From personalized timely notification to healthy habit formation: A feasibility study of reinforcement learning approaches on synthetic data

Aneta Lisowska[1], Szymon Wilk[1] and Mor Peleg[2]

[1]*Institute of Computing Science, Poznań University of Technology, Poznań, Poland*
[2]*Department of Information Systems, University of Haifa, Haifa, Israel*

Abstract

Cancer patients may struggle with mental wellbeing issues such as distress and depression. As a part of the CAPABLE project, we aim to develop a digital behaviour-change intervention that helps them build positive health habits and improve their wellbeing. The main challenge to the evaluation of the system is the lack of access to real data prior to intervention start. Therefore, first, we created a simulator that mimics patient responses to activity suggestions based on Fogg's behaviour model. Later we used supervised and reinforcement learning methods to learn the best time of sending the patient prompts. We found that the reinforcement learning methods learn quickly not to over-notify patients and find prompt policies that are more effective in facilitating users in performing target activity than a random notification strategy, but are less effective than adaptive supervised learning method trained to predict patient responsiveness.

Keywords

Fogg behaviour model, reinforcement learning, digital behaviour change intervention

1. Introduction

Cancer patients may struggle with mental well-being issues such as distress and depression at all stages of their cancer journey [1, 2]. Poor mental health not only impacts their quality of life but also reduces treatment adherence and cancer survival [3].

As a part of the Horizon 2020 CAncer PAtient Better Life Experience (CAPABLE) project, we aim to develop a digital behaviour change intervention [4] that could help cancer patients build emotional resilience and form positive health habits. The patients will be equipped with a coaching system (Virtual Coach), which contains three components: backend (with most of the logic and processing), mobile health app (for functions that need to be performed in proximity to the patient, such as supporting, and communicating) and a smartwatch (for sensing). The Virtual Coach (VC) will have the capacity to interact with the patient through notifications and will suggest multiple activities from the domain of mindfulness and positive psychology.

In this first proof of concept work, we focus on stimulating the development of a daily

walking habit. We chose this activity for initial exploration because of its evidence-based benefits to psychological wellbeing [5] and to the overall health of cancer survivors [6, 7]. We draw inspiration from Fogg's Behaviour Model [8] (see Section 2.1) to simulate the patients' responses to the notification (**trigger**) sent by the VC (reminding the patients to perform the daily activity), depending on the patient's **motivation** and **ability**; these three components are the cornerstones of Fogg's theory. The responsiveness of the patients to the notifications depends also on the context e.g., the time of the day [9], location of the patient [10] and physiological state of the patient [11] (e.g. stress level).

Previously, researchers investigated the possibility of using reinforcement learning (RL) for the identification of the appropriate time to send notifications in order to optimize user engagement with mobile applications [12]. Our goal is to facilitate patient habit formation through timely notification. We investigate the use of supervised and reinforcement learning approaches to learning the best pattern of interacting with patients through notifications (see Section 3). A starting point is constructing a simulator that would allow us to test learning-based techniques without access to real-life data.

Thus, in this feasibility study we:

- Propose a simulator that mimics patients' responsiveness to activity suggestions, based on Fogg's behaviour model and on findings from multiple health intervention studies and mobile notification research.
- Apply supervised and reinforcement learning approaches to learn the best time of sending the patient notifications in order to achieve positive habit formation and compare the patient's responsiveness to prompts against that achieved with randomly timed notifications.

2. Related work

2.1. Modeling behaviour

According to Fogg, habit formation depends on three components: the person's motivation, her ability to perform the task, and the existence of a trigger reminding her to perform the target behaviour [13]. Fogg expresses this dependency as "B=MAT" [8] and suggests that **behaviour** occurs when the person's level of motivation and **ability** and the existence of the trigger place her above the action threshold (i.e., performing the behaviour), which we here interpret as:

$$Behaviour = \begin{cases} 1 & \text{if } (Motivation \times Ability \times Trigger) > \text{action threshold} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Each of the three components may in turn be influenced by a range of internal (personal) and external factors. We describe them below in the context of our digital behaviour change intervention system.

Motivation. Jowsey *et al.* [14] interviewed Australian patients with chronic illness to gain an understanding of what motivates them to engage in self-management behaviours. The authors report that maintaining a positive attitude (i.e., positive emotional valence, *Valence*) was one of

the most important internal factors motivating patients to control their health. The external factor impacting motivation was the presence of family and friends (*Family*). On the other hand, the patients got demotivated when they were perceiving self-management behaviour as having limited benefit (*Perceived Benefit*).

Interestingly, sleepiness might decrease motivation for behaviours that are not oriented toward assuring sufficient sleep [15]. Conversely, sufficient sleep (*Sufficient Sleep*) might have a positive effect on motivation, for example Dolsen *et al.* suggests that improving sleep benefits patient adherence to treatment [16].

In our proof of concept behaviour simulator, the above mentioned factors have an equal contribution, although we acknowledge that this is a very simplified view of motivation:

$$Motivation = Valence + Family + Perceived\ Benefit + Sufficient\ Sleep \qquad (2)$$

We adopt the same simple cumulative factors representation for the remaining components of behaviour. Note that for behaviour equation multiplication is used because it is enough that a single component is 0 and the target behaviour will be not performed. We assume this is not true for the computation of the components, e.g., a person might currently experience negative valence, but the remaining motivation factors might be present and hence motivation should not be nullified.

Ability. Chan *et al.* found that app users are more receptive to respond to memory-training suggestions when they are under low cognitive load [17]. High cognitive load might suggest that the person pays attention to another difficult task e.g., driving a car [18] and is not available to engage with the suggested activity (*Load*).

The ability to perform a task can be also affected by self-efficacy, i.e., perception of one's capability to execute the target activity, and whether the person has performed this activity successfully before. Self-efficacy has been shown to relate to treatment adherence [19] (*Self-efficacy*).

Finally, patient's ability to perform the target behaviour may be affected by the patient being tired [20] or bored of [21] repeating the same activity (*Strained*).

$$Ability = Low\ Load + Self\text{-}efficacy + Unstrained \qquad (3)$$

Trigger. In a laboratory study, Goyal *et al.* [11] found that the best moment to interrupt people is when they are in a state of increasing arousal (*Arousal*).

Bidargadi *et al.* found that the notifications delivered on weekends or during midday are more effective [22]. On the other hand, participants in the Kunzler *et al.* study were more receptive to intervention on weekdays rather than weekends between 10am-6pm[23] (*Day*). Saikia *et al.* [9] also found that the timing of the notification is important; in their study, the largest percentage of users engaged with notification around 11am (*Time*).

In addition, multiple studies considered persons' current location and motion activity in finding the most appropriate notification delivery context [12, 23] (*Location, Motion*) .

$$Trigger = Arousal + Day + Time + Location + Motion \qquad (4)$$

We set the trigger to zero when patient is sleeping in order not to interrupt sleep with activity suggestions.

Table 1

Factors impacting motivation, ability and trigger

Factors	States (scores)	Function of	Based on
Valence	positive (1), negative (0)	sleep (last 24h) physical activity (last 24h) no. notifications (last 24h)	Poor sleep quality correlates with high negative and low positive emotions [26] Physical activity has a positive within-day association with pleasant core affects [27] Stress is a state of high arousal and negative valence:
Arousal	low (0), mid (1), high (0)		- physical inactivity increases chances of high stress [28] - smartphone user experience stress from receiving frequent notifications [29] The optimal physiological arousal resulting in flow-experience lies between stress and relaxation [30]
Cognitive load	low (1), high (0)	notifications timing	Notifications at detected breakpoint timing resulted in lower cognitive load compared to randomly timed notifications [31]
Awake	awake (1), asleep (0)	time of the day physical activity (last 24h)	The daily amount of activity positively relates with sleep quality [32] High arousal is associated with reductions in slow-wave sleep and negative valence with disruptions to REM sleep [33]
Sufficient sleep	yes (1), no (0)	arousal valence	
Perceived benefit	yes (1), no (0)	constant (between each performed activity) valence (during activity)	During-behaviour affect is predictive of concurrent and future physical activity behaviour [34]
Self-efficacy (confidence)	low (0), high (1)	no. activity performed (overall)	
Unstrained	yes (1), no (0)	no. activity performed (last 24h) time since the activity	
Location	home (1), other (0)	time	
Motion	stationary (1), walking (0)	time successful notification past walking frequency	
Time of the day	morning (0), midday (1), evening (0), night (0)	time (one hour time-step)	
Day	weekend(1), week day (0)		
Has family	yes (1), no (0)	constant	

Some of the above-described factors affecting motivation, ability, and trigger are constant for the patient (e.g., the presence of family members); others vary with time and can be obtained through self-reporting or inferred from signals gathered by wearable devices (e.g., cognitive load [24] or stress[25]). Table 1 lists the factors affecting behaviour along with their states. To compute the behaviour components, we assign the value of 1 to the factor state that benefits the behaviour and 0 otherwise. For example, motivation = 4 if the patient is in a positive valence (1), has support from family (1), perceives benefit in performing the activity (1) and had sufficient amount of sleep last night (1). Let's say that the patient also has a low cognitive load (1) at the moment and feels confident that they can perform the task (1) but by now they are tired of and bored from repeating the activity (0) so their ability = 2. In terms of trigger the patient is in a low arousal state (0), it is a weekday (0)[1], midday (1), home (1) and the patient is sitting (1) = 3. So we have $motivation(4) \times ability(2) \times trigger(3) = 24$. When the *action threshold* $= 25$ the behaviour would not be performed. But let's say that if the patient was not yet strained from repeating activity today their ability would be 3 and behaviour computation would result in 36, which is above the action threshold and the behaviour would occur (activity performed).

---

[1]Trigger scores (except arousal) are affected by preference, e.g. if patient prefers to perform activity on weekday the score for weekday will be 1 rather than 0.

2.2. Machine learning for notifications

Ho *et al.* [12] framed the task of notifying users at the right time as an RL problem. In Ho's *et al.* framework, the goal is to maximize the effectiveness of the notifications. The notification system is an agent that interacts with the app user (environment). Based on the user's context (the state received from the environment), the agent chooses to notify or not (action). The agent is rewarded when the user responds to the notification. Ho *et al.* conclude that the Q-Learning (QL) [35] RL method yields higher notification response rate on crowd-sourced data than support vector machine or shallow neural network trained in a supervised fashion. In follow up real-life five-week study, the authors compared Advance Actor Critic (A2C) RL approach against Random Forest (RF) that has been trained on data gathered with random notification policy in the first three weeks of the study [36]. The authors concluded that A2C is more effective in reducing dismissed notifications, but the supervised approach achieved a higher notification answer rate. They suggested that the benefits of RL method might be more prominent in the context of users whose preferences change. We consider this condition in our experiments (see Section 4).

Kunzler *et al.* [23] utilised RF to detect receptivity to Just-In-Time Adaptive Interventions and showed that it increased receptivity over biased random classifier, however, they did not consider RL based approaches. In the follow-up study Mishran *et al.* [37] compared the static supervised learning approach against the adaptive approach. In adaptive set up the supervised model is retrained every time new receptivity data are made available. Mishran *et al.* observes that receptivity to notifications from the adaptive model improved over the course of the study. We include the adaptive supervised learning method in our experiments and compare against RL-based approaches (see Section 3.2).

Sutton *et al.* [38] proposed that RL could be used for management of notifications coming from various applications. Their goal was to learn which notifications are important and which should be blocked. They created Open AI gym environment for the training of QL and Deep QL (DQL) agents on synthetic data. When the models were applied to real data, they could predict a user's action toward notifications better than a random benchmark. The authors suggested that the synthetic data could be applied for RL training in cases where the use of real data is not possible. Encouraged by Sutton's *et al.* findings we use Open AI gym [39] to create an environment (see Section 3.1) and synthesize patients behaviour according to the behaviour model described in Section 2.1.

3. Methods

We draw inspiration from works described above and formulate our problem similarly to Ho *et al.* [36], where the VC is an agent, a patient is an environment and the patient's context is the environments state. However, we use a richer description of the context which includes the patient's physiological data that will be captured by the smartwatch (thirteen variables described below).

3.1. Patient environment

We used Open AI gym [39] to create a patient environment and simulate patients response to prompt according to the equation 1^2. We simulate a patient who has a family, prefers being notified at midday, and does not want to receive more than 3 notifications daily. The patient's action threshold is set to 20^3 and their initial state shows issues with sleeping (less than 5h a day), negative valence for majority of the day and no physical activity, to improve patient sleep and mood they are recommended to develop a walking habit.

The patient state available to the VC includes thirteen variables: time of day, weekday, benefit score (since the last action performed), location, awake state, valence, arousal, cognitive load, motion, time since activity performed, number of hours slept (in the last 24h), number of times notified (in the last 24h) and number of time patient performed the activity (in the last 24h).

3.2. Prompt strategies

We consider three different prompt strategies that are described below.

Random: At each time step (each hour) the VC randomly decides whether to send or not a prompt to the patient, except at night time.

Supervised Approaches: Following [36, 23, 37] we selected RF model and train it on data gathered during the first three weeks (as in [36]) of simulated intervention. We use scikit-learn [40] implementation of the model with the default parameters and balanced class weighting. The model is trained in a supervised fashion on pairs of patient states and their response to prompt (the activity was performed or not). During the sample gathering phase, the prompts are sent every hour except at night. After the initial training phase, the model prompts the patient when it predicts that they will perform the activity, given their current context (the thirteen variables). Following [37] we consider both *static* and *adaptive* model training approaches. In the former case, the model is applied as it is and no further training after the initial training phase is performed. In the latter case, after each prompt, a new example describing the context and the response is added to the training set and the model is retrained.

RL Approaches: Every hour the VC decides whether to send a prompt or not using RL model (agent). If suggested by the model, it sends the prompt to the environment (patient). The environment calculates the value of eq. 1 based on the results simulates the action of walking and then responds to agent with a reward. The agent always receives a reward of 20 units for notifications which result in the target walking behaviour (even if the number of prompts exceed the daily notification threshold), -1 in cases where the agent notified the simulated patient but the patient did not perform the activity, -10 when the agent sent a notification which resulted in exceeding the tolerated number of daily notifications and the patient did not perform the activity, and 0 when the agent did not send a notification.

We employ three RL approaches: DQL, A2C and proximal policy optimization (PPO) algorithm, which is the state of the art for many continuous control problems [41] . For all we use implementation from stablebaseline3 [42] with default parameters, except that DQL

---

[2]Simulation code at https://github.com/Capable-project/capable-rl4vc.git

[3]At this threshold 70% of runs with hourly notification results in target behaviour being performed at least once during the course of intervention.
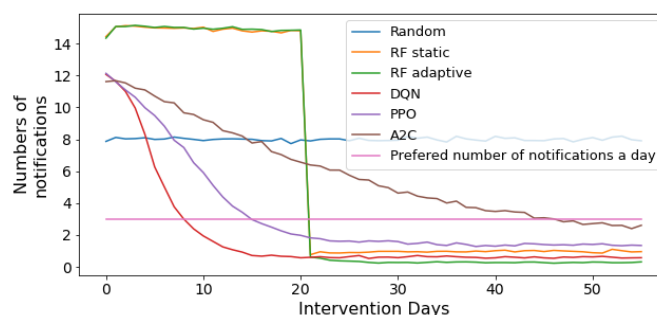
Figure 1: Comparison of the prompt strategies in terms of a number of prompts daily thought the intervention (this excludes runs with failed training).

$learning\_start = 24$ (1 day of data gathering), A2C and PPO $n\_steps = 24$ (update every day).

4. Experiments

We simulate eight week long digital intervention and compare all prompt strategies in terms of their effectiveness (activity performed to prompt ratio) and potential annoyance to the user (number of prompts daily) in four conditions:

- Stable responses pattern – patient preferences and activity threshold stay constant across the intervention.
- Habituation – patient habituates to prompts (its activity threshold increases every day by 0.15). Given how we define eq. 1 and specific factors (all integers), this corresponds to increasing the threshold by 1 after each week.
- Preference shift – patient changes their preferred time of notification to the evening after week 5.
- Preference shift + Habituation – patient habituates to prompts and changes their preferred time of notification to the evening after week 5.

Given the variation in the results between runs, we ran each strategy 500 times. Figure 1 shows the algorithms' learning of the number of daily notifications (fewer notifications are desirable to patients) and Figure 2 shows the mean prompt effectiveness across all runs.

After 3 weeks (21 days) of initial training, the supervised approaches (both static and adaptive) send the prompts at more appropriate times than other methods, which manifest in the higher prompt effectiveness (Fig. 2). The main disadvantage of the supervised learning methods is that they might be causing annoyance to the user at the initial stage of the intervention as they require frequent prompts to gather positive and negative training samples. The model performance also depends on how many positive samples have been gathered during that initial training phase. Runs in which none of the prompts resulted in an activity being performed in first 21 days lead to failed model training (no convergence) and the model always predicting not to prompt the user (these runs are excluded from Fig. 1 and 2).
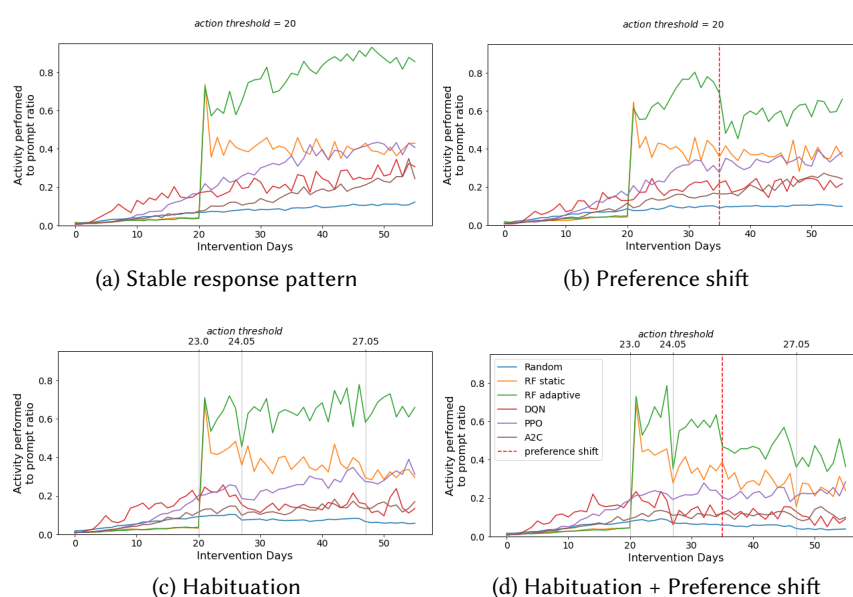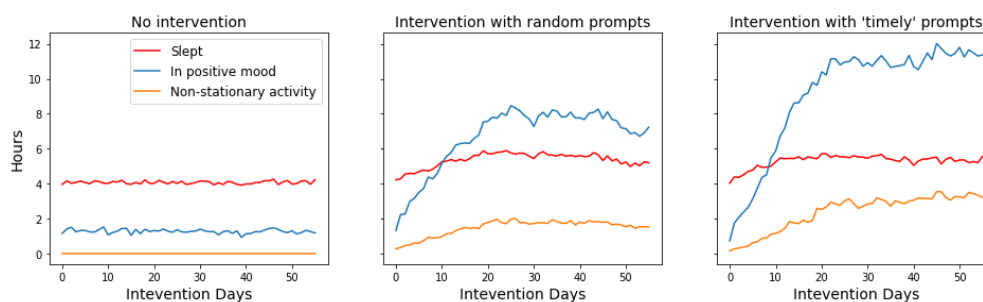
Figure 2: Comparison of the prompt strategies in terms of prompt effectiveness. Note the mean is computed from runs when at least 1 prompt has been sent.

All the reinforcement learning approaches learned to prompt the user less than 3 times daily in the course of the intervention (Fig. 1). DQN reaches desired number of prompts the fastest within around 9 days and the ratio of prompts that results in the activity performed is higher than in the case of randomly timed prompts (Fig. 2a). The best reinforcement learning method in our experiments is PPO which slowly reaches the prompt effectiveness of supervised model but requires fewer prompts in the initial three weeks of intervention.

Changes in simulated patients' prompt response patterns (Fig. 2b,c,d) have the most visible impact on supervised models. Nevertheless, if initially successfully trained, the supervised methods remain more effective in prompting patients than other approaches. The drop in the response effectiveness in the case of the adaptive supervised training approach is due to reusing the training samples gathered before the simulated patient preferences shift. There were not enough samples to train the new model after detecting performance drop. Up-weighting of the more recent samples during model retraining or use of ensemble learning method designed to tackle concept drift[43] could lead to better prompt strategy adaptation.

We also checked if the simulated patient environment, even in the hardest simulation conditions (habituated + preference shift), captures the better habit formation and consequently results in improvements in patients' state when interventions are administered. Fig. 3 shows patient state without intervention, with intervention when the prompts are sent hourly during the day and intervention with timely notifications using PPO. The simulated patient changed their state as a result of intervention with walking activity, sparking improvements in mood and number of hours slept. Note that in our simulation positive valence depends on both the physical activity and the number of notifications (Table 1), therefore **the timely notification causes a longer time of being in a positive mood** than randomly timed prompts.

Figure 3: Effect of intervention on simulated patient.

5. Discussion

The comparison of different prompt strategies in a simulated patient environment suggests that the supervised learning approaches to patient's notifications might send the most appropriately timed prompts. However, their training could potentially tire patients with a large number of prompts during the initial intervention stage. It might be interesting to investigate if a smart sampling strategy, such as diversity sampling could reduce the notification burden in the initial intervention stage but keep model performance high in the remainder of the intervention.

In our experiments, PPO was the most effective RL-based prompting method. It almost reached activity performed to prompt ratio of supervised models at the end of the intervention period, and it required sending fewer prompts than supervised trained models overall, which might burden less patients. PPO displayed stable improvement in prompt effectiveness over time, regardless of the simulated patient prompt habituation or preference shift, making it potentially an interesting candidate approach for real-life study.

All learning methods considered in this study could benefit from a larger number of training samples, especially those capturing the context in which prompt results in the performed activity. Here we simulated a single patient but in reality, there might be multiple patients with similar behaviour patterns and leveraging learning from them could provide a boost in model performance. Currently, one limitation of our simulated environment is that we do not capture personality differences, which have been shown to impact user receptivity to notification [23]. In future work, we aim to include this information in our simulation and investigate learning from multiple different users.

Acknowledgments

References

[1] J. C. Holland, Y. Alici, Management of distress in cancer patients., The Journal of Supportive Oncology 8 (2010) 4–12.

[2] A. Pitman, S. Suleman, N. Hyde, A. Hodgkiss, Depression and anxiety in patients with cancer, BMJ 361 (2018).

[3] M. R. DiMatteo, K. B. Haskard-Zolnierek, Impact of depression on treatment adherence and survival from cancer, Depression and Cancer (2011) 101–124.

[4] E. B. Hekler, S. Michie, M. Pavel, D. E. Rivera, L. M. Collins, H. B. Jimison, C. Garnett, S. Parral, D. Spruijt-Metz, Advancing models and theories for digital behavior change interventions, American Journal of Preventive Medicine 51 (2016) 825–832.

[5] P. Kelly, C. Williamson, A. G. Niven, R. Hunter, N. Mutrie, J. Richards, Walking on sunshine: scoping review of the evidence for walking and mental health, British Journal of Sports Medicine 52 (2018) 800–806.

[6] D. B. Wilson, J. S. Porter, G. Parker, J. Kilpatrick, Peer reviewed: Anthropometric changes using a walking intervention in african american breast cancer survivors: A pilot study, Preventing Chronic Disease 2 (2005).

[7] L. J. Frensham, G. Parfitt, J. Dollman, Effect of a 12-week online walking intervention on health and quality of life in cancer survivors: a quasi-randomized controlled trial, International Journal of Environmental Research and Public Health 15 (2018) 2081.

[8] B. J. Fogg, Tiny Habits: The Small Changes That Change Everything, Houghton Mifflin Harcourt, 2019.

[9] P. Saikia, M. Cheung, J. She, S. Park, Effectiveness of mobile notification delivery, in: 2017 18th IEEE International Conference on Mobile Data Management (MDM), IEEE, 2017, pp. 21–29.

[10] H. Sarker, M. Sharmin, A. A. Ali, M. M. Rahman, R. Bari, S. M. Hossain, S. Kumar, Assessing the availability of users to engage in just-in-time intervention in the natural environment, in: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014, pp. 909–920.

[11] N. Goyal, S. R. Fussell, Intelligent interruption management using electro dermal activity based physiological sensor for collaborative sensemaking, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1 (2017) 1–21.

[12] B.-J. Ho, B. Balaji, M. Koseoglu, M. Srivastava, Nurture: notifying users at the right time using reinforcement learning, in: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, 2018, pp. 1194–1201.

[13] B. J. Fogg, A behavior model for persuasive design, in: Proceedings of the 4th International Conference on Persuasive Technology, 2009, pp. 1–7.

[14] T. Jowsey, C. Pearce-Brown, K. A. Douglas, L. Yen, What motivates australian health service users with chronic illness to engage in self-management behaviour?, Health Expectations 17 (2014) 267–277.

[15] J. Axelsson, M. Ingre, G. Kecklund, M. Lekander, K. P. Wright Jr, T. Sundelin, Sleepiness as motivation: a potential mechanism for how sleep deprivation affects behavior, Sleep 43 (2020) zsz291.

[16] M. R. Dolsen, A. M. Soehner, C. M. Morin, L. Bélanger, M. Walker, A. G. Harvey, Sleep the night before and after a treatment session: A critical ingredient for treatment adherence?, Journal of Consulting and Clinical Psychology 85 (2017) 647.

[17] S. W. Chan, S. Sapkota, R. Mathews, H. Zhang, S. Nanayakkara, Prompto: Investigating receptivity to prompts based on cognitive load from memory training conversational agent, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4 (2020) 1–23.

[18] Y.-C. Lee, J. D. Lee, L. Ng Boyle, The interaction of cognitive load and attention-directing cues in driving, Human Factors 51 (2009) 271–280.

[19] A. Luszczynska, R. Schwarzer, The role of self-efficacy in health self-regulation, The Adaptive Self: Personal Continuity and Intentional Self-development 15 (2005) 137–152.

[20] S. F. Chastin, N. Fitzpatrick, M. Andrews, N. DiCroce, Determinants of sedentary behavior, motivation, barriers and strategies to reduce sitting time in older women: a qualitative investigation, International Journal of Environmental Research and Public Health 11 (2014) 773–791.

[21] S. Bucci, J. Ainsworth, C. Barrowclough, S. Lewis, G. Haddock, K. Berry, R. Emsley, D. Edge, M. Machin, A theory-informed digital health intervention in people with severe mental health problems, in: MEDINFO 2019: Health and Wellbeing e-Networks for All, IOS Press, 2019, pp. 526–530.

[22] N. Bidargaddi, D. Almirall, S. Murphy, I. Nahum-Shani, M. Kovalcik, T. Pituch, H. Maaieh, V. Strecher, To prompt or not to prompt? a microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app, JMIR mHealth and uHealth 6 (2018) e10123.

[23] F. Künzler, V. Mishra, J.-N. Kramer, D. Kotz, E. Fleisch, T. Kowatsch, Exploring the state-of-receptivity for mhealth interventions, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3 (2019) 1–27.

[24] A. Lisowska, S. Wilk, M. Peleg, Is it a good time to survey you? cognitive load classification from blood volume pulse, in: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2021, pp. 137–141.

[25] A. Lisowska, S. Wilk, M. Peleg, Catching patient's attention at the right time to help them undergo behavioural change: Stress classification experiment from blood volume pulse, in: Proceedings of the 19th International Conference on Artificial Intelligence in Medicine, AIME 2021, Springer, 2021, pp. 72–82.

[26] C. Baglioni, K. Spiegelhalder, C. Lombardo, D. Riemann, Sleep and emotions: a focus on insomnia, Sleep Medicine Reviews 14 (2010) 227–238.

[27] A. Ivarsson, A. Stenling, K. Weman Josefsson, S. Höglind, M. Lindwall, Associations between physical activity and core affects within and across days: a daily diary study, Psychology & Health 36 (2021) 43–58.

[28] X. Zhai, M. Ye, C. Wang, Q. Gu, T. Huang, K. Wang, Z. Chen, X. Fan, Associations among physical activity and smartphone use with perceived stress and sleep quality of chinese college students, Mental Health and Physical Activity 18 (2020) 100323.

[29] S. Yoon, S.-s. Lee, J.-m. Lee, K. Lee, Understanding notification stress of smartphone messenger app, in: CHI'14 Extended Abstracts on Human Factors in Computing Systems, 2014, pp. 1735–1740.

[30] C. Peifer, A. Schulz, H. Schächinger, N. Baumann, C. H. Antoni, The relation of flow-experience and physiological arousal under stress—can u shape it?, Journal of Experimental Social Psychology 53 (2014) 62–69.

[31] T. Okoshi, K. Tsubouchi, M. Taji, T. Ichikawa, H. Tokuda, Attention and engagement-awareness in the wild: A large-scale study with adaptive notifications, in: 2017 IEEE International Conference on Pervasive Computing and Communications (PERCOM), IEEE, 2017, pp. 100–110.

[32] A. N. S. Bisson, S. A. Robinson, M. E. Lachman, Walk to a better night of sleep: testing the relationship between physical activity and sleep, Sleep Health 5 (2019) 487–494.

[33] C. P. Fairholme, R. Manber, Sleep, emotions, and emotion regulation: an overview, Sleep and Affect (2015) 45–61.

[34] D. M. Williams, S. Dunsiger, E. G. Jennings, B. H. Marcus, Does affective valence during and immediately following a 10-min walk predict concurrent and future physical activity?, Annals of Behavioral Medicine 44 (2012) 43–51.

[35] C. J. Watkins, P. Dayan, Q-learning, Machine Learning 8 (1992) 279–292.

[36] B.-J. Ho, B. Balaji, M. Koseoglu, S. Sandha, S. Pei, M. Srivastava, Quick question: Interrupting users for microtasks with reinforcement learning, arXiv preprint arXiv:2007.09515 (2020).

[37] V. Mishra, F. Künzler, J.-N. Kramer, E. Fleisch, T. Kowatsch, D. Kotz, Detecting receptivity for mhealth interventions in the natural environment, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5 (2021) 1–24.

[38] R. Sutton, K. Fraser, O. Conlan, A reinforcement learning and synthetic data approach to mobile notification management, in: Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia, 2019, pp. 155–164.

[39] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, W. Zaremba, OpenAI Gym, 2016. `arXiv:arXiv:1606.01540`.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[41] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, arXiv preprint arXiv:1707.06347 (2017).

[42] A. Raffin, A. Hill, M. Ernestus, A. Gleave, A. Kanervisto, N. Dormann, Stable baselines3, https://github.com/DLR-RM/stable-baselines3, 2019.

[43] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, M. Woźniak, Ensemble learning for data stream analysis: A survey, Information Fusion 37 (2017) 132–156.