Using Adaptive Bandit Experiments to Increase and Investigate Engagement in Mental Health

Harsh Kumar[1], Tong Li[2], Jiakai Shi[1], Ilya Musabirov[1], Rachel Kornfield[3], Jonah Meyerhoff[3], Ananya Bhattacharjee[1], Chris Karr[4], Theresa Nguyen[5], David Mohr[3], Anna Rafferty[6], Sofia Villar[7], Nina Deliu[7,8], Joseph Jay Williams[1]

[1] Department of Computer Science, University of Toronto
[2] Department of Statistics, University of Toronto
[3] Center for Behavioral Intervention Technologies, Northwestern University
[4] Audacious Software
[5] Mental Health America
[6] Carleton College
[7] MRC - Biostatistics Unit, University of Cambridge
[8] MEMOTEF Department, Sapienza University of Rome

Abstract

Digital mental health (DMH) interventions, such as text-message-based lessons and activities, offer immense potential for accessible mental health support. While these interventions can be effective, real-world experimental testing can further enhance their design and impact. Adaptive experimentation, utilizing algorithms like Thompson Sampling for (contextual) multi-armed bandit (MAB) problems, can lead to continuous improvement and personalization. However, it remains unclear when these algorithms can simultaneously increase user experience rewards and facilitate appropriate data collection for social-behavioral scientists to analyze with sufficient statistical confidence. Although a growing body of research addresses the practical and statistical aspects of MAB and other adaptive algorithms, further exploration is needed to assess their impact across diverse real-world contexts. This paper presents a software system developed over two years that allows text-messaging intervention components to be adapted using bandit and other algorithms while collecting data for side-by-side comparison with traditional uniform random non-adaptive experiments. We evaluate the system by deploying a text-message-based DMH intervention to 1100 users, recruited through a large mental health non-profit organization, and share the path forward for deploying this system at scale. This system not only enables applications in mental health but could also serve as a model testbed for adaptive experimentation algorithms in other domains.

Introduction

Enhancing digital mental health interventions can contribute significantly to improved well-being and support in today's society (Boardman 2011). The global prevalence of mental health problems increased by an unprecedented 25% in 2020, the first year of the COVID-19 pandemic. This increased prevalence and disruptions to traditional mental health care emphasized preexisting gaps in access to support. Before, during, and after the pandemic, many cannot access adequate mental health support and are increasingly turning to DMH tools (Torous et al. 2020). There is an opportunity for AI to help address these challenges and enhance mental health support across various populations (Kumar et al. 2023; Malgaroli et al. 2023).

Digital Mental Health (DMH) interventions hold promise in extending the availability of support for individuals as they take steps to manage their mental health. Automated messaging is a central component of many DMH interventions as a modality for delivering support and information. However, messaging interventions typically deliver the same content across users despite the users' varied needs. This in turn reduces the relevance of such messages as well as compromises engagement and effectiveness. This is problematic because while engagement with interventions is not sufficient, it is necessary to benefit users. Moreover, building, optimizing, and evaluating multicomponent interventions is a multi-stage process that is complex and requires data-driven decisions from experts (Collins, Murphy, and Strecher 2007).

To address these challenges, we developed a system for adaptive experimentation to improve people's engagement with text messages providing a mental health intervention, using Multi-Armed Bandit (MAB) algorithms. The system was built in partnership with Mental Health America (MHA)[1], a large non-profit organization dedicated to promoting mental health and preventing mental illness through advocacy, education, research, and services. The system supports an eight-week DMH intervention in which users receive various contents via text messages to help manage their mental health (see Figure 1). To continually test out improvements, our software system instruments components of the text messages to function like intelligent agents that test out different actions (arms), by using MAB algorithms to conduct adaptive experiments.

The purpose of using this system is twofold:

- First, we are able to run algorithms in a real-world mental health setting to optimize efficacy, as has been done in
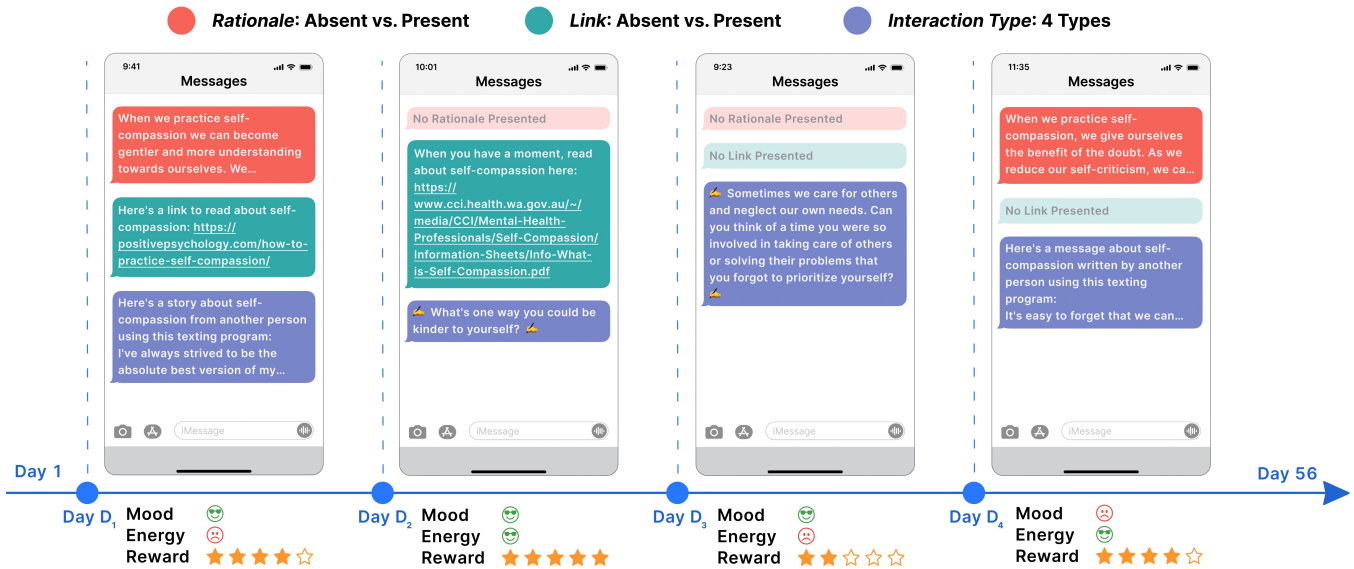
[1]https://mhanational.org/

Figure 1: Schematic representation of an example sequence of messages a user could receive during days D_i, where D_i is any random day within the 8-week-long intervention when the user receives a message. 'Mood' and 'Energy' during each day D_i represent a subset of contexts describing the user during that particular day, and 'Reward' indicates a response from the user to the question *"How helpful were these messages? Reply with a number 1 (not at all helpful) to 5 (very helpful)"* after receiving the messages during the day. The messages are composed of *three* modular components in a 2 (*Rationale*: present vs absent) x 2 (*Link*: present vs absent) x 4 (*Interaction type*: 4 options) factorial design.

product development.

- Second, this system allows us to understand how well the results of such algorithms can be analyzed by clinical scientists with statistical rigor.

For example, consider an Action (experimental) variable that is testing whether providing a rationale for using a particular psychological strategy impacts the Reward (outcome) – a measure of message helpfulness. The impact of the Action variable on the Reward may depend on contextual variables, such as self-reported Mood. Domain scientists, like clinical psychology researchers, may want to analyze such data to understand how the Action variable impacts the Reward, and how that impact is mediated by covariates or contextual variables like Mood.

Traditional experiments with uniform random assignment generate data that can be used to provide answers to these questions. However, traditional experiments do not dynamically provide a better user experience by rapidly using existing data to inform what a user will receive in the future. Using the most fundamental formulation of exploration-exploitation tradeoffs, (contextual) MAB algorithms allow for such a dynamic, adaptive experimental design. This can have the benefit of increasing the probability that users are assigned to actions that work best for them, as well as the benefit of decreasing the probability that users receive burdensome or ineffective actions.

However, there might be a trade-off between collecting data to answer domain scientist questions and using data to try to optimize for users (Yao et al. 2021). For example, research has suggested that there may be biases in estimates of

arm means, increased false positive rates, and reduced statistical power. Our system not only enables the use of contextual MAB algorithms to optimize for users, but also provides a testbed for evaluating how different scenarios can impact false positive rates, statistical power, and reward maximization.

Related Work

Digital Mental Health and Text-Message Based Interventions

Digital mental health (DMH) interventions provide a solution for narrowing the service provision gap, wherein the need for mental health support far outweighs the availability of traditional care (Gan et al. 2021). By utilizing tools like computers and smartphones, psychological treatment can become widely accessible. The majority of Americans own smartphones, and DMH interventions using this modality allow users to customize how they wish to receive and engage with mental health support. Previous work has found DMH interventions for depression to be effective across randomized control trials (Firth et al. 2017). However, user engagement levels with DMH interventions are low, barring adoption into the real world (Gan et al. 2021). Text messaging as a modality has proven to be effective in promoting behavior change. Using automated messaging, short content that is repeatedly delivered throughout the day can lay the foundation for a larger behavior change process (Abroms et al. 2015). Further, automated messaging has been successful in promoting and maintaining positive mental health (Armanasco et al. 2017).

Adaptive Experiments and Multi-Armed Bandit Algorithms

Contextual MAB algorithms are one solution for adaptive experiments. These algorithms base the probability of assignment on previous data, ensuring that more users are exposed to more effective conditions, and fewer users are exposed to less effective conditions (Li et al. 2010). Bandit algorithms have been extensively applied in real-world settings to tackle the exploration-exploitation dilemma in sequential decision-making (Fouché, Komiyama, and Böhm 2019; Zhan et al. 2021). However, there are issues that arise from MAB algorithms that complicate data analysis. Primarily, these algorithms have been shown to produce measurement errors, increases in false positive rates, and decreases in statistical power (Russo 2020). These hindrances work directly against the objective of the algorithm by decreasing the chance that effective conditions will be identified and increasing the probability that unhelpful conditions will be deployed (Rafferty et al. 2019). Therefore, it may be helpful to use systems that enable comparisons of different algorithms, such that researchers can make informed decisions about these statistical trade-offs.

Design of Intervention

A series of design workshops were conducted with non-treatment seeking participants (Kornfield et al. 2022; Bhattacharjee et al. 2023) to generate content which could be tailored according to users' preferences, backgrounds, and contexts. In the next section, we deep dive into a particular component of the intervention to show as an example how contextual multi-armed bandit algorithms are used throughout the numerous experimentation points.

Adaptive Components in DMH Intervention

Modular dialogues are brief, self-contained interactions that support a single psychological strategy, and can be delivered at one contact point without the need to reference messages earlier or later in the day. Our software system enables embedded adaptive experiments at a number of *intelligent* decision points in these modular dialogues.

Figure 1 illustrates an example of message sequences a participant may receive during an 8-week intervention to practice self-compassion. The messages contain three modular components (decision points) in a factorial design: *Rationale* (present vs. absent), *Link* (present vs. absent), and *Interaction type* (four options). Our software system enables adaptive experiments at various intelligent decision points in these modular dialogues.

We use contextual multi-armed bandit algorithms throughout numerous experimentation points to adapt the assignment of arms based on the collected data. The system captures various contextual variables to tailor messages according to users, such as user profiles, preferences, and interactions with the system during the intervention (Table 1). The coding of these variables can be adjusted as needed.

Rewards in the form of message ratings and link clicks are used to optimize content quickly. For instance, at the end of each message sequence, users rate the helpfulness of the

Variable	Type	Description
Mood & Energy	Binary	Takes 0 when users report Low or Medium levels, and 1 for High levels
K10	Ordinal	Based off of the Kessler Psychological Distress Scale. Bounded between 1 and 4.
Recent Activity Last 48 hours	Binary	Takes 0 when user is inactive in the last 48 hours of receiving the message, and 1 if user interacts in any way with the system (e.g. types something, clicks on any link, etc).

Table 1: Some user contexts captured by the system (out of 12 total contextual variables).

messages on a scale of 1 to 5. The bandit algorithm also considers whether users click the embedded hyperlinks in the messages.

The bandit algorithm adapts the assignment of action variables based on the collected data. For example, it decides whether to provide a *Rationale* for introducing a psychological strategy, include a *Link* to web content, or select one of the four *Interaction types*. The system is designed to capture rewards of varying natures to enable rapid prototyping and testing of different rewards for different contexts.

Algorithms for Adaptive Experimentation

The system supports a range of algorithms for adaptive experimentation, which dynamically update the assignment probability of each intervention based on the observed data. Below we describe the application of one of the algorithms.

Problem Formulation Assuming an experiment over a horizon of size T, the problem is to choose a sequence of T actions $\{a_t\}_{t=1,\dots,T}$ that maximize the expected cumulative reward over time $E[\sum_{t=1}^{T} r_{a_t}(t)]$, with $r_{a_t}(t)$ being the reward associated to action a_t at time t. One way to formalize the problem of dynamic decision-making about what type of message to deliver *to whom* is by using a *contextual* MAB problem (Li et al. 2010; Agrawal and Goyal 2013). In contextual MABs, the reward is conceived as a function of users' contexts, in addition to allocated arms, so it allows *personalization* as well. We formulate each dimension of the messaging protocol (framing, hyperlink, interaction type) as a separate intelligent decision point, so we solve three contextual bandit problems. Here, we illustrate the setup of a popular MAB strategy based on contextual linear TS (Agrawal and Goyal 2013).

Algorithm Suppose that the likelihood of the observed reward $r_i(t)$ of arm i at step t, given a context vector $b_i(t) \in \mathrm{R}^d$ and the unknown parameter vector $\mu \in \mathrm{R}^d$, is given by a Gaussian model with mean $b_i(t)\mu$ and standard deviation $v = R\sqrt{\frac{24}{\epsilon}d\ln(\frac{1}{\delta})}$ with $\epsilon \in (0,1)$. Denoting with $B(t) = I_d + \sum_{\tau}^{t-1} b_{a_\tau}(\tau)b_{a_\tau}(\tau)^T$ and with $\hat{\mu}(t) = B(t)^{-1}\left(\sum_{\tau}^{t-1} b_{a_\tau}(\tau)r_{a_\tau}(\tau)\right)$, if the prior at step t

is $\mathcal{N}(\hat{\mu}(t), v^2 B(t)^{-1})$, then the posterior at $t+1$ is given by $\mathcal{N}(\hat{\mu}(t+1), v^2 B(t+1)^{-1})$. In practice, at each step t, a sample $\tilde{\mu}(t)$ is be generated from $\mathcal{N}(\hat{\mu}(t), v^2 B(t)^{-1})$, and the i-th arm which maximizes $b_i(t)^T \tilde{\mu}(t)$ is played.

Evaluation of System

After extensive internal testing, we deployed this system with 50 participants (recruited from the MHA website) to undergo the eight-week intervention. The study was approved by the local university ethics board. This helped us design simulation scenarios and make adjustments to ensure data is being appropriately collected, and that the algorithms are adequately adapting the experiments. Based on data from pilot deployments, we illustrate the following simulated scenarios. The code to reproduce the following analysis is made publicly available[2].

Simulation Settings

We generate each user's context randomly (with all possible values being equally likely). To resemble our real-world scenario, where the users' feedback rating has 5 levels (see Figure 1), we simulate reward values from $\mathcal{R} = \{0, 0.25, 0.5, 0.75, 1\}$. The generating process involves first, generating a raw reward from a Normal distribution, and then rounding it to the closest value in $\mathcal{R}$. We consider 'Rationale' framing (arm $i = 1$) versus 'No Rationale' framing (arm $i = 0$) arms, and context.

Scenario 1: No arm difference In this case, the raw-reward $r_i(t)$ follows the same distribution for both arms and does not depend on the context:

$$r_i(t) \sim N\left(0.5, (1/6)^2\right), \quad i = 0, 1.$$

Scenario 2: Substantial arm difference Here, we assume that arm $i = 1$ works better than the arm $i = 0$:

$$r_i(t) \sim N\left(0.5 + 1/8 \times i, (1/6)^2\right), \quad i = 0, 1.$$

Scenario 3: Optimal arm changes based on the context We assume that the reward for each arm $i = 0, 1$ changes based on an interaction between arm and context. Denoting the interaction with $*$, this is given by:

$$r_i(t) \sim N\left(0.5 + \frac{3}{8} \times i - \frac{1}{4} \times m_t - \frac{5}{8} \times i * m_t, (\frac{1}{6})^2\right)$$

where m_t is the context variable assuming value of 1 if the t-th participant has a high mood, and 0 otherwise. According to our setting, when a participant has a high mood, the arm 'Rationale' has a higher expected raw reward (0.875 with 'Rationale' versus 0.5 with 'No Rationale'), while the opposite is true for participants with a low mood (0 with 'Rationale' versus 0.25 with 'No Rationale').

Simulation Analysis

For statistical analysis, the data collected can be used to run simulations of the algorithm's behavior if an experiment was

---

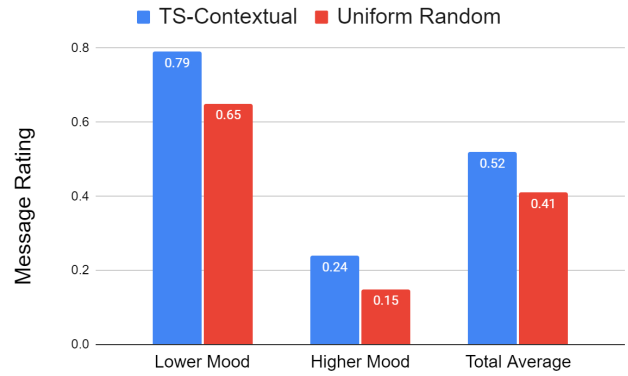[2]https://github.com/harsh-kumar9/bandit_simulation

Figure 2: Average rewards using Thompson Sampling for Contextual Bandits (Contextual TS) versus Uniform Random in different cases. The first pair of bars compares the reward in the group of participants having *low mood*, the second pair compares in *high mood* group, and the last pair takes the average among all participants.

run thousands of times. We present the Reward, False Positive Rate (FPR; the probability of a statistical test to incorrectly report an arm difference, when one does not exist), and Statistical Power (the probability of a test to correctly report an arm difference, when one exists). The hypothesis tests related to FPR and Power are conducted by sampling from the joint posterior distribution of the parameters of interest and estimating the 95% Confidence Intervals. Compared with traditional uniform random experimentation, these reveal the extent to which an algorithm's adaptive data collection increases, reduces, or has no effect on reliably detecting different kinds of effect - like whether or how much better an arm is or which arm is best based on a contextual variable.

From Figure 2, we see that Contextual TS can adapt to the data and lead to a higher reward than Uniform Random when there is a difference in the expected arm rewards. For people in a high mood, Contextual TS assigns most of them to the 'Rationale' arm, and gets 0.14 more average reward than Uniform Random; for people in a low mood, Contextual TS assigns the arm with 'No Rationale' more frequently.

Table 2 shows the FPR results in our first simulation setting (Scenario 1). The FPR for Contextual TS is almost as good as the Uniform Random algorithm when the sample size is N = 100, and it increased to 0.07 when the sample size increases to N = 1000. This is because we chose a relatively uninformative prior, which forces Contextual TS to explore more and reduces FPR. Increasing the sample size, the effect of the prior is reduced, and an exploitative algorithm like TS may allocate the two arms unevenly, impacting statistical inference ability. However, a value of 0.07 may be regarded as a good trade-off given the higher average reward. This highlights how one could further reduce the FPR by using a wider confidence interval, or choosing a more uninformative prior – making the choice to give up some reward.

Table 3 summarizes the results for Power in simulation

	N=100	N=1000
Contextual TS	0.03	0.07
Uniform Random	0.04	0.04

Table 2: False Positive Rate (FPR) in simulation Scenario 1 where there are no differences between the arms.

	N=100	N=1000
Scenario 2		
Contextual TS		
Rationale	0.72	0.97
Uniform Random		
Rationale	0.87	1.00
Scenario 3		
Contextual TS		
Rationale	0.92	1.00
Mood	0.41	0.47
Rationale * Mood	1.00	1.00
Uniform Random		
Rationale	1.00	1.00
Mood	0.95	1.00
Rationale * Mood	1.00	1.00

Table 3: Calculation of Power for Scenario 2 (substantial difference between the arms) and 3 (optimal arm changes based on context).

Scenarios 2 and 3. In Scenario 2, when the sample size is small (N = 100), the Contextual TS algorithm has 15% less Power than Uniform Random. However, when the sample size is fairly large, Power is not much of a concern, as both approaches produce Power close to 1. In Scenario 3, where the reward is a more complicated function of 'Rationale' and 'Mood', the Contextual TS algorithm has good discovery rates on detecting effects in terms of 'Rationale' and the interacting effect 'Rationale * Mood'. The only drawback for Contextual TS, in this case, is its low Power on the single effect on 'Mood'. This occurs because Contextual TS assigns most people with a high mood to the arm 'Rationale' and those with a low mood to 'No Rationale' arm, increasing Reward for both groups, whereas UR more accurately evaluates the effect of Mood across both arms.

The data can also be used to understand the detailed behavior of an algorithm over time, and reveal violations of assumptions and interesting extensions based on the structure of the data, like unforeseen subgroups, non-stationarity, or correlations between observations.

These simulations are an important part of the intervention design process. First, they help evaluate the effectiveness of adaptive algorithms for exploration-exploitation, such as MAB-based strategies, in terms of how they balance reward and rigor of statistical analysis. Second, when tuned for the specific intervention design problem, they serve as a foundation for fine-grained intervention design discussion between domain scientists, developers, and machine learning specialists, allowing exploration of multiple alternative

adaptive experimentation scenarios in mental health even before data collection.

Real-World Deployment

To evaluate the system and the intervention experimentally, we deployed the system with 1100 users recruited through an online ad on the MHA website in multiple batches. Each user consented to enroll in the 8-week program, where they received *modular* messages (visually represented in Figure 1) 2-3 times a week, as part of a larger text-message-based intervention.

We structure the analysis around two factors of the intervention, *Rationale* (present vs absent) and *Link* (present vs absent).

	Provided Response		
	No	Yes	Total
Link Present	610 (78%)	171 (22%)	781 (100%)
Link Absent	628 (80%)	160 (20%)	788 (100%)
Total	1,238 (79%)	331 (21%)	1,569 (100%)

Table 4: Response rate by arm for the factor "Link". We do not observe large differences between arms in terms of response rate

Engagement For 8,521 total arm assignments, we received 813 ratings (9.54% engagement rate), contributed by 230 unique users (20.9% of total users). Overall engagement is low, as is characteristic of DMH interventions (Gan et al. 2021). To check if the engagement depends on the experimental arms, we analyzed the rate by arms (see Table 4 for the Link arm), observing no large differences.

Efficiency of Policies Table 5 shows the summary of data for the different factors and arms of the experiment. We observe an increase in Mean rewards for both levels of each factor for Contextual TS, compared to Uniform Random assignment.

Contextual Effects Checks on potential contextual effects are difficult in the early stages due to higher power requirements for detecting interactions, but they are still important. Here we present two summary analyses for promising contextual variables. In Figure 3, we can see that Contextual

Factor		Contextual TS		Uniform Random	
		Present	Absent	Present	Absent
Link	N	232	175	202	204
	Mean	0.790	0.716	0.719	0.640
	SEM	(0.018)	(0.024)	(0.021)	(0.024)
Rationale	N	282	167	192	171
	Mean	0.736	0.728	0.710	0.703
	SEM	(0.018)	(0.025)	(0.022)	(0.024)

Table 5: Overall summary of the rewards collected for the Link and Rationale factors.
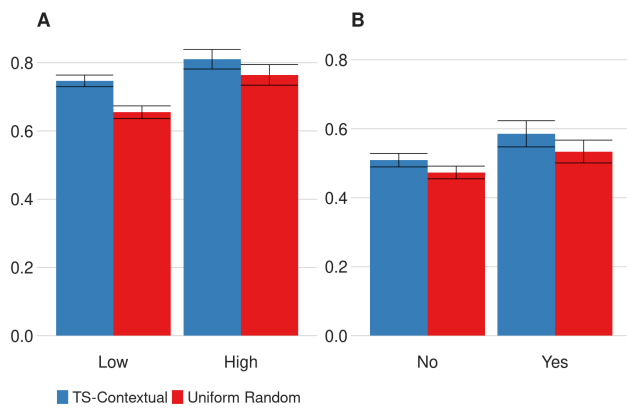
Figure 3: Average reward (rating of 1 to 5 scaled) using Contextual TS versus Uniform Random for "Link" rating for different levels of contextual variables. Figure A shows the distribution for contextual variable *Mood* (Low vs High). Number of participants (N) from left to right is [322, 316, 83, 87]. Figure B shows the distribution for *Activity in last 48 hours* (Yes vs No). N from left to right are [67, 75, 338, 329].

TS with Mood as a contextual variable was able to achieve a slightly higher average reward compared to Uniform Random. This corresponds to the results of one of the simulation scenarios we developed in the earlier stage of the analysis (Figure 2), contributing to the reinforcement of our hypothesis about the role of mood in moderating reactions. Figure 4 shows how Contextual TS adapts the assignments for Link vs No Link decisions based on users' interactions. For the Recent Activity variable, the effect is less pronounced for not recently active participants but observed for those recently active. One design decision to consider based on these results might be to revise the 48h cut-off or make it adaptive.

Discussion and Conclusion

A key insight in the design of our system was to build flexibility in how algorithms are applied. First, in open source testbed where algorithms can be directly uploaded and used in production within days[3]. Second, in allowing for algorithms to flexibly adjust which reward and contextual variables are being used and how, with rapid redeployment. For example, the current reward for the *Rationale* bandit problem is the message rating provided by the user. However, other variables could be used as reward to optimize the messages. The system also allows for new arms to be added by the social-behavioural science design team, as the adaptive experiments reveal that some arms are less effective, as we scale to 5000 users.

Our system provides a testbed for optimizing user experience in mental health while also facilitating critical analysis of algorithms, particularly understanding how algorithms balance reward maximization with data collection for statistical analysis. This places more emphasis on support-

---
[3]https://github.com/Intelligent-Adaptive-Interventions-Lab/mooclet-engine
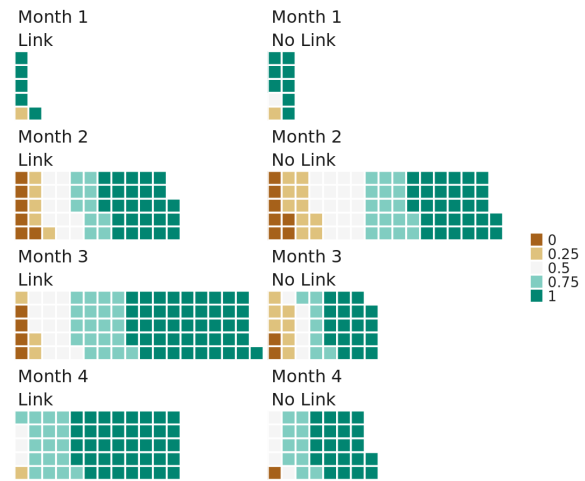
Figure 4: Arm allocation dynamics for Contextual Thompson Sampling. Columns represent arms, and grid rows artificially split experiments by approximately one-month periods, allowing to compare arm allocation in different stages of the experiment. Each small square is one reward we receive with a fill color representing how helpful the participant found the message. In the Period 1, there was a marginally better response for "No Link," leading to more allocations to this arm in Period 2. However, the algorithm was able to adjust based on responses, consistently allocating more interactions to the "Link" arm in periods 3–4.

ing domain scientists to answer socio-behavioral research questions and to draw conclusions that can be generalized to many future users, which is increasingly being emphasized in applications of bandit algorithms (Yao et al. 2021). There is a growing understanding in the behavioral science and policy communities that behavioral interventions require taking into account the heterogeneity of treatment effects (Bryan, Tipton, and Yeager 2021), including adapting to the contexts on the participant or situational level. This is also relevant for vulnerable and underprivileged populations, providing intervention designers with the tools to break the vicious circle of optimizing for the average participant. Future work can look into incorporating fairness metrics to evaluate the fairness of these algorithms when applied in a mental health setting (Wang et al. 2022; Joseph et al. 2016; Huang et al. 2022).

We showed how the algorithms trade-off optimization of reward (by giving the better arm on average or providing a personalized user experience) while collecting data that enables statistical inference. These just begin to scratch the surface of the complex real-world scenarios. Our system opens doors for future investigation of rewards that are dependent on multiple interactions between different action variables, between contextual variables, non-stationarity, when data are missing in different ways, and with repeated observations from the same person. The accompanying appendix to this manuscript is available online.[4]

---
[4]https://arxiv.org/abs/2310.18326

Acknowledgements

References

Abroms, L. C.; Whittaker, R.; Free, C.; Mendel Van Alstyne, J.; and Schindler-Ruwisch, J. M. 2015. Developing and pretesting a text messaging program for health behavior change: Recommended steps. *JMIR MHealth UHealth*, 3(4): e107.

Agrawal, S.; and Goyal, N. 2013. Thompson Sampling for Contextual Bandits with Linear Payoffs. In *International Conference on Machine Learning*. JMLR:W&CP.

Armanasco, A. A.; Miller, Y. D.; Fjeldsoe, B. S.; and Marshall, A. L. 2017. Preventive health behavior change text message interventions: a meta-analysis. *American journal of preventive medicine*, 52(3): 391–402.

Bhattacharjee, A.; Williams, J. J.; Meyerhoff, J.; Kumar, H.; Mariakakis, A.; and Kornfield, R. 2023. Investigating the Role of Context in the Delivery of Text Messages for Supporting Psychological Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19.

Boardman, J. 2011. Social exclusion and mental health–how people with mental health problems are disadvantaged: an overview. *Mental Health and Social Inclusion*, 15(3): 112–121.

Bryan, C. J.; Tipton, E.; and Yeager, D. S. 2021. Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature human behaviour*, 5(8): 980–989.

Collins, L. M.; Murphy, S. A.; and Strecher, V. 2007. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American journal of preventive medicine*, 32(5): S112–S118.

Firth, J.; Torous, J.; Nicholas, J.; Carney, R.; Pratap, A.; Rosenbaum, S.; and Sarris, J. 2017. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry*, 16(3): 287–298.

Fouché, E.; Komiyama, J.; and Böhm, K. 2019. Scaling multi-armed bandit algorithms. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1449–1459.

Gan, D. Z. Q.; McGillivray, L.; Han, J.; Christensen, H.; and Torok, M. 2021. Effect of engagement with digital interventions on mental health outcomes: A systematic review and meta-analysis. *Front Digit Health*, 3: 764079.

Huang, W.; Labille, K.; Wu, X.; Lee, D.; and Heffernan, N. 2022. Achieving user-side fairness in contextual bandits. *Human-Centric Intelligent Systems*, 1–14.

Joseph, M.; Kearns, M.; Morgenstern, J. H.; and Roth, A. 2016. Fairness in learning: Classic and contextual bandits. *Advances in neural information processing systems*, 29.

Kornfield, R.; Meyerhoff, J.; Studd, H.; Bhattacharjee, A.; Williams, J. J.; Reddy, M.; and Mohr, D. C. 2022. Meeting users where they are: User-centered design of an automated text messaging tool to support the mental health of young adults. In *CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM.

Kumar, H.; Wang, Y.; Shi, J.; Musabirov, I.; Farb, N. A.; and Williams, J. J. 2023. Exploring the Use of Large Language Models for Improving the Awareness of Mindfulness. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–7.

Li, L.; Chu, W.; Langford, J.; and Schapire, R. E. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, 661–670.

Malgaroli, M.; Hull, T. D.; Zech, J. M.; and Althoff, T. 2023. Natural language processing for mental health interventions: a systematic review and research framework. *Translational Psychiatry*, 13(1): 309.

Rafferty, A.; Ying, H.; Williams, J.; et al. 2019. Statistical consequences of using multi-armed bandits to conduct adaptive educational experiments. *Journal of Educational Data Mining*, 11(1): 47–79.

Russo, D. 2020. Simple bayesian algorithms for best-arm identification. *Operations Research*, 68(6): 1625–1647.

Torous, J.; Myrick, K. J.; Rauseo-Ricupero, N.; Firth, J.; et al. 2020. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. *JMIR mental health*, 7(3): e18848.

Wang, Y.; Ma, W.; Zhang*, M.; Liu, Y.; and Ma, S. 2022. A survey on the fairness of recommender systems. *ACM Journal of the ACM (JACM)*.

Yao, J.; Brunskill, E.; Pan, W.; Murphy, S.; and Doshi-Velez, F. 2021. Power Constrained Bandits. In *Machine Learning for Healthcare Conference*, 209–259. PMLR.

Zhan, R.; Hadad, V.; Hirshberg, D. A.; and Athey, S. 2021. Off-policy evaluation via adaptive weighting with data from contextual bandits. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2125–2135.