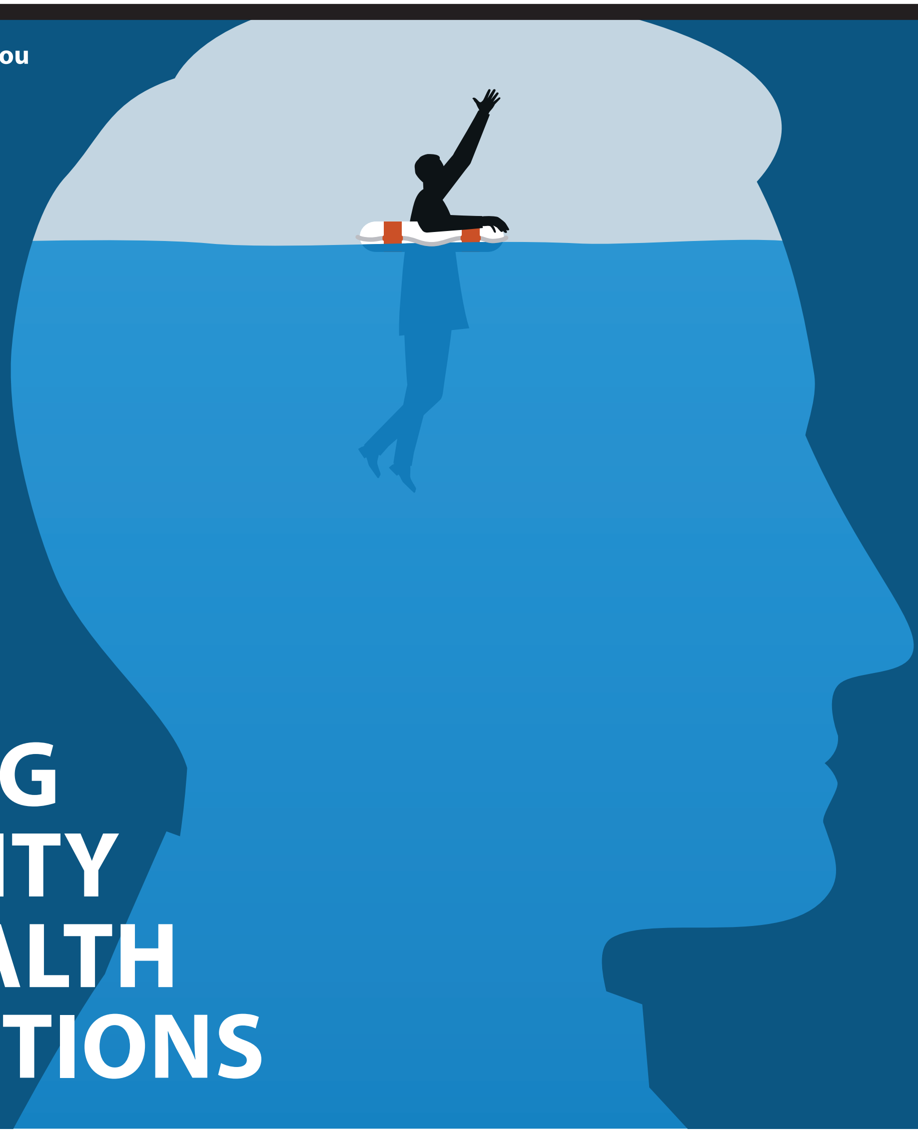
Varun Mishra *Northeastern University, Boston, MA*
Florian Künzler *Nash Exchange* **Jan-Niklas Kramer** *CSS Health Insurance, Switzerland*
Elgar Fleisch *ETH Zürich* **Tobias Kowatsch** *University of Zurich, University of St. Gallen, Switzerland*
David Kotz *Dartmouth College, Hanover, NH*

Editors: Nicholas D. Lane and Xia Zhou



DETECTING RECEPTIVITY FOR mHEALTH INTERVENTIONS

Excerpted from "Detecting Receptivity for mHealth Interventions in the Natural Environment," from *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Volume 5, Issue 2, with permission. https://dl.acm.org/doi/10.1145/3463492

Just-In-Time Adaptive Interventions (JITAI) have the potential to provide effective support for health behavior by delivering the right type and amount of intervention at the right time. The timing of interventions is crucial to ensure that users are receptive and able to use the support provided. Previous research has explored the association of context and user-specific traits on receptivity and built machine-learning models to detect receptivity after the study was completed. However, for effective intervention delivery, JITAI systems need to make in-the-moment decisions about a user's receptivity. In this study, we deployed machine-learning models in a chatbot-based digital coach to predict receptivity for physical-activity interventions. We included a *static model* that was built before the study and an *adaptive model* that continuously updated itself during the study. Compared to a control model that sent intervention messages randomly, the machine-learning models improved receptivity by up to 36%. Receptivity to messages from the adaptive model increased over time.

Illustration, istockphoto.com

The ubiquitous presence of mobile technologies has enabled a wide array of research into mobile health (mHealth), from sensing health conditions to providing behavior-change interventions. In the past, ubiquitous technologies like smartphones and wearables have shown promise in detecting stress, anxiety, mood, depression, personality change, addictive behavior, physical activity and a host of other conditions. Furthermore, several studies have demonstrated the potential of smartphone-based digital interventions to affect positive behavior change for a range of conditions like smoking, alcohol disorder, eating disorders, and physical inactivity. The eventual goal in mHealth is to be able to combine the two components of accurate sensing and effective interventions to improve the quality of life amongst people suffering from various conditions.

Just-In-Time Adaptive Intervention (JITAI) is a novel intervention design that aims to deliver the right type and amount of support, at the right time, while adapting as-needed to the users' internal and external contextual change [6]. Several studies have employed JITAI-like interventions for various outcomes, e.g., improving physical inactivity [3], and reducing alcohol use [1]. For JITAIs to be effective, the intervention should be delivered at "the right time." Two key concepts determine the "right time": (1) when a person needs support, i.e., at or before the onset of a negative outcome, or a psychological or contextual state that might lead to that outcome (*state-of-vulnerability*); and (2) when a person is able and willing to receive, process, and use the support provided (*state-of-receptivity*).

In our prior study, we developed the *Ally* app to deliver physical-activity interventions and explored how the passively collected contextual factors associated with receptivity in a study with 189 participants [4]. We also evaluated the feasibility of building machine-learning models to detect receptivity and achieved a 77% improvement in F1-score over a biased random model. Other researchers have also explored discriminative features and built machine-learning models to detect receptivity to just-in-time interventions [7, 2]. Most of these studies, however, focused on data collection followed by post-study analysis and evaluation of post-study machine-learning models, with the expectation that the models would perform similarly when deployed in real-life conditions.

In this paper, we go beyond post-study analysis: we deployed two different machine-learning models to predict in-the-moment receptivity and used that prediction to decide when to deliver the intervention. We deployed these models in a physical-activity app used by 83 participants in free-living conditions over a period of 3 weeks. Our goal was to evaluate whether such models actually helped increase receptivity to interventions.

Given the promising results from our previous study, we decided to build upon that work [4]. We used the data from the previous study to build two different machine-learning models, which we later deployed in our field study (a) a *static* model that remained constant for all participants throughout the study, and (b) an *adaptive* model that continuously learned the receptivity of individual participants from their enrollment in the study and updated the model as the study progressed; we delayed activation of this model until the participant had been in the study for 7 days, however, to ensure that enough data was collected for that participant before using the model's predictions. To compare the utility of these models, we also included (c) a *control* model that would send the intervention messages at a random time. We extended the original Ally app [4], to incorporate the different models and enable in-the-moment detection of receptivity.[1] We deployed these models in a 3-week study with 83 participants and observed that the static model led to significantly higher receptivity than the control model and while the adaptive model did not show significant improvements over the control model, the receptivity to messages delivered by the adaptive model improved as participants progressed in the study.

THE ALLY STUDY

The Ally app – based on the open-source MobileCoach framework – was a chat-based digital coach (for Android and iOS phones) that delivered an actual behavior-change intervention aimed at increasing the participant's daily step count. The previous study was conducted with 189 participants in Switzerland over a period of 6 weeks. Participants received notifications that encouraged them to engage in conversation with the digital coach, which was a German-speaking chatbot motivating participants to increase their physical activity as measured by daily step count [4].

Given the promising results in the post-study analyses, we decided to build upon that work by deploying in-the-moment receptivity-detection models to evaluate how these models perform in real-world situations.

OPERATIONALIZING RECEPTIVITY

Before discussing our methodology, it is important to establish precise metrics about what the models are trying to achieve. These definitions are consistent with the metrics we used in our prior work [4].

- **Just-in-time response:** If a user views and responds to the initiating message within 10 minutes. We chose a 10-minute window to remain consistent with our prior work, where we also used a 10-minute window to define the receptivity metrics [4].
- **Response:** If the user responds to the initiating message at any time, even after the first 10 minutes, it counts as a "response."
- **Response delay:** The time (in seconds) elapsed between receipt of the initiating message and the user's first reply to it.
- **Conversation engagement:** If the user replies to more than one message in a 10-minute window following the initiating message, it counts as "conversation engagement."

THE ALLY+ APP

We modified the iOS version of the Ally app to create a new app we call *Ally+*. Similar to Ally, Ally+ is a chat-based digital coach aimed at increasing daily step count. We show a screenshot of the app in Figure 1. The intervention components were chat-based conversational messages that were delivered by the digital coach and the participants had to choose from a set of pre-defined responses. The coach initiated

[1] During the first 7 days, the app randomly chose between the *control* and *static* models. After 7 days, the app randomly chose between the *control, static,* and *adaptive* models.

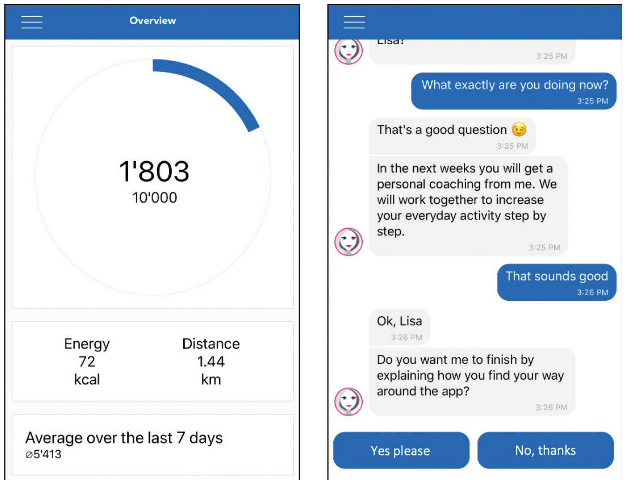


FIGURE 1. Two screenshots showing Ally+'s dashboard (left) and the chat screen for the interventions (right).

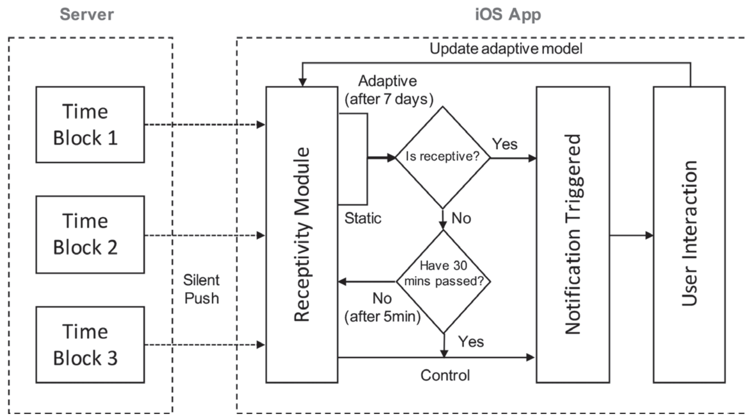


FIGURE 2. System design of the Ally+ app.

the starting message of each conversation to each user at random times within certain time periods.

Further, Ally+ had a context-based receptivity module that continuously tracked several contextual features; Ally+ used this module to time the delivery of notifications, as follows. For each day, for each participant, the server randomly chose three times (one in each of the three time blocks) to send a *silent* push notification to that participant's app. When Ally+ received the silent push from the server, it triggered the receptivity module to determine when to deliver that notification to the participant. During the first seven days, the receptivity module randomly selected either the control or static model, with equal weight. On the eighth day and after, the receptivity module randomly selected one of three models, with equal weight. (The seven-day "warm-up" period allowed accumulation of participant-specific receptivity data before enabling the adaptive model.) For each initiating message received, the app recorded which model was used to time its delivery – control, static or adaptive.

Ally+ then delivered the notification about the initiation prompt if and only if the selected model inferred the user would be receptive at the current time. The control model always agreed. The static and adaptive models used their classifier to determine whether the current moment is "receptive." If the models did not find the current moment to be receptive, the app would try again by asking the same model every 5 minutes. If, after 30 minutes, the model never inferred an opportune moment, Ally+ delivered the notification on the 31st minute; in this case, it recorded the delivery mechanism as "control," since the notification was delivered at a random time, and not at an opportune moment.

We used the Ally+ app to conduct a within subjects' study with three experimental conditions for delivering the interventions: control, static, and adaptive. It is important to note that the intervention delivery conditions did not affect the actual content of the interventions delivered by the app.

Regardless of the chosen delivery model, the participant's response to any initiating message provided new data for use by the adaptive model. There were three cases: (a) just-in-time response: the contextual state at the time of notification delivery was added with label "receptive"; (b) later response: the contextual state at the time of notification delivery was added with label "non-receptive," and the contextual state at the moment of response was added with label "receptive" (since the participant was in a state-of-receptivity when they responded); (c) no response: the contextual state at the time of notification delivery was added with label "non-receptive." Whenever the adaptive model was selected as the delivery model, it first retrained its model using any new data points added. We diagram the system design in Figure 2.

THE *Static* AND *Adaptive* MODELS

We implemented two machine-learning models in Ally+. We trained the static model before deployment (using data from the previous the 141 iOS users in Ally study) and used it, unchanged, for all participants and all days throughout the study. The adaptive model used the receptivity data of individual participants as they progressed through the study; it was rebuilt (within the app) every time a new receptivity in the system triggered the adaptive model.

Both these models were trained to predict *just-in-time response*. While we use several metrics of receptivity in our work, the main emphasis is on the presence of a just-in-time response. For completeness, however, we report the effect of our models on the various receptivity metrics.

Static Model: We used CoreML to build and integrate the static model with the iOS app. We split the original Ally iOS data (with 141 users) into five equal non-overlapping groups. We used Leave-One-Group-Out (LOGO) cross-validation to evaluate two built-in models within CoreML – MLRandomForestClassifier and MLSupportVectorClassifier. These classifiers are CoreML's implementation of RandomForest and SVM, respectively.

We tuned the models to have higher recall, since we wanted Ally+ to recognize most opportune moments, even if it was at the cost of precision. We compared the models with a random classifier as a baseline and chose the model that demonstrated a greater improvement in F1 score. The SVM classifier achieved a mean F1 score of 0.36, whereas the random baseline classifier achieved only F1 score of 0.25, which is an improvement of 40% over the baseline. The RandomForest classifier achieved a mean F1 score of 0.33, only 32% improvement over baseline. We thus chose the SVM classifier as the static model to be included in our app.

Adaptive Model: In the adaptive model, the participant's recent receptivity data was added to the model's training dataset to help with future detection. Given the structure of our study, however, each participant was prompted at most three times per day and there were thus few data points even after seven days. We thus followed a "dual-model" approach: the adaptive model's output probability was the average of the output probability from "P1," a model trained on data from the prior Ally study, and "P2," a Logistic Regression (LR) model trained on the participant's personalized data accumulated thus far. If the output probability was greater than 0.50, the adaptive model classified that instance as "receptive." This dual-model approach enabled us to introduce a degree of personalization without being concerned about high variance of the personalized model developed from a limited set of data points.

THE ALLY+ STUDY

Since the goal of the study was to evaluate participant receptivity, we did not want to bias the participant's interaction and usage of the app by providing monetary incentives for using the app or for engaging with the app. Instead, our study strategy used "deception" to mask the actual goals of the study. During recruitment, we told the participants the goal of the study was to understand how different contexts affect the physical activity levels of a person throughout their day. We asked participants to interact naturally with the Ally+ app and compensated them the equivalent of USD 25 if they installed the app for at least two-thirds of the study duration, i.e., 14 days.

The study protocol (including the use of deception) was approved by the Institutional Review Board (IRB) of the respective institutions. As required by the IRB, at the end of the 3-week period, we emailed the participants informing them of the real goal of the study with an explanation of why deception was needed.

We used Facebook advertisements to reach potential participants. A total of 83 participants (64 female; 30±10.8 years) downloaded the app and started the intervention. We did not have any exclusion criteria and report results from all participants. Across the 83 users, we had 1091 messages delivered by the control model, 691 messages delivered by the static model, and 241 messages delivered by the adaptive model; resulting in a total of 2023 delivered messages.

EVALUATION

We used generalized linear mixed effects models for our analysis. We observed that the model type had a significant effect on the just-in-time response rate ($\chi^2(2) = 13.433$, $p = 0.001$). On post-hoc analysis, we observed that the static model showed a significant improvement of over 36% in just-in-time receptivity when compared to the control model ($p = 0.002$). This result suggests that if a participant received a prompt from the static model, they were more likely to be receptive than if the same participant received the prompt through the control model. The adaptive model led to an increase of almost 10% over the control model, but the result was not significant ($p = 0.558$).

For the secondary metrics, we observed that the type of model had an effect on the likelihood of response ($\chi^2(2) = 8.364$, $p = 0.00$). Post-hoc analysis revealed that only the static model had a significant improvement over the control model, with an improvement of almost 10% ($p = 0.015$). Further, our analysis showed that the model type had a significant effect on the likelihood of conversation engagement ($\chi^2(2) = 10.407$, $p = 0.017$), with post-hoc analysis revealing that the static model led to an improvement of over 32% in the

TABLE 1. Detailed analysis to understand *within-participant* differences. We report the absolute change of the static and dynamic models over the control model, along with the percentage improvement in brackets.

Comparison	Mean Difference %	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Adj.	*p*-value
Just-in-time response (as likelihood; control = 0.276)						
static – control	+0.101 (+36.60%)	0.033	0.035	0.170	0.002	**
adaptive – control	+0.027 (+9.58%)	0.041	−0.044	0.109	0.558	
Overall response (as likelihood; control = 0.738)						
static – control	+0.072 (+9.75%)	0.028	0.015	0.116	0.015	*
adaptive – control	+0.031 (+4.20%)	0.038	−0.046	0.092	0.493	
Conversation engagement (as likelihood; control = 0.261)						
static – control	+0.084 (+32.18%)	0.034	0.021	0.153	0.007	**
adaptive – control	+0.009 (+3.44%)	0.040	−0.057	0.089	0.819	
Response delay (as minutes; control = 99.500)						
static – control	−19.950 (−20.05%)	11.725	−39.500	3.500	0.124	
adaptive – control	−13.830 (−13.89%)	13.585	−41.000	13.500	0.439	

. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

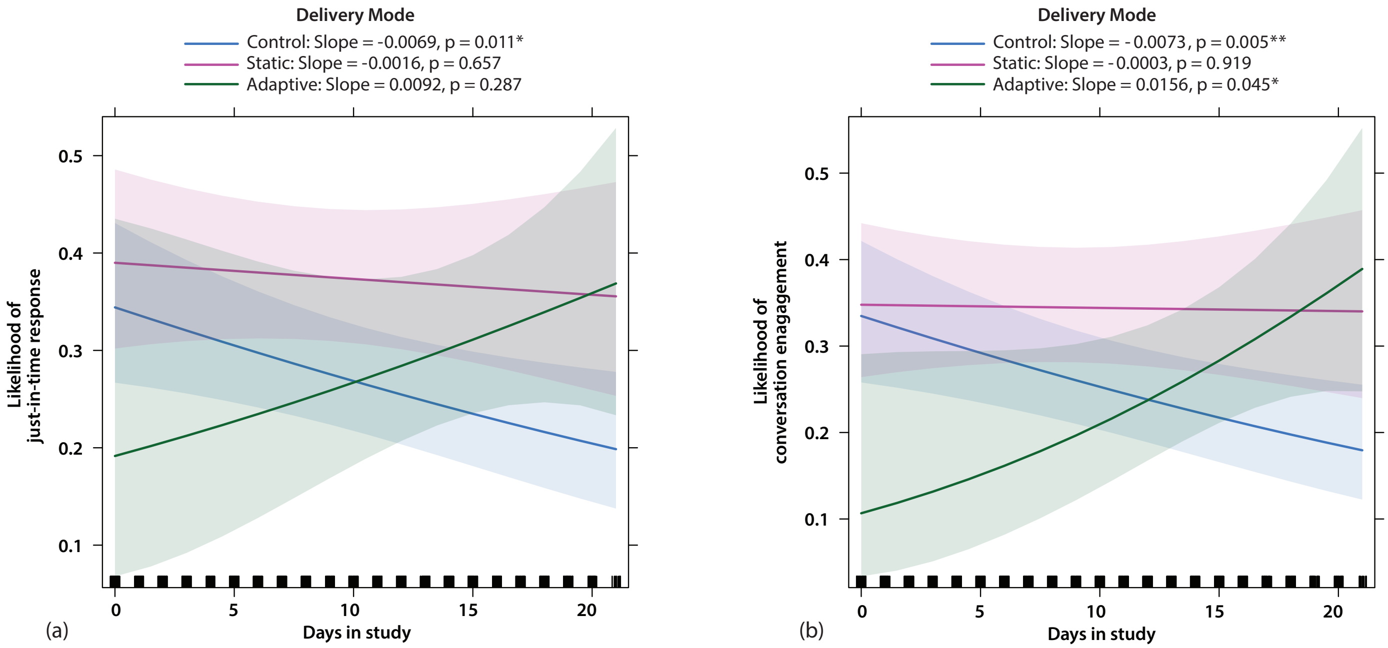


FIGURE 3. The performance over time of the models on the receptivity metrics. The adaptive model was only activated starting Day 8; the dotted lines represent the projection of the trend for the adaptive model from Day 1 to Day 7.

likelihood of conversation engagement over the control model ($p = 0.007$). We present the detailed findings in Table 1.

As we observe from Table 1, for most receptivity metrics, messages delivered by the static model led to significantly higher receptivity than the control model. The adaptive model, however, did not seem to perform significantly better. Those results were based on an analysis across the full study period. A day-by-day analysis, however, may provide more insights regarding whether and how the adaptive model's performance changed over the days – in short, whether it adapted well to each participant.

Hence, we added a new variable – the participants' day in study (from Day 1 to Day 21) – as an interaction effect to the generalized linear mixed effects models used earlier. To best understand and visualize the results, we plot the effects of the model types over time as estimated from the mixed effect model in Figure 3. While visualizing the trends, it is important to note that the confidence interval for the adaptive model is quite wide for the first few days, because the adaptive model was not triggered until Day 8 and hence no actual data points for the adaptive model during that period.

As the study progressed, we found that the just-in-time response rate dropped significantly for the control model ($p = 0.011$) (Figure 3a). For the static model, there was a slight downward trend, but it was not significant. For the adaptive model there was a steep upward trend with a slope of 0.0092, suggesting that the just-in-time response to adaptive model increased by almost 1 percentage-point each day; this trend was not statistically significant ($p = 0.287$). The observation – although not statistically significant – is encouraging, suggesting that the adaptive model was able to learn and personalize over time, and eventually improving the just-in-time response. In fact, after day 19, the adaptive model seems to have had higher just-in-time response rate than the static model. Further, on Day 21, the adaptive model had an increase of over 51% in just-in-time response rate as compared to Day 8.

We observe similar trends for the conversation-engagement rate (Figure 3b), with the adaptive model having a significant positive trend ($p = 0.045$), with a slope of 0.0156, which translates to a 1.56 percentage-point increase in conversation-engagement rate each day.

THE UBIQUITOUS PRESENCE OF MOBILE TECHNOLOGIES HAS ENABLED A WIDE ARRAY OF RESEARCH INTO MOBILE HEALTH (mHEALTH)

IMPLICATIONS ON JITAI DESIGN

JITAIs have six major components : a distal outcome, proximal outcomes, decision points, intervention options, tailoring variables, and decision rules [6]. Our results show that it is indeed possible to detect receptivity in real-time. Hence future studies could design JITAIs such that the intervention components and decisions rules can account for receptivity as a tailoring variable before deciding on whether to deliver an intervention.

In our work, we considered receptivity as a binary outcome, i.e., a person is either receptive or not. This is just the first step towards enabling effective delivery of interventions. We argue that receptivity is a spectrum and not an absolute yes/no. It could be possible that – in each moment – a person is receptive to a particular type of intervention and be non-receptive to a different type of intervention. Given the promising results in our study, we lay solid groundwork for future researchers to move forward to other dimensions of receptivity. The treatment of receptivity as a spectrum would enable intervention designers to decide not only *if* an intervention should be delivered but also what interventions to deliver in that moment – JITAIs could be developed that consider the degree of vulnerability (tailoring variable), the level of receptivity (tailoring variable), and the expected effectiveness of various interventions (intervention options) and decide which intervention to maximize the distal and proximal outcomes.

Although our results are promising, they are still preliminary. We had 83 users in our study who participated for only 3 weeks; most behavior change programs last longer than 3 weeks. Hence, more research is needed to evaluate model performance and how receptivity changes over a longer period. For more detailed discussions, please refer to the full paper [5]. ■

Acknowledgments
This research was supported by the NIH National Institute of Drug Abuse under award number NIH/NIDA P30DA029926, and by CSS Health Insurance, Switzerland.

Varun Mishra is an assistant professor at Northeastern University with a joint appointment in the Khoury College of Computer Sciences and the Department of Health Sciences. He directs the Ubiquitous Computing for Health and Well-being (UbiWell) group, which works on sensing and intervention systems for various mental and behavioral health outcomes. He received his PhD from Dartmouth College.

Florian Künzler is currently the Chief Compliance Officer at Nash Exchange. He received his PhD in Technology Management and Artificial Intelligence from ETH Zürich, where he worked on states-of-receptivity for JITAIs.

Jan-Niklas Kramer is the Innovation Manager at CSS Health Insurance in Switzerland. He received his PhD in Business Innovation from the University of St. Gallen, Switzerland. His interests are in building novel interventions for positive health behavior change, with a focus on physical activity.

Elgar Fleisch is professor of Information and Technology Management at the ETH Zürich and the University of St. Gallen (HSG) and Director of the Institute of Technology Management. His research interests focus on the current fusion of the physical and digital world into an Internet of Things. With his transdisciplinary team, his goal is to understand this fusion in the dimensions of technology, applications and social implications and, based on this, to develop new technologies and applications for the benefit of the economy and society.

Tobias Kowatsch is an associate professor for Digital Health Interventions at the Institute for Implementation Science in Health Care, University of Zurich, and Director at the School of Medicine, University of St. Gallen, Switzerland. In collaboration with his interdisciplinary team and research partners, he designs digital therapeutics (DTx) at the intersection of information systems research, computer science, and behavioral medicine.

David Kotz is the provost, and the Pat and John Rosenwald Professor in the Department of Computer Science, at Dartmouth College, Hanover, NH. His current research involves security and privacy in smart homes, and wireless networks.

REFERENCES

[1] David H. Gustafson, Fiona M. McTavish, Ming-Yuan Chih, Amy K. Atwood, Roberta A. Johnson, Michael G. Boyle, Michael S. Levy, Hilary Driscoll, Steven M. Chisholm, Lisa Dillenburg, et al. 2014. A smartphone application to support recovery from alcoholism: a randomized clinical trial. *JAMA Psychiatry*, 71(5):566–572.
[2] Kevin Koch, Varun Mishra, Shu Liu, Thomas Berger, Elgar Fleisch, David Kotz, and Felix Wortmann. March 2021. When do drivers interact with in-vehicle well-being interventions? An exploratory analysis of a longitudinal study on public roads. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, 5(1).
[3] Jan-Niklas Kramer, Florian Künzler, Varun Mishra, Bastien Presset, David Kotz, Shawna Smith, Urte Scholz, and Tobias Kowatsch. 2019. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: Protocol of a microrandomized trial. *JMIR Research Protocols*, 8(1):e11540.
[4] Florian Künzler, Varun Mishra, Jan-Niklas Kramer, David Kotz, Elgar Fleisch, and Tobias Kowatsch. Exploring the state-of-receptivity for mhealth interventions. December 2019. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. (IMWUT)*, 3(4.
[5] Varun Mishra, Florian Künzler, Jan-Niklas Kramer, Elgar Fleisch, Tobias Kowatsch, and David Kotz. June 2021. Detecting Receptivity for mHealth Interventions in the Natural Environment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. (IMWUT)*, 5(2):1–24.
[6] Inbal Nahum-Shani, Shawna N. Smith, Bonnie J. Spring, Linda M. Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A. Murphy. 2016. Just-in-time adaptive interventions (JITAIS) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, pages 1–17.
[7] Hillol Sarker, Moushumi Sharmin, Amin Ahsan Ali, Md Mahbubur Rahman, Rummana Bari, Syed Monowar Hossain, and Santosh Kumar. 2014. Assessing the availability of users to engage in just-in-time intervention in the natural environment. *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 909–920.