinforcement (audio, visual, audio-visual) on the process of habit formation. The main contribution of our work is a better understanding of how these types of positive feedback affect the development of new habits. We show that combined audio-visual positive reinforcement is more effective at supporting repetition and habit completion compared to visual- or audio-only displays. However, it does not seem to affect habit automaticity. This has implications for the design of habit formation technologies that rely on audio and visual rewards as means of positive reinforcement.

## BACKGROUND

### Role of positive reinforcement in habit formation

Changing behaviour permanently requires forming a new habit [24], which in turn depends on consistent repetition [23]. The easier the action, the shorter the time before it turns into a habit. For example, tasks as simple as drinking water may require 18 days to become automatic, but for going to the gym that period may as well take 254 days [23]. To ensure this process is shorter, the new habit should be linked with a regularly occurring event. Building the habit around an existing routine

facilitates repetition [2], as the task that is already a part of daily routine and the context in which it takes place both serve as triggers to action [26]. Such event-based, contextual cues also make it easier to remember to repeat the task [37, 19].

Motivation further supports the process of habit formation. Positive reinforcement at the end of the task rewards the person by encouraging them to perform the action again, until it becomes habitual [5]. The reward increases intrinsic motivation by inducing the feeling of satisfaction [22] and increasing the person's perception of their performance [35], especially when the task itself is interesting [14]. Therefore, when designing behaviour change technology, the delivery of the reward must be carefully considered because it can affect the formation of new habits.

### Providing positive reinforcement with technology

Smartphone apps are often used to support the process of habit formation [34]. However, even though most of them are highly rated, they are not grounded in theory [33, 31]. Moreover, several studies suggest that users may become dependent on the app and the behaviour is not sustained after the app is removed [38, 31, 25], due to the lack of automaticity developed during the habit tracking process. Building habit automaticity, while avoiding developing dependency is therefore key, but requires the desired action to be built around an existing routine [23, 37]. Technology should help with routine creation by having additional checks to guard against changes in routine that check in with people about their habit if their situation changes [32]. While this could be facilitated by apps (e.g. see [33]), there are other possibilities.

Chatbots are becoming increasingly popular. They have been integrated into messaging apps, such as the *Telegram* app ([www.telegram.org](http://www.telegram.org)) to track food habits [20] and are seen as a significant component of HCI's increasing shift towards new interaction paradigms involving natural language user interfaces [16]. Applying the design guidelines for building habit formation systems [33] to a chatbot presents a new opportunity for delivering positive reinforcement.

Positive reinforcement rewards can be delivered via different modalities; audio, visual and audio-visual modes are common choices [10]. For example, audio feedback has been shown to improve task success when used as a notification for a task [27], but little research has looked at its impact on the completion of habits. On the other hand, fast visual feedback has been shown to be preferred over audio when the relative importance of the task is high [7] and to improve the consistency of a medication-taking habits [25]. However, when the visual feedback is removed, the performance can drop, which suggests a dependency on the technology and the visual feedback it provides. Although audio and visual feedback has been shown to improve task performance, the combination of both tends to be more effective than audio or visual alone [15, 39, 28]. For example, a meta-analysis of 43 multi-modal studies [9] revealed that a combined audio-visual modality was the most effective at increasing performance of a single task, compared with only visual or only audio feedback. However, care needs to be taken when adding an extra modality: this

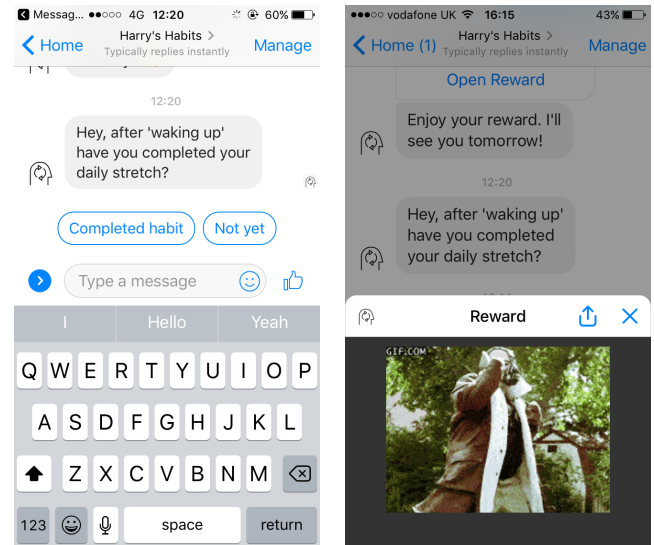


Figure 1. Bot's "check in" message (left) and an example of a reward (right).

study also showed that while visual feedback improved task performance, it reduced task quality [9].

Overall, there is evidence that these three types of feedback increase task completion in different ways. However, how – and whether – this type of feedback influences habit formation is unclear. Therefore, below we describe an in-the-wild study that explored how audio, visual and audio-visual positive reinforcement influences habit formation.

### COMPARING TYPES OF POSITIVE REINFORCEMENT

The aim of the study was to explore what type of positive reinforcement (audio, visual, audio-visual) would be the most effective in supporting regular habit completion and the development of automaticity. We conducted a 4-week situated study followed by optional semi-structured interviews to evaluate each mode of reward. Positive reinforcement was delivered by a chatbot.

### Method

For the purpose of the study, we developed a chatbot that was used to deliver positive reinforcement to participants wanting to start a new healthy habit. To ensure they were motivated to pursue their habit, we gave them a choice: they could select a physical activity habit (stretching, press ups, the plank) or a relaxing habit (reading, writing, meditation). We selected these specific habits, as they are easy to complete and generally do not require any special equipment. Moreover, simple tasks become automatic quicker than complex actions [23] and such simple habits allow to observe changes in automaticity of behaviour within four weeks [33]. The study was approved by the University Ethics Committee.

### Participants

Fifty-four participants were recruited on social networks. They were 18-63 years old (mean age=26.7 years old, SD=10); 32 (59%) were men, 14 (26%) were women, and eight (15%)

Condition	Number of Participants	Positive Reinforcement
Audio	14	15s audio
Visual	15	15s GIF
Audio-Visual	14	15s GIF and audio
None (control group)	11	Confirmation

**Table 1. Study conditions and their corresponding types of positive reinforcement.**

chose not to disclose their gender; 16 (29%) were university students. As their habit of choice, 15 participants selected meditation, 12 selected stretching, 11 selected press ups, six selected reading, five selected writing and other five selected the plank. In terms of interactions with the bot, 21 participants used a web browser, 18 used iOS devices, 12 used Android and three used another mobile device.

### Design

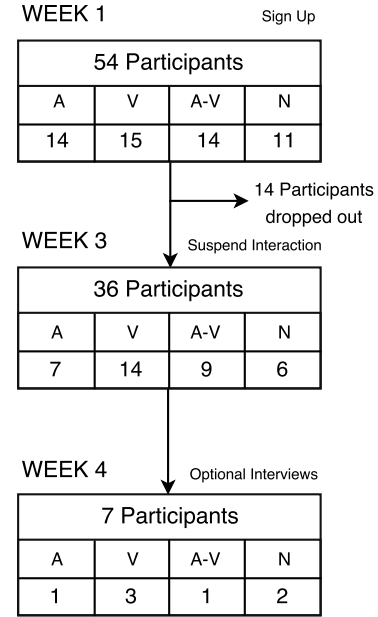
Participants were randomly assigned to one of four conditions: Audio, Visual, Audio-Visual and the Control group. Each condition differed by the type of positive reinforcement it provided after participants reported completion of their daily habit: participants in the Audio condition received a fifteen-second audio clip as a reward for completing their habit; in the Visual condition they received a fifteen-second GIF; the Audio-Visual condition combined the audio clip with a GIF; and participants in the Control group received a confirmation message (“Thanks.”). Study conditions are summarised in Table 1.

For each condition, we measured the number of times the selected habit was completed during the study. “Habit completion rate” was defined as the number of times the person repeated and reported the behaviour, divided by 21 days during which participants interacted with the bot. We also measured the change in automaticity of the behaviour during the final weeks, to assess habit strength.

### Materials

To deliver different types of positive reinforcement, we built a custom Facebook Messenger bot (see Figure 1). Bot’s functionality and approach to support habit formation was informed by [33]: rather than providing reminders, the bot first allowed participants to define a routine and during the study provided “check in” notifications to see if the routine was followed; if the routine was not working and participants did not complete their habit on time, the bot then suggested changes to the plan. For example, if a participant regularly snoozed the check in notification, the bot would ask if they wanted to move the check in time for later in the day.

We decided to use a bot as it can easily send notifications, is cross-platform and thus highly available, and also offers simple user interface and easy interactions already familiar to the users. Moreover, the integration with the platform meant that interactions with the bot would be visible alongside participant’s other conversations, making it easier to report completion of the habit [6].



**Figure 2. Participant drop-out during the four-week study. A: audio, V: visual, A-V: audio-visual, N: no reward.**

As the aim of the rewards was to motivate the participants to keep coming back every day and completing their habits, we identified several “motivational” GIFs and audio files, and each GIF file was tweaked to match the audio frequency. We decided to use popular GIFs with internet memes, i.e. content that is passed along from person to person via social media posts [12]. Humorous GIF memes make people laugh and are popular on some social media sites, where they usually are the most engaging content [4]. Given that our bot was integrated with Facebook, and that Facebook introduced built-in support for animated GIFs [1], memes would integrate well with the bot environment.

We also used the bot to collect responses to the Self-Report Behavioural Automaticity Index (SRBAI; [18]). This is a validated four-item instrument for measuring habit automaticity, which can be used as an indicator of habit strength.

### Procedure

Recruitment posts instructed potential participants to connect with our bot via Facebook Messenger. By connecting with the bot, completing the set-up and providing demographic information, participants expressed their consent to participate in the study.

During the set-up process, participants were asked to select one of the healthy habits they would like to develop and, as recommended in [32] for technologies supporting habit formation, to specify contextual cues associated with that habit and the time they would want to complete it (morning, afternoon,

evening). Once their habit routine was specified, participants were automatically assigned to one of the study conditions.

Next, for the first three weeks of the study, participants were asked to complete their chosen habit as part of the specified routine. During that period, the bot “checked in” with the participants after their specified time (see Figure 1, left). Participants then could indicate whether they completed the habit or not. If they reported habit completion, they would receive their positive reinforcement message (see Figure 1, right). If they selected “not yet”, the bot would check on the participant again an hour later. This allowed for the checks to be snoozed, to ensure the new habit fits into participant’s routine. If participants constantly told the bot they had not completed their habit yet, the bot would suggest making changes to the routine. Throughout the study, information about how well each participant was performing was not revealed, as streaks can provide extrinsic motivation [31], which might have confounded our results.

After interacting with the bot for three weeks, participants were asked to complete the SRBAI questionnaire and answer a set of questions about the positive reinforcement they received (e.g. how useful it was, whether they liked it). Next, during the final week of the study, bot interactions were suspended, but participants were still asked to continue with their habit. At the end of that period, participants were asked to complete the SRBAI questionnaire again along with a few questions about their experience with the rewards.

Participants also had an option to opt-in for an interview to discuss their experience. The interviews were conducted after the 4-week study and explored participants experience with trying to maintain their habit after the interactions with the bot ended, their attitudes towards positive reinforcement and towards the bot itself.

## RESULTS

Thirty-six participants completed the study (see Figure 2); seven of them were from the Audio condition, 14 from the Visual condition, nine were from Audio-Visual and six from the Control group. They were 18-63 years old (mean=27 years old, SD=12); 23 (64%) were men, 11 (30%) were women and two (6%) didn’t disclose their gender.

Among participants who completed the study, meditation was the most popular habit (12 participants, or 33%), followed by press ups (eight participants, or 22%), then stretching (six participants, or 15%). Reading and writing were the least common, only selected by four (11%) and two (5%) participants, respectively.

### Habit Completion

Overall, completion rates varied from 5% to 100% across individual participants (mean=30%). The average habit completion rate was the highest for the Audio-Visual condition, with 33%, followed by 29% for the Visual condition, 28% for the control condition and the lowest for Audio, with 14%. Figure 3 shows the average completion rate on a weekly basis for all conditions.

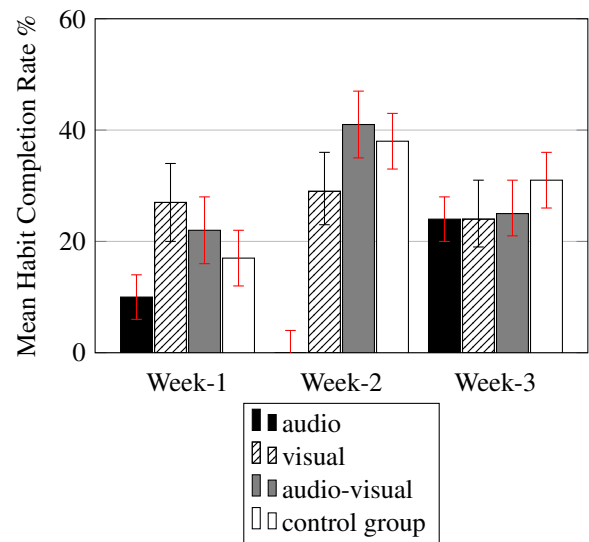


Figure 3. The effect of each condition on habit completion.

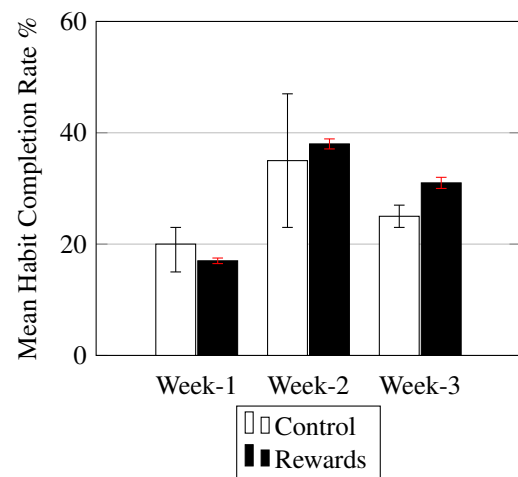


Figure 4. The rewards effect on habit completion.

#### The effect of each condition on habit completion

To investigate the levels of habit completion per condition, we conducted a one-way ANOVA across each week. The analysis (see Figure 3) revealed that there was a statistically significant difference between each reward ( $F(2,624) = 27.007, p < 0.05$ ). A Tukey post hoc test revealed that the Audio ( $0.081 + 0.274, p < 0.005$ ), Control ( $0.182 + 0.386, p < 0.005$ ) and Visual ( $0.267 + 0.484, p = 0.025$ ) feedback conditions were significantly lower when compared to the Audio-Visual ( $0.373 + 0.484, p < 0.005$ ) condition.

#### The rewards effect on habit completion

To explore the effect of rewards on the number of habits completed, compared with the control group, a one-way between-groups ANOVA with planned comparisons was conducted (see Figure 4). Participants were divided into two groups: group one with rewards, group two without rewards. There was a statistically significant difference at the  $p < 0.005$  level in both groups for week one and three: week one,  $F(1,23.20) = 9.48, p = 0.005$ , week two,  $F(1,33.35) = 4.46, p = 0.42$  and week three,  $F(1,0) = 17.01, p \leq 0.005$ . The effect size for week one, week two and week three are large, calculated using  $\eta^2$ , were 0.25, 0.39 and 0.43 respectively, reporting a large difference in the mean scores.

#### Multiple vs. single-mode feedback on habit completion

The results from our study report a significant increase in habit completion for multiple modes (audio-visual). The significance of this finding was tested by conducting a one-way between-groups ANOVA with planned comparisons (see Figure 5). Participants were divided into two groups according to their mode. Group one: audio rewards and visual rewards, group two: audio-visual rewards. There was a statistically significant difference between groups each week ( $p < 0.005$ ) week one,  $F(1,50) = 0.69, p = 0.410$ , week two,  $F(1,50) = 23.04, p < 0.005$  and week three,  $F(1,50) = 8.85$ . The effect size for week one, week two and week three are large, calculated using  $\eta^2$ , were 0.25, 0.39 and 0.43 respectively, reporting an increasing difference in the mean scores throughout weeks.

#### Habit Automaticity

Participants completed habit automaticity questionnaires (SRBAI) at the end of week three and week four. Eleven participants completed both SRBAI questionnaires: two from the Audio condition, five from Visual, two from Audio-Visual and two from the control group.

#### The effect of each condition on habit automaticity

There was no statistically significant difference between scores as determined by a one-way ANOVA (see Figure 6) for SRBAI one ( $F(3,7) = 0.333, p = 0.802$ ) and SRBAI two ( $F(3,7) = 0.639, p = 0.614$ ). For each reward specifically a Tukey post hoc test revealed that there was no significant difference in scores for each reward.

#### The effect of rewards on habit automaticity

Comparing participants SRBAI results' with the control group shows the impact rewards had on automaticity (see Figure 7).

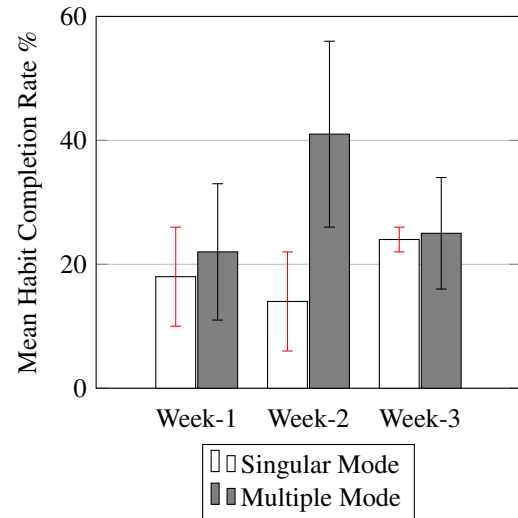


Figure 5. Multiple vs. single-mode feedback on habit completion.

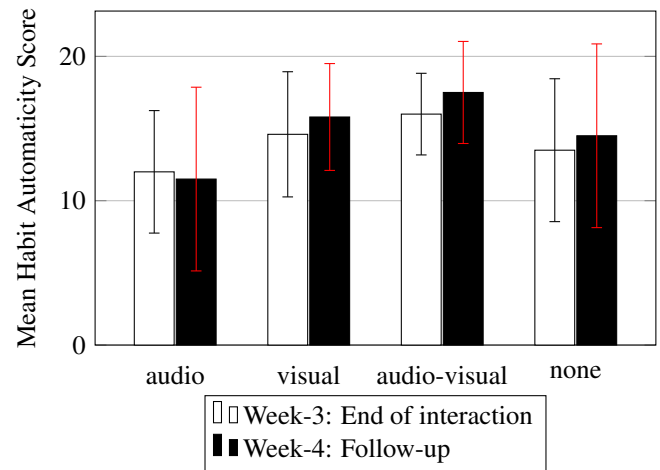


Figure 6. The effect of each condition on habit automaticity.

A one-way between-groups ANOVA with planned comparisons was conducted, but there was not a statistically significant difference for the two groups at SRBAI one:  $F(1,9) = 0.02, p = 0.88$ , and SRBAI two:  $F(1,9) = 0.07, p = 0.78$ . The difference in mean scores between the groups had, at SRBAI one: a medium effect with an effect size  $\eta^2 = 0.11$ , and at SRBAI two: a large effect, with an effect size of  $\eta^2 = 0.17$ .

#### Multiple vs. single-mode feedback on habit automaticity

To compare the effect audio-visual had on automaticity a one-way between-groups ANOVA with planned comparisons was conducted (see Figure 8). Participants were divided into two groups according to their reward condition. Group one: audio rewards and visual rewards and group two: audio-visual rewards. There was not a statistically significant difference for the two groups at SRBAI one:  $F(1,9) = 1.04, p = 0.33$ , and SRBAI two:  $F(1,9) = 0.64, p = 0.44$ . The difference in mean scores between the groups had, at SRBAI one: a medium



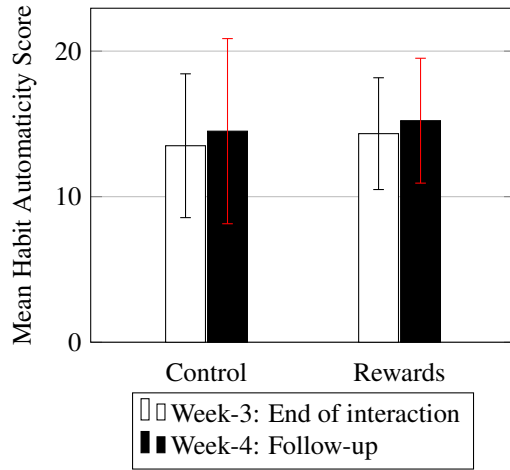


Figure 7. The effect of rewards on habit automaticity.

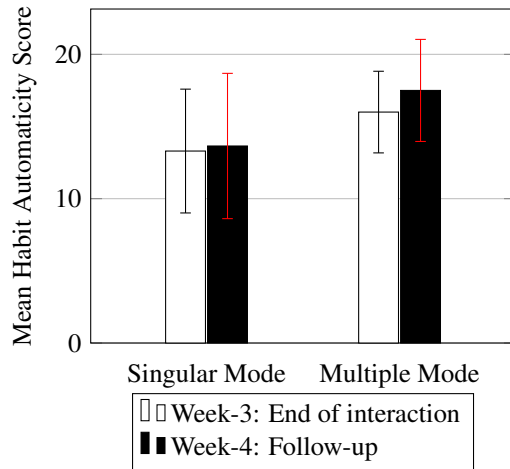


Figure 8. Multiple vs. single-mode feedback on habit automaticity.

effect with an effect size of  $\eta^2 = 0.11$ , and at SRBAI two: a large effect, with an effect size of  $\eta^2 = 0.17$ .

### Interview Feedback

Seven participants agreed to participate in optional interviews (their details are summarised in Table 2). The interviews were recorded, transcribed and grouped into three themes: continuation with habit formation, attitudes towards rewards and attitudes towards the chatbot.

#### Continuation with habit completion

When discussing the experiences with habit formation, we asked participants for their reasons of selecting their habits. P4 and P5 admitted choosing their habit because they had wanted to start that particular habit for a long time. P2 said that it was 'not too much effort' and P6 said it was 'something successful people do'. P7 wanted 'to be more active' P2 wanted a habit that was 'less time consuming'. Finally, P1, P5 and P6 wanted to 'relieve stress'.

In terms of habit completion, while most participants were able to do it regularly, P6 reported always putting the message off and eventually their habit completion got 'worse and worse' until they stopped altogether. After the bot was removed, the same was true for all other interviewed participants, who admitted finding it difficult to continue with their habit. They 'kept forgetting' (P3), found it 'harder to remember' (P1) and lacked motivation, not performing the action if it had 'been a long day' (P5). P5 also tried to do it 'every now and again', but usually they would only complete it if 'they remembered'. P1, P4, P6, and P7 all discussed not being able to remember to complete their habit after the bot was removed. This reveals the dependency between technology and habits, suggesting that the bot did not increase habit automaticity, or that the existing routine participants chose was not suitable for new habits, or that they were not given enough time to develop automaticity.

#### Attitudes Towards Rewards

Participants had mixed feelings about the rewards. Some 'did not like the [visual] rewards' (P1), skipping over them after the first few, they 'just wanted to get rid of the notification dot' (P1). Another participant (P3) said 'some of them [audio-visual rewards] were funny', but they did not like them overall and mentioned the audio rewards were 'too random'.

One participant (P5) thought the rewards did not give them an incentive towards their habit, just a 'nice little extra' and they 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 audio rewards at certain times of the day. Finally, an upbeat participant (P7) talked about 'always wanting to open them' and 'the combination was perfect'. However, they said they also found them 'repetitive' (P7) and one participant stated that they simply wanted to get rid of the notification (P1). This implies that the bot was not tracking how many habits that participant completed, but instead how many times wanted the bot to stop messaging them.

#### Attitudes Towards the Chatbot

Participants were asked about the chatbot as the method of interaction. They found the method 'pretty good' (P2), they 'liked it' (P3) and 'would have liked more interaction' (P4). Some suggested expanding the bot with additional features, such as 'help and support throughout' (P4), 'ideas on how to improve your habit' (P4) and 'advice on how to set aside

	Age	Gender	Condition	Habit	Completion rate
P1	24	M	Visual	Meditation	86%
P2	19	F	Auditory	Stretch	43%
P3	28	F	Visual	Writing	38%
P4	57	F	No reward	Meditation	80%
P5	25	M	Visual	Meditation	19%
P6	21	F	No reward	Meditation	33%
P7	32	M	Visual-Auditory	Stretch	90%

Table 2. Details of participants who were interviewed.

*time for your habit*' (P4). Others were neutral, some expecting *'different messages, such as Hey [name], a bit more care about the person, a bit less like a robot'* (P7).

Four participants (P2, P4, P6, P7) enjoyed the check-in notification aspect, but one found it *'repetitive'* (P7) and *'got annoying if I pressed Not Yet'* (P5). P1 and P3 wanted to see their progress as they tracked their habits, and talked about wanting to reflect on their data. They mentioned that they would feel *'more encouraged to keep doing it, rather than random music [audio rewards]'* (P3).

Six out of the seven participants wanted the prototype to come back with a few modifications (P1 was neutral): *'enclosed with Fitbit so it is all in a single place'* (P2), *'fine without rewards'* (P1, P3), *'more interaction'* (P4) and *'with statistics about my progress'* (P1, P3). Participants wanted the bot as more of a *'constant persistent reminder'* (P6) with additional tracking elements to remind them to perform their habit to fit into their busy schedule.

Five participants mentioned *Headspace* — the popular meditation app ([www.headspace.com](http://www.headspace.com)), saying that they wanted a combination of the bot and meditation. It prompted one participant to download the Headspace app. Overall, participants wanted the bot to keep track of their habit and they would use the Headspace app to help them perform their meditation.

Participants also gave additional, unexpected feedback by simply messaging the bot. They asked inquisitive questions, such as *'what kind of thing are you looking to find out'* (P8), or provided feedback: *'this is not working for me'* (P9). Others simply sent *'stop'* (P10) — to try and stop the daily messages (the participant then blocked the bot).

Negative comments towards the rewards and the 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'* (P11), and another said *'never message me'* (P12). Yet another posted that this was the *'lame same band'* (P13) after receiving an audio reward.

## Summary

Our results showed a statistically significant drop in habit completion for the control group without rewards. The findings suggest that participants are more likely to complete their habit if given one of the rewards, especially if the reward is audio-visual. Participants with audio-visual rewards also had higher habit automaticity scores compared with only audio or only visual rewards; however, this difference was not statistically significant. Finally, participants picked habits they wanted to perform, but when the bot stopped notifying them, they lacked motivation to continue. Even though participants enjoyed the rewards at the beginning, this passed and most participants reported disliking them after the first week.

## DISCUSSION

This research aimed to understand how different types of positive reinforcement — audio, visual and audio-visual — influence the process of habit formation, especially how they facilitate regular habit completion. We found that participants who received bot-delivered rewards completed more habits than

the control group without rewards, which suggests that positive reinforcement delivered by a chatbot may have a better effect on habit completion than SMS reminders: in [33] the control group performed better than the group with positive reinforcement delivered by text messages.

There was also an increase in completed habits for participants with combined modalities when compared with the other reward modes. This matches existing literature for using multiple modalities to improve task completion rates [9, 15, 39, 30, 29]. Although during the interviews participants reported a drop in habit completion after removing the prototype, this is likely due to the lack of habit automaticity developed during the three week period [31, 25]. Developing automaticity would enable the continuation of habit completion after one week without the prototype [33].

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. A dependency could have been developed with the bot which would hinder habit automaticity [33]. Participants discussed negative feelings towards rewards during the interviews, however, participants attitudes towards interaction with the chatbot was mostly positive. All interviewed participants suggested features they would like, likely because the bot did not meet their needs. Moreover, each participant seemed to have mixed feelings towards the rewards. However, although the rewards did not significantly increase habit automaticity, they did improve the number of habits participants completed. Therefore, we can conclude that the chatbot has a potential as a tool for supporting habit formation.

The prototype was based on the existing literature, i.e. it helped participants stick to a routine by reminding them of their routine and using positive reinforcement [35] and back up notifications [33]. However, because of the nature of interactions, participants might have treated the bot as a regular reminder. Therefore, the bot may have simply reinforced repetitive actions rather than facilitating the development of automaticity [11].

## Limitations and Future Work

This research has a few key limitations that open up future work avenues. 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. Nevertheless, our findings already suggest a significant difference between combined modalities and singular to improve habit completion, which matches previous studies [15, 39, 3] that compared of audio and visual modalities.

Second, the measurement of habit formation past the lifetime of the bot relied on the interviews. But only seven participants responded to the follow up interviews and their discussions relied on participants recall, which could be inaccurate. Future work conducting similar research with a larger sample size for the SRBAI questionnaire would validate the effect these rewards had on habit formation without the prototype.

Third, this research relies on participants self-reporting and some of them could have lied to remove the alert instead of performing the habit, just to get the reward. This is particularly true with the snooze function, as the voluntary interviews showed that some participants found the bot annoying and stopped using it over time. Therefore, it is difficult to draw any valid conclusion on actual habit performance. Future work into how quickly participants responded to the alerts and if the device delivery (browser or smartphone) affected these findings would allow us to better understand how participants interacted with the alerts.

Finally, future work could focus on comparing bot-delivered positive reinforcement to more traditional approaches, such as app notifications or SMS. With their ease of implementation and access to a large pool of participants, chatbots are a promising platform for delivering behaviour change interventions. Further research into the effect of bot interactions on habit formation is needed to explore the full potential of this approach.

## CONCLUSIONS

We found that the type of reward did not influenced habit automaticity, but it nevertheless increased throughout the study. However, the rewards did improve the number of habits people completed, with the highest completion rates in the audio-visual condition, which suggests that bots may play a useful, albeit limited, role in supporting new behaviours.

## ACKNOWLEDGEMENTS

[Removed for peer review.]

## REFERENCES

1. 2017. Celebrating 30 Years of the GIF at Facebook Newsroom. <https://newsroom.fb.com/news/2017/06/celebrating-30-years-of-the-gif/>. Accessed: 2017-09-12.
2. Marieke A Adriaanse, Peter M Gollwitzer, Denise TD De Ridder, John BF De Wit, and Floor M Kroese. 2011. Breaking habits with implementation intentions: A test of underlying processes. *Personality and Social Psychology Bulletin* 37, 4, 502–513.
3. Motoyuki Akamatsu, I. Scott Mackenzie, and Thierry Hasbroucq. 1995. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics* 38, 4, 816–827. DOI: <http://dx.doi.org/10.1080/00140139508925152>
4. Saeideh Bakhshi, David A. Shamma, Lyndon Kennedy, Yale Song, Paloma de Juan, and Joseph 'Jofish' Kaye. 2016. Fast, Cheap, and Good: Why Animated GIFs Engage Us. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 575–586. DOI: <http://dx.doi.org/10.1145/2858036.2858532>
5. A. Bandura. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev* 84, 2, 191–215.
6. Frank Bentley and Konrad Tollmar. 2013. The Power of Mobile Notifications to Increase Wellbeing Logging Behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, 1095–1098. DOI: <http://dx.doi.org/10.1145/2470654.2466140>
7. Duncan P. Brumby, Samantha C.E. Davies, Christian P. Janssen, and Justin J. Grace. 2011. Fast or Safe?: How Performance Objectives Determine Modality Output Choices While Interacting on the Move. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, 473–482. DOI: <http://dx.doi.org/10.1145/1978942.1979009>
8. Jennifer L Burke, Matthew S Prewett, Ashley A Gray, Liuquin Yang, Frederick RB Stilson, Michael D Coovert, Linda R Elliot, and Elizabeth Redden. 2006a. 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*. ACM, 108–117.
9. Jennifer L. Burke, Matthew S. Prewett, Ashley A. Gray, Liuquin Yang, Frederick R. B. Stilson, Michael D. Coovert, Linda R. Elliot, and Elizabeth Redden. 2006b. 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)*. ACM, 108–117. DOI: <http://dx.doi.org/10.1145/1180995.1181017>
10. Fang Chen. 2005. *Designing Human Interface in Speech Technology*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
11. Catalina Danis, Katherine Vogt, Justin D. Weisz, Yu Ma, and Russell Olsen. 2017. 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)*. ACM, 2480–2486. DOI: <http://dx.doi.org/10.1145/3027063.3053206>
12. R. Dawkins. 2006. *The Selfish Gene: 30th Anniversary Edition*. OUP Oxford. <https://books.google.co.uk/books?id=go0e5sBRznYC>
13. Roelof A.J. de Vries, Khiet P. Truong, Sigrid Kwint, Constance H.C. Drossaert, and Vanessa Evers. 2016. 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)*. ACM, 297–308. DOI: <http://dx.doi.org/10.1145/2858036.2858229>
14. Edward L. Deci, Richard Koestner, and Richard M. Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125, 6. DOI: <http://dx.doi.org/10.1037/0033-2909.125.6.627>



15. Jan B.F. Van Erp and Hendrik A.H.C. Van Veen. 2004. Vibrotactile in-vehicle navigation system. *Transportation Research Part F: Traffic Psychology and Behaviour* 7, 4, 247 – 256. DOI : <http://dx.doi.org/10.1016/j.trf.2004.09.003>
16. Asbjørn Følstad and Petter Bae Brandtzæg. 2017. Chatbots and the New World of HCI. *interactions* 24, 4, 38–42. DOI : <http://dx.doi.org/10.1145/3085558>
17. Benjamin Gardner. 2012. Habit as automaticity, not frequency. 14, 32–36.
18. Benjamin Gardner, Charles Abraham, Phillippa Lally, and Gert-Jan de Bruijn. 2012. Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity* 9, 1, 102. DOI : <http://dx.doi.org/10.1186/1479-5868-9-102>
19. Peter M Gollwitzer. 1999. Implementation intentions: strong effects of simple plans. *American Psychologist* 54, 7, 493.
20. Bettina Graf, Maike Krüger, Felix Müller, Alexander Ruhland, and Andrea Zech. 2015. Nombot: Simplify Food Tracking. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia (MUM '15)*. ACM, 360–363. DOI : <http://dx.doi.org/10.1145/2836041.2841208>
21. Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. 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)*. ACM, New York, NY, USA, 3063–3072. DOI : <http://dx.doi.org/10.1145/1978942.1979396>
22. Phillippa Lally and Benjamin Gardner. 2013. Promoting habit formation. *Health Psychology Review* 7, sup1, S137–S158. DOI : <http://dx.doi.org/10.1080/17437199.2011.603640>
23. Phillippa Lally, Cornelia H. M. van Jaarsveld, Henry W. W. Potts, and Jane Wardle. 2010. How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology* 40, 6, 998–1009. DOI : <http://dx.doi.org/10.1002/ejsp.674>
24. Phillippa Lally, Jane Wardle, and Benjamin Gardner. 2011. Experiences of habit formation: A qualitative study. *Psychology, Health & Medicine* 16, 4, 484–489. DOI : <http://dx.doi.org/10.1080/13548506.2011.555774>
25. Matthew L. Lee and Anind K. Dey. 2014. Real-time Feedback for Improving Medication Taking. In *Proceedings of the 32<sup>nd</sup> Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 2259–2268. DOI : <http://dx.doi.org/10.1145/2556288.2557210>
26. Mark A McDaniel and Gilles O Einstein. 2000. Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied cognitive psychology* 14, 7.
27. Marilyn R. McGee-Lennon. 2012. Reminders that make sense: designing multisensory notifications for the home. *Journal of Assistive Technologies* 6, 2, 93–104. DOI : <http://dx.doi.org/10.1108/17549451211234957>
28. Oussama Metatla, Nuno N. Correia, Fiore Martin, Nick Bryan-Kinns, and Tony Stockman. 2016a. 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)*. ACM, 1055–1066. DOI : <http://dx.doi.org/10.1145/2858036.2858456>
29. Oussama Metatla, Nuno N. Correia, Fiore Martin, Nick Bryan-Kinns, and Tony Stockman. 2016b. 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)*. ACM, 1055–1066. DOI : <http://dx.doi.org/10.1145/2858036.2858456>
30. Sharon Oviatt, Rachel Coulston, and Rebecca Lunsford. 2004. When Do We Interact Multimodally?: Cognitive Load and Multimodal Communication Patterns. In *Proceedings of the 6th International Conference on Multimodal Interfaces (ICMI '04)*. ACM, 129–136. DOI : <http://dx.doi.org/10.1145/1027933.1027957>
31. Ian Renfree, Daniel Harrison, Paul Marshall, Katarzyna Stawarz, and Anna Cox. 2016. Don't Kick the Habit: The Role of Dependency in Habit Formation Apps. 2932–2939. DOI : <http://dx.doi.org/10.1145/2851581.2892495>
32. Katarzyna Stawarz, Anna L. Cox, and Ann Blandford. 2014. Don't Forget Your Pill!: Designing Effective Medication Reminder Apps That Support Users' Daily Routines. In *Proceedings of the 32<sup>nd</sup> Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 2269–2278. DOI : <http://dx.doi.org/10.1145/2556288.2557079>
33. Katarzyna Stawarz, Anna L. Cox, and Ann Blandford. 2015. Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation. In *Proceedings of the 33<sup>rd</sup> Annual ACM Conference on Human Factors in Computing Systems. Proceedings of the 2015 Conference on Human Factors in Computing Systems (CHI-2015)*, 2653–2662. DOI : <http://dx.doi.org/10.1145/2702123.2702230>
34. Logan T Cowan, Sarah Van Wagenen, Brittany A Brown, Riley Hedin, Yukiko Seino-Stephan, P Hall, and Joshua H West. 2012. Apps of Steel: Are Exercise Apps Providing Consumers With Realistic Expectations. A Content Analysis of Exercise Apps for Presence of Behavior Change Theory. 40.
35. Robert J. Vallerand and Greg Reid. 1984. On the Causal Effects of Perceived Competence on Intrinsic Motivation: A Test of Cognitive Evaluation Theory. *Journal of Sport Psychology* 6, 1, 94–102. DOI : <http://dx.doi.org/10.1123/jsp.6.1.94>

36. Yolanda Vazquez-Alvarez, Ian Oakley, and Stephen A Brewster. 2012. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous computing* 16, 8, 987–999.
37. B Verplanken. 2005. Habits and implementation intentions. In *The ABC of behavioural change*. Elsevier, 99–109. <http://opus.bath.ac.uk/9438/>
38. H. Joshua West, Cougar P. Hall, L. Carl Hanson, D. Michael Barnes, Christophe Giraud-Carrier, and James Barrett. 2012. There's an App for That: Content Analysis of Paid Health and Fitness Apps. *J Med Internet Res* 14, 3. DOI:<http://dx.doi.org/10.2196/jmir.1977>
39. Christopher D. Wickens, Juliana Goh, John Helleberg, William J. Horrey, and Donald A. Talleur. 2003. Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. *Human Factors* 45, 3, 360–380. DOI:<http://dx.doi.org/10.1518/hfes.45.3.360.27250>