PopTherapy: Coping with Stress through Pop-Culture

Pablo Paredes
EECS
UC Berkeley
pablo.paredes@berkeley.edu

**Ran Gilad-Bachrach,
Mary Czerwinski, Asta Roseway,
Kael Rowan**
Microsoft Research
{rang, marycz, astar, kael} @microsoft.com

Javier Hernandez
Media Lab
MIT
javierhr@mit.edu

ABSTRACT

Stress is considered to be a modern day "global epidemic"; so given the widespread nature of this problem, it would be beneficial if solutions that help people to learn how to cope better with stress were scalable beyond what individual or group therapies can provide today. Therefore, in this work, we study the potential of smart-phones as a pervasive medium to provide therapy for the general population - "popular therapy". The work melds two novel contributions: first, a micro-intervention authoring process that focuses on repurposing popular web applications as stress management interventions; and second, a machine-learning based intervention recommender system that learns how to match interventions to individuals and their temporal circumstances over time. After four weeks, participants in our user study reported higher self-awareness of stress, lower depression-related symptoms and having learned new simple ways to deal with stress. Furthermore, participants receiving the machine-learning recommendations without option to select different ones showed a tendency towards using more constructive coping behaviors.

Author Keywords

Interventions, pervasive, mental health, stress, depression, coping, contextual, mobile, recommender, learning, behavior change, crowd therapy, stress awareness, machine learning

ACM Classification Keywords

H.5.2. User Interfaces

INTRODUCTION

Stress is considered the modern-day killer epidemic [11]. Many physiological and mental disorders are associated with stress [1, 16]. Sadly, although many people (69%) recognize that stress is a big problem, only a small number (32%) actually know how to deal effectively with it [1]. Recent discoveries have shown the importance of coping with stress in a constructive way in order to reduce its damaging effects [23]. When undergoing stress, our body experiences a series of physiological changes, colloquially named the "fight or flight" response [21]. In the past, our ancestors living in the prairies relied on this mechanism to survive threats, such as being chased by an animal, or attacked by another tribe, etc. Modern humans are exposed to many psychological stressors - such as an imminent paper deadline, an interview, an important presentation, etc. However, often there is no practical way to "fight" or "flight" from these stressors. Coping constructively with such modern stressors is in many ways a skill we have to learn.

Several theories have driven the creation of effective therapeutic interventions [6, 9]. However, despite their

efficacy, these interventions are not always efficient. Among several of the challenges, the interventions often suffer from two delivery problems: low adherence and low engagement rates. Research into how to improve these efficiency metrics is actively being pursued in the psychological community. In practical terms, the challenge to deliver effective interventions in real life can be summarized with the following question: how can we design the "right" intervention(s) to be delivered at the "right" time(s)?

In this paper we focus on the first half of this challenge, i.e., "what" should a mobile app recommend when the user needs an intervention in any real life setting; leaving the "when" as future research. We conducted a study to verify three main questions:

1. *Can we repurpose popular applications and web-sites as stress management micro-interventions?*
2. *Can the efficiency of such interventions be greatly improved by personalizing them to each individual and their context?*
3. *Can we gently move people's stress coping tendencies from destructive to constructive ones over time?*

We designed a system based on an adaptive "learn-by-doing" model that allowed us to verify these claims.

In the following sections we explain the current attempts used in the HCI community to create computerized mental health interventions. We explain our novel intervention authoring process and the resulting group of micro-interventions derived from it; we present the details of the machine-learning (ML) system, its sensor inputs and algorithms to successfully match a user's intervention request; we describe the mobile app we designed to record context data, enable an Experience Sampling Method (ESM) and deliver interventions; and we finally present study results and their implications for design and future research for recommender systems that leverage popular media.

BACKGROUND WORK

Contemporary research of technology for mental health has been mostly focused on sensing symptoms. Movements such as the Quantified Self and Wearable Computing are driving research focused on the development of new and adequate sensors that will enable clinical research. Much less research is focused on the delivery or the enablement of new therapies. We cite below a few examples of these studies, most of them extending current therapies and others exploring novel technologies.

Cognitive Behavioral Therapy (CBT) based technology

The most notable and successful use of technology for mental health is Computerized-CBT (CCBT) systems. The most relevant are: MoodGYM [26] FearFighter [27] and Beating the Blues [28]. These systems provide effective treatment even covered by health insurance in some countries. Another interesting effort is the use of gaming platforms for CBT enhancement [7]. Online technology has also been utilized for CBT-based smoking cessation [19], and recently for personalized CBT treatment [9]. An example of a

mobile system leveraging CBT concepts is the PTSD coach [29]. This app teaches patients with Post Traumatic Stress Disorder (PTSD) skills to manage their anxiety or depressive episodes.

Stress Technology Intervention R&D
Specifically focused on stress, Paredes and Chang's CalmMeNow [20] presents an experiment that measures effectiveness and usability of four interventions: social messages, breathing exercises, mobile gaming and acupressure. Among the relevant findings was a confirmation that there is a fine line between an intervention being effective and actually becoming a stressor, if applied in the wrong context. MoodWings [15] explores a wearable biofeedback design that helps people become aware of their stress to help regulate it. Maybe not surprisingly, during a driving task, stress actually went up, but driving performance improved significantly. This underlines the importance of stress as a normal reaction when performing demanding tasks, and that it is not always advisable to reduce it, but to perhaps simply keep it under control. More recently, wearable devices have been developed to help regulate breathing patterns and increase breathing mindfulness [17]. It is worthwhile mentioning also the existence of various meditation and breathing commercial applications that help people learn relaxation and mindfulness.

Unlike previous studies, the goal of this work is not to find "the best" intervention. Instead, we argue that there is not a one-size-fits-all intervention. Therefore, we present methods that allow the authoring of many interventions and matching them to individuals based on their personalities and current needs.

STRESS MANAGEMENT SYSTEM DESIGN
We designed and implemented an application for Windows Phone 8.1 and cloud based services to support the delivery of micro-interventions, providing recommendations on the interventions,

collecting user feedback and collecting contextual information. In the following sub-sections we provide more details about its main components.

Micro-Intervention Authoring System

Design Objectives
Our intervention authoring system design objectives were two: 1) maximize engagement and 2) discover online activities that could be used by the general population to reduce stress.

1) Maximize Engagement: Engagement is a key component of therapy adoption. Eysenbach's work on attrition science [10] explains the importance of understanding the differences of intervention adoption between traditional drug trials and eHealth systems. He proposes metrics that capture not only the intrinsic efficacy of the interventions, but also its usability efficiency. Additionally, Schueller's research on personalized behavioral technology interventions has shown the importance for patients to choose their own interventions as one way to improve engagement [24]. Finally, Doherty describes four strategies for increased engagement in online mental health: interactivity, personalization, support and social technology [9].

2) Stress Reduction Online Activities: Psychology research has shed some light on people's natural abilities to deal with stressful situations during daily life. Bonanno has studied people's innate characteristics to recover from stressful situations [3], Lazarus has described the strategies people use to cope [13] and the field of positive psychology studies the ways people use their strengths to reduce the impact of stress [25]. However, people deal on a daily basis with stress using simple physiological and psychological activities, such as breathing before reacting negatively, giving meaning to hardship, laughing, etc.

Therapy Group	Therapy Techniques	Group Icons and Names	Micro-intervention Samples
Positive Psychology Focus on wellness and well being, and making the positive aspects of life more salient.	- Three good things - Best future self - Thank you letter - Act of kindness - Strengths - Affirm values	☾ **Food for the Soul** (Individual) ☾☾ **Social Souls** (Social)	• Individual: **Prompt**: *"Everyone has something they do really well... find an example on your FB timeline that showcases one of your strengths."* **+ URL**: http://www.facebook.com/me/ • Social: **Prompt**: *"Learn about active constructive responding and practice with one person"* **+ URL**: http://youtube.com/results?search_query=active+constructive
Cognitive Behavioral Observe thoughts, their triggers and their consequences, entertain alternatives, dispute them, etc.	- Cognitive reframing - Problem solving therapy - Cognitive Behavioral Therapy - Interpersonal Skills - Visualization	☺ **Master Mind** (Individual) ☺ ☺ **Mind Meld** (Social)	• Individual: **Prompt**: *"Challenge yourself! Replace an unpleasant thought with two pleasant ones. Write the pleasant ones down."* **+ URL**: http://www.shrib.com/ • Social: **Prompt**: *""Try to think what new perspective these news bring to your life and share them with others."* **+ URL**: http://www.huffingtonpost.com/good-news/
Meta-cognitive Respond to ongoing experience episodes with emotions that are socially tolerable and flexible to permit spontaneous reactions or delay them as needed.	- Dialectic Behavioral Therapy - Acceptance and Commitment Therapy - Mindfulness - Emotional Regulation	♥ **Wise Heart** (Individual) ♥ ♥ **Better Together** (Social)	• Individual: **Prompt**: *"Shall we play a short game?"* **+ URL**: http://www.magicappstore.com • Social: **Prompt**: *"Write down a stressful memory of another person, imagine it flies away and disappears, and then destroy it.* **+ URL**: http://privnote.com
Somatic Exercises to shift physiological signs of arousal.	- Relaxation - Sleep - Exercise - Breathing - Laughter	🏃 **Body Health** (Individual) 🏃🏃 **Social Time** (Social)	• Individual: **Prompt**: *"Time for a quick stretch! Try some of these for a few minutes…"* **+ URL**: http://m.pinterest.com/search/pins/?q=office stretch • Social: **Prompt**: *"Cats are hilarious except when they want to eat me. Check out a few of these and show it to your friends."* **+ URL**: http://m.pinterest.com/search/pins/?q=funny cats

Table 1: Micro-interventions design matrix. Therapy groups are subdivided into individual and social groups. Each therapy group has a friendly icon and name. The micro-intervention format consists of a Prompt plus a URL.

It is reasonable then to assume that the recent trend towards mobile apps should reveal people using these apps for stress reduction activities. Some apps do share some characteristics similar to psychotherapy interventions, such as the following: keep us distracted away from our problems, record personal progress, organize thoughts, socialize, etc. Given this observation of the online world, we decided that it was worthwhile to explore the use of web apps as proxies to psychotherapy micro-interventions.

Mapping the Design Space

Our design space can be mapped as the intersection of stress management psychotherapy and popular web apps. We started by mapping the most commonly used stress management psychotherapy approaches. Then we grouped these approaches into four categories: Positive Psychology, Cognitive Behavioral, Meta-cognitive and Somatic (see Table 1). We chose this classification based on two premises: a) it corresponded to a *theoretical framework* as a good approximation to therapeutic approaches and was accepted by clinical psychology collaborators and b) it was simple enough to be *presented to users* using friendly nametags. As mentioned earlier, socialization is an element associated with improved engagement [9]. Therefore we further sub-divided our four intervention groups into interventions that could be performed alone (individual) or with or for others (social). In parallel, we mapped the top web apps [30] and top (Windows Phone) apps [31] and games [32]. We chose best rating metric as a proxy for popular/engaging apps, i.e., those with high levels of adoption and user satisfaction.

Micro-Intervention Structure

We wanted to design micro-interventions that followed some of the effective usability characteristics described by Olsen [18], i.e. a micro-intervention that could be designed by diverse *design populations* (i.e., psychologists, caretakers or even users), be used in *combination,* and *scaled up* easily. We boiled down the micro-intervention format to a minimal expression using only two components: a text *prompt* that tells the user what to do and a *URL* that launches the appropriate tool to execute the micro-intervention (see Table 1 for examples). Furthermore, we constrained the micro-interventions to be representative of one of the psychotherapy categories, and performed in a short time (approximately less than 3 minutes) to maximize usage scenarios.

Web apps and psychotherapy intersection

With these design elements in hand; we proceeded to brainstorm a long list of potential micro-interventions that mapped into the psychotherapy groups. This was a two way process; we used the psychotherapy descriptions and techniques as a guide to "harvest" activities that could be applied using popular web apps (or one of their features); and vice versa, we choose some "cool" web app (or one of their features) that could be categorized into one of the psychotherapy groups. We chose two micro-interventions per group to account for a total of 16. Table 1 shows 8 of them.

Friendly Titles and Icons

To finalize our design process, we substituted the theoretical therapy group names with "friendly" names and icons that could be accepted by the users. We wanted to avoid names that would make people feel as if they were in therapy, and rather use names that were fun and memorable. For example, we changed Positive Psychology (individual) to "Food for the Soul" and Somatic (social) to "Social Time". See Table 1 for the 8 Group Names and Icons list and Figure 1b for a screenshot. We added a one-sentence motivational slogan per group (not shown in Table 1).

Intervention Recommender System

The goal of the recommender system was to match interventions to the personal traits of each individual and their temporal context. For example, asking someone to join you for a drink of water may be an efficient coping strategy, but one may not be able to exercise it if he or she is at home by him or herself. In order to learn the matching, we proposed modeling this problem as a contextual multi-armed bandit problem [5]. In this setting, the learning algorithm tries different interventions and learns from the feedback it gets. More specifically, we have trained a model to predict the expected stress reduction of each intervention for an individual at a given context. Based on these estimates, the recommender selects an intervention by leveraging a tradeoff between *exploiting* (refining) the best interventions and *exploring* those interventions that were not used enough to gauge their effectiveness.

Input Features

The recommender system receives both user and contextual data. *User data*, such as Personality and trait data, was obtained from a pre-study survey and self-reported mood data was obtained by implementing an Experience Sampling Method (ESM) (see next section for details). Table 2 shows the user's parameters that were used. We also used the phone sensors and APIs to capture *contextual data* (see Table 3) Sensor data was collected during 5 seconds every 30 minutes. This was done to prevent battery drainage and in alignment with the operating system policies.

Output - Intervention Type Features

Five binary features were used to describe the type of intervention being selected. One feature was used as a signal to choose individual vs. social interventions and the other four features were used to select each of the four therapy groups (See Table 1).

Data Type	Parameters
User Traits	- Personality: BIG5 (agreeableness, conscientiousness, extraversion, neuroticism, openness) - Affect: Positive and Negative Affect - PANAS - Depression: PHQ-9 - Coping Strategies: CSQ - Demographics: gender, age, marital status, income, education, employment, professional level - Social network usage: Facebook usage, size of online social network and number of good friends
Self-Reports	- Last reported energy/arousal and mood value and time - Energy/arousal and mood (average and variance) - Number of self reports

Table 2. User data and their parameters.

Sensor / API	Feature
Calendar	- Number of (free, not free) calendar records (before, during and after an intervention) - Time until the next meeting
GPS	- Number of records (at home, at work, null) - Time since GPR record at work - Signal quality (average, last record) - Location (distance to home, distance to work)\ - Distance traveled
Time	- Day of the week and Time - Lunch or Night time
Accelerometer	- X, Y, Z average, variance (jerk) – 30, 120 min Number of accelerometer records (30, 120 min)
Screen Lock	- Number of events - Time since last lock event

Table 3. Sensory features collected on the phone.

Machine Learning Model

We trained ensembles of regression trees using the Random Forest algorithm [4]. After training, the model was capable of taking into account user information to predict the expected *reduction of stress* if a certain intervention was performed. We measured such reduction by calculating the delta between the subjective stress assessment (SSA) before and after the intervention. The Random Forest algorithm creates an ensemble of trees that are diversified by allowing each node to use only a subset of the available features. Since the goal of the model is to learn the differences between the interventions, the 5 intervention type features were given higher probability to be enabled in every node.

Following the Upper Confidence Bounds (UCB) algorithms [2], we used optimistic predictions using the standard deviation computed by the deviation on the leaves of the trees that conform to the random forests of the average. Furthermore, we used this score to find the intervention that is expected to reduce stress as much as possible and/or tell us more about which micro-interventions to use in the future for such purposes. We retrained the Random Forest model on a daily basis. However, we did implement incremental changes during the day to the scores on the tree leaves without changing the tree structure.

Mobile App Interface

A mobile app was designed to interface with the user, deliver the interventions and gather input data for the ML algorithms. Additionally an Experience Sampling Method (ESM) messaging interface was used to gather daily emotion self-reports.

App Design and Implementation

The design of the app was based on the following constraints: 1) support appropriate interactivity to drive engagement, 2) deliver content seamlessly, and 3) gather user and sensor data needed for the ML algorithms (see the following section). We chose a web-based format on Windows Phone v.8.1 with an embedded version of Internet Explorer v. 10 supporting HTML5. We used this system to track the metadata associated with the phone interaction and URL usage. We used a custom Azure cloud service to implement the data collection and user management modules.

User Flow

The app flow was designed using a dialogue-based schema. We presented the intervention prompts as a dialogue between the user and an agent (we chose an owl as a symbol of companionship and intelligence). We never used the word "intervention" in the dialogues, but rather the word "activity". The process followed five simple steps: 1) the user clicked on the app icon to request for an intervention; 2), the user was prompted to enter his/her current stress level (Figure 1a); 3) a micro-intervention (title, slogan and prompt) was presented (Figure 1b); 4) the interventions were all web-based and implemented in HTML5, so users simply needed to click on the play button to get to the URL page (or, depending on the condition, choose from a list of links suggested); 5) once the user experienced the intervention, they were asked to rate their stress level again.

During the first week of operation we observed a lot of unrealistic data (i.e., very high or very low stress reports with practically no time spent in interventions). We assumed this was due to people trying to take advantage of the incentives in place to encourage adoption. So we added a smaller footnote to the slider to elicit people's sense of duty (moral code) [14] in the hope this would encourage more practical use of the app (Figure 1a).

Experience Sampling Method (ESM)

ESM was used to track emotional variation during the day as an input to the ML algorithms. A two-dimensional Circumplex model of emotion [22] was used as the self-report rating interface. Users were prompted via a pop up message (Figure 2a) to use the Circumplex model (Figure 2b) and self report their emotional state approximately every 90 minutes (+/- 30 minutes) from 8am until 10pm. Users simply had to drag the circle to the quadrant that they felt they were in at that time (left to right for negative to positive valence, and bottom to top for low to high energy). A user who wanted to perform an intervention was not required to self-report their mood. However, if users wanted one intervention right after a self-report, they were prompted to do so (Figure 2c).

STUDY DESIGN

During the study we wanted to answer the three broad questions mentioned in the Introduction section. We chose four weeks as the length of our study for practical and monetary reasons. The study protocol was approved by the ethical and legal committee of the institution in which it was conducted. This section provides details about screening, experimental design and participation incentives.

Figure 1. Mobile app sample screens: a) subjective stress assessment; b) intervention title, icon, slogan and prompt
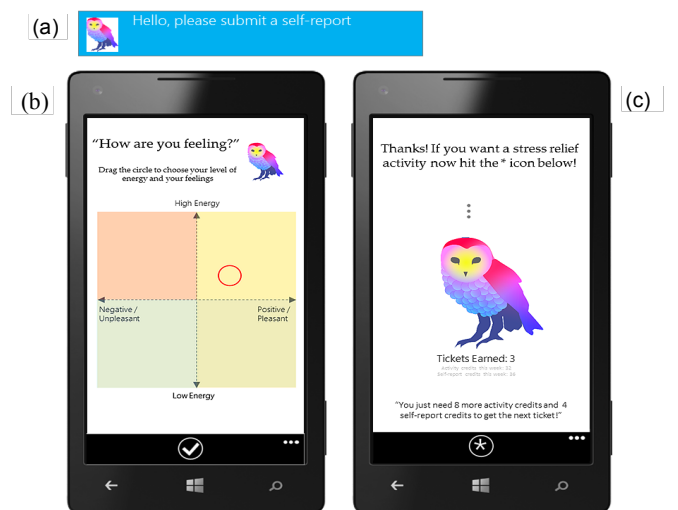

Figure 2. ESM: a) Self-report message to be clicked by user; b) Circumplex quadrant selection; c) Self-rating completion with option to launch micro-intervention if desired.

Phones and Screening Procedures

Per design, our users had to own a Windows 8 phone. In addition, all of our users were screened to be between 18-60 years old, use social media and the web. We recorded information (but did not screen) about presence of any mental illness and also whether or not other family members had been diagnosed. After screening we ended up with 95 participants (25 women), with an average age of 30. As part of the initial recruiting process, we had participants fill out validated scales for: depression (Patient Health Questionnaire - PHQ-9) [12], coping with stress (Coping Styles Questionnaire - CSQ) [8], affective states (Positive Affect and Negative Affect Scale—PANAS short) [26] and gathered demographics info.

Experiment Design

The participants were divided into 4 groups for a 2 x 2 between subjects' experimental design: *ML v. Random Recommended Interventions* and *Self-selection from a Menu or Not* (users could take the recommended intervention offered or choose from a list). Table 4 shows the different conditions with the number of participants assigned to each category and the number of interventions performed at the end of the study.

Gratuity and Incentive Policies

Participants opted into the study by accepting an email invitation after asserting that they would like to take part in the study. The email included a username and a password for downloading and installing our application, which was hidden in the Windows app store from the general public but available to our participants. Every week, for every 10 activities and 10 self-reports, each participant received a ticket to a weekly lottery of 3 x $100 gift cards; an additional ticket was awarded to participants who filled the weekly survey. On top of that, any participant that had at least 10 activities and 10 self-reports and had completed the survey on each of the 4 weeks was awarded a standard gratuity (~$300).

RESULTS

The study generated 26 days of data collection. First we present some descriptive statistics on interventions, stress deltas, drop out ratios, etc. Next, we present qualitative and quantitative results for the 20 users that used the app and filled out surveys for all four weeks, i.e. the group that completed the study in its entirety. Further qualitative analysis of the data from those users that did not complete the study is not included in this paper.

Descriptive Data

Recommendations and Selections
Figure 4 shows the distribution of the interventions recommended by the random recommender compared to the ML one. Per design, the random recommender delivered uniform recommendations (320-380 times each), while the ML converged towards recommending mostly 4 types of interventions: Social Souls, Food for the Soul, Social Time and Body Health. Indeed, one should not expect the interventions to have equal benefits for all users, this is also demonstrated in the distribution of interventions for each of the participants in the ML groups (Figure 5).

The groups who could self-select used the recommended interventions the vast majority of the time, despite having the freedom not to. The ML group used the recommendations in 97% of the cases versus 98% of the time for the random group (See Figure 6). However, it seems as if during the last 10 days of the experiment, the participants in ML/self-select group used the selection option more often. This change towards the end of the study may be explained by seeking novelty effects when the ML

became too "locked in" (i.e. stopped providing new types of interventions).

Stress Deltas
For each intervention completed, we have computed the delta between the stress reported before and the stress reported after the intervention. Figure 7 presents the average stress delta for the different groups on a daily basis. A paired t-test without the assumption of uniform variances was carried out for the post-pre stress deltas, t(46)=2.06, p(one-tailed)=0.02. Users in the ML group reported significantly greater differences in stress reduction.

	Random Choice	ML recommendation
Cannot self-select	22 users (23.1%) 1307 interventions (24%)	21 users (22.1%) 1176 interventions (22%)
Can self-select	26 users (27.4%) 1444 interventions (26%)	26 users (27.4%) 1550 interventions (28%)

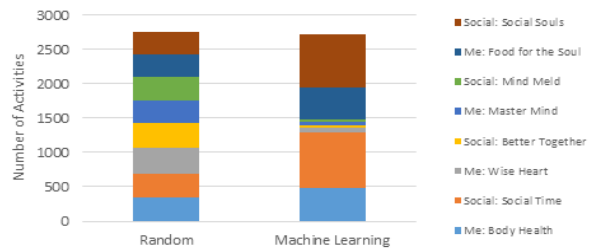**Table 4. Distribution of participants and interventions for the different conditions of the study**

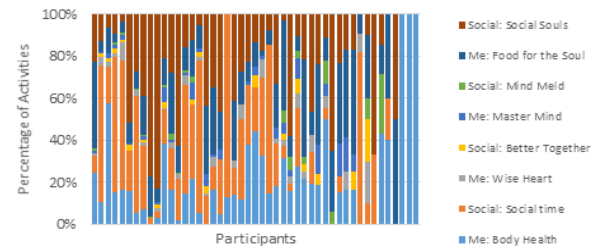**Figure 4. Machine Learning selected interventions**

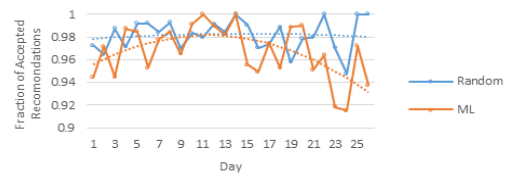**Figure 5. Distribution of interventions per participant**

Figure 6. Fraction of interventions for which users selected the recommended intervention
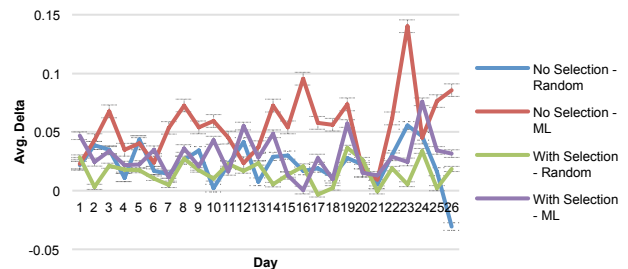

Figure 7. Stress Deltas (Stress before – Stress after)

The group that received the machine learning interventions with no self-selection had the largest delta on most days. In particular, the average delta for this group (0.054) is greater than the other groups (0.016-0.021) and even greater than any single intervention (0.04). This shows that the use of the machine-learning algorithm increased the effectiveness of the interventions by doing better matching of the interventions to the participants and their context.

Drop-out
In terms of number of unique users per day, we saw a steady decline during the experiment; however, we did not notice large differences between the different groups (Figure 8).

"Ideal" Users
During the study introduction we prompt users to use the app whenever they felt stressed. So, we were intrigued by a couple of "abnormal" behaviors: a) extremely short intervention usage time (< 3 sec) with very high (~1) or very low self-rating scores (~0); and b) reporting low pre-intervention stress levels (< 0.5), i.e., maybe not being stressed in the first place. We believe that part of this behavior could be explained by users trying to take advantage of the incentives. We observed "ideal" users, i.e. those with stress > 0.5 and using interventions for more than 60 seconds. Figure 9a shows that stress reduction is increased with a higher stress precursor. The black line represents the "no effect" line, i.e., stress delta = 0. The red line shows the average stress delta. Users that were not stressed reported lower gains in terms of stress relief (avg. delta = 0.037) than stressed people (avg. delta =0.096). We ran a 2 x 2 (ML or Random vs. stress <0.5 or stress >0.5) RM-ANOVA. There was a significant effect of having stress >0.5 before an intervention, $F(1,18)=21.6$, $p<0.001$. No other effects or interactions were significant. In other words, having a stressful precursor resulted in a larger reduction of stress after performing a micro-intervention. Additionally, Figure 9b shows that interventions with a usage time larger than ~200 sec offered diminishing results, showing a preliminary indication that longer interventions could have less efficacy. This is an interesting marker for the suggested optimal length of the interventions of the type we chose to use.

Qualitative Results
We asked a series of questions to gather information about user's interaction with the interventions and their learning process.

Subjective Data
We obtained ratings of what users considered to be the intervention they liked and disliked the most, and the interventions they thought were or less effective. Figure 10 shows the comparison of each aggregated number of counts for all the weeks. Clearly, Body Health (somatic-individual) received the highest scores for effectiveness and likeability, closely followed by Food for the Soul (positive psychology-individual), Social Time (somatic-social), and Better Together (meta-cognitive-social). The lowest rated were the cognitive-behavioral ones, Master Mind and Mind Meld. The positive psychology groups, Food for the Soul, and Social Souls, although liked, received low grades on perceived efficacy. It is also interesting to note that the most rated interventions align with the ML recommendations.

Awareness and Coping Learning
Most users reported higher levels of stress awareness due to the use of the app. Comments like: *"Although I did not do a good job of using the app this week, by using it in weeks past I am still aware of when I become stressed and try to deal with it"* or *"Doing the study helped me spotlight it"* showcase the way people

extrapolated the benefits of the study beyond the use of the app. A number of participants also reported having learned that simple methods can help manage stress if performed regularly. Comments like: *"I breathe and take time for myself to clear my mind"* or *"(I) take time to take care of my body and soul"* showcase the way some people found inspiration in the micro-interventions to do something about stress.

Table 5 shows the answers to the question *"What have you learned from this study?"* 70.3% of the users reported a higher stress self-awareness; however, it is interesting to observe that 34% reported stress awareness as stressful. 65.6% reported having learned simple ways to control stress. A paired t-test without the assumption of uniform variances was carried out for the post-pre stress deltas, $t(46)=2.06$, $p(one-tailed)=0.02$. Users in the ML group reported significantly greater differences in stress reduction.
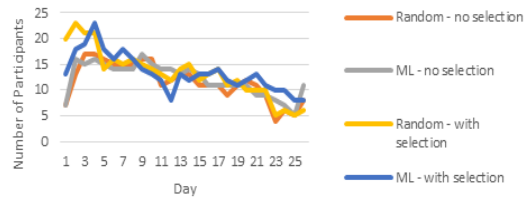
Figure 8. Daily users per day per experiment condition
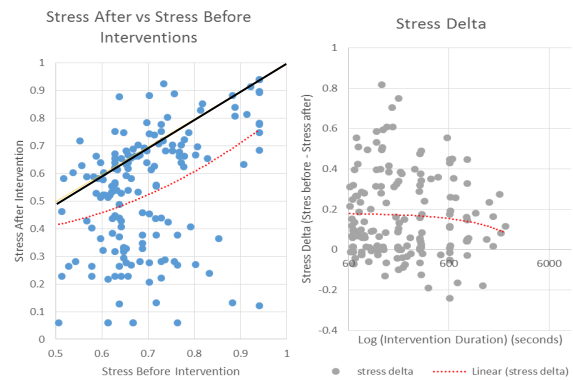
Figure 9. "Ideal" users (stress self-rating before intervention >0.5 and intervention duration >60 sec) a) Stress after vs. stress before interventions; b) stress delta vs. intervention duration
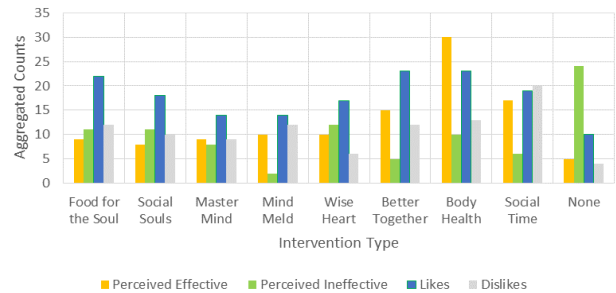
Figure 10. Subjective ratings (likeability and efficacy)

"What have you learned from this study?" (Multiple choice question)	%
To be more aware of my stress levels	70.3%
Simple ways to control my stress	65.6%
That being more aware of my stress level is stressful	34.4%
Nothing	7.8%
Other	4.7%

Table 5. Reported Learning

Quantitative Results

Depression – PHQ9

The PHQ-9 response data was analyzed for the 20 participants who used the app all 4 weeks. A 2 (ML or not) x 2 (Selection or Not) RM-ANOVA was carried out on the average score values for the initial survey week (pre-application baseline) and the 4 weeks after using the app, for 5 replications. A significant effect of week, $F_{(4,76)}=2.9$, $p=0.026$, was found, and ML was borderline significant, but no effect was observed for the Selection variable. When the data was collapsed across the Selection variable, there was a borderline significant effect interaction for week x ML ($p=0.06$). This means that, regardless of ML condition, participants showed statistically significantly less depression level (DL) while they used our tool. In addition, the ML group added to this improvement more than Random selection (borderline). In clinical terms, ML condition users ended week 4 with no signs of depression (DL < 5), while the Random condition ones showed mild depression symptoms (5 < DL < 10) [12] (Figure 11).

Coping – CSQ

A 2 (Random v. ML) x 2 (No selection v. selection from a menu) x 4 (week) RM ANOVA was performed on the differences between constructive and destructive coping behaviors to see if our participants were *learning and incorporating* new coping strategies via our interaction tool. A significant 3 way interaction was observed, $F_{(1,16)}=4.4$, $p=0.003$. No other significant effects emerged from the analysis. While a 3-way interaction can be difficult to understand, as observed in Figure 12, the group with ML without selection reported significantly more constructive (positive) coping behaviors over time. This is an encouraging finding as it indicates that those users were willing to trust the personalized intervention offered by the ML algorithm by using it and indicating greater stress relief.

Overall, results support the hypotheses (see Intro); a *recommender* system was able to deliver a suite of *web apps* that helped people reduce stress locally and *learn to cope* with stress over time.

IMPLICATIONS FOR DESIGN

Several design implications were identified during the study and can be presented in two groups: intervention design and intervention recommendation.

Intervention Design

Simplicity & Friendliness

Our simple intervention format (short prompt + familiar URL) and our friendly interaction style motivated people to use the app repeatedly throughout the day. Furthermore, the proposed format can easily scale up to other interventions that may work better for different people and/or situations. A limitation of complex formats is that people lose track of the task, loose interest, or find it harder to attribute their gains to the system. Complexity can also result in lower adoption rates and/or a more difficult learning process.

Small Size

As shown in the results, shorter interventions may result in higher stress efficacy gains. This could be due to the fact that people do not have more time to engage in a stress reducing activity; so longer ones may actually induce anxiety. Short interventions are also easier to use in more contexts. Despite their short duration, the long-term effect in depression and coping shows us that micro-interventions can result in long-term behavior change. Further work may focus on the optimal intervention time depending on the person and the context.

Incentives

A limitation of a usage-based incentive system is that users end up using the system without needing it. One way to overcome this is to make the app public so that no incentives are needed for its use. The limitation of this model is that an advertisement campaign should be properly crafted to drive initial awareness and adoption.

Stress Awareness

As reported, stress awareness was a driver for people's use of the application. However, it was itself a source of stress to 1/3 of the users. This is a limitation of ESM as a source of data. Tailoring self-reporting frequency, or even eliminating its use, should be considered when developing stress management systems.

Intervention Recommendation

Ensuring Novelty

Our experiment suggests that using ML helps in matching interventions to the user's context and hence, improves the overall outcomes for users, in terms of stress level depression and coping behaviors (for the non-selection users). There are several ways in which the ML algorithm can be improved. For example, the algorithm reduced the diversity of the interventions sent to the participants. This might have resulted in boredom, and in the long run, might lead to a high attrition rate. This may be improved by increasing the number of interventions in each group and adding diversity as an objective to the ML algorithm. Another approach would be to use periodic surveys to update participants' models, which can lead to changes in the types of interventions presented.

Exploration vs. Exploitation

The ML algorithm presented here addressed the problem of exploration (searching for new options) vs. exploitation (refining existing procedures); however, the group that was given random interventions did most of the exploration. In a sense, the design of the experiment dictated that, for at least 50% of the time, the model was exploring. This was needed in order to validate our assumption that the ML matching algorithm works for the intervention selection problem. However, now that we have validated this conjecture, in future studies, one may not wish to use 50% of the interventions for exploration.

Targeting "Ideal Users"

As described, targeting "ideal" users, i.e., people aware of their need for a stress management recommender and who are able and willing to use it should increase the local effect of interventions.
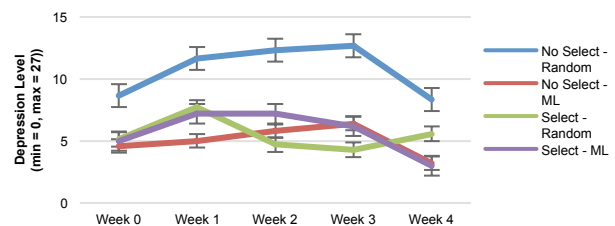

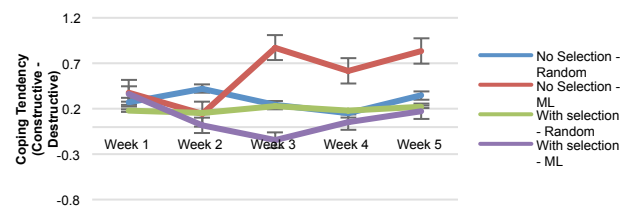**Figure 11. Depression (PHQ-9) quantitative results**

Figure 12. Coping Styles Questionnaire (CSQ)

ML algorithms could adapt weights for these users' inputs. Targeting populations that need stress management could teach us more about the efficacy of the recommender system and the interventions themselves. However, the challenge remains to create systems that help prevent stress in the general population.

Future Research

As mentioned in our introduction, the challenge to deliver effective interventions in real life was framed as: how can we design the "right" intervention(s) to be delivered at the "right" time(s)? Many interesting questions still remain in terms of "what" interventions should be delivered. In a new iteration of this system, we plan to explore the authoring problem. We want to explore crowd and self-authoring as direct sources of new interventions and social media data mining as an indirect source. We will further explore the types and duration of the interventions as a factor of adoption, as well as a new variation of the app based on complementary qualitative analysis of elements such as the mascot, the interaction with the experience sampling method, the flow, among others. With regards to "when" is the right moment to intervene, we plan to do experiments where we use psycho-physiological sensors to trigger the interventions. We want to study not only if the sensors can determine the best time to intervene, but also if they drive awareness and motivation in users.

CONCLUSION

In this paper we have shown the potential for popular web apps to provide an "unlimited" source of not only inspiration, but also actual stress management interventions. We showed that ML algorithms could be used to improve engagement and local efficacy by matching the right intervention to the context of the user. Finally we observed a tendency from users to adopt constructive coping strategies, not only by using the interventions suggested, but also by understanding that simple activities can actually help them to manage their stress. We find these results encouraging with regards to continuing research to enable "popular" therapies, mechanisms to assist large populations to cope with daily stress and drive sustained behavior change.

ACKNOWLEDGMENTS

REFERENCES

[1] American Psychological Association, "Stress in America 2010", http://www.apa.org/news/press/releases/stress/

[2] Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. Machine learning, 47(2-3),235-256.

[3] Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events?.American psychologist, 59(1), 20.

[4] Breiman, L. and Cutler, A. Random forests-classification description. Department of Statistics, Berkeley

[5] Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. arXiv preprint arXiv:1204.5721.

[6] Carson, J., & Kuipers, E. (1998). Stress management interventions.Occupational stress: Personal and professional approaches, 157-74.

[7] Coyle, D., McGlade, N., Doherty, G., & O'Reilly, G. (2011, May). Exploratory evaluations of a computer game supporting cognitive behavioural therapy for adolescents. In Proc. of SIGCHI Conference on Human Factors in Computing Systems (pp. 2937-2946). ACM.

[8] Davis, M., Robbins, E., McKay, M. (2000) The Relaxation and Stress Reduction Workbook, 5th Edition. Adapted from Boyers, J., (1999). "Coping Styles Questionnaire" ©, Kaiser-Permanente Medical Center and Health Styles, Santa Clara, CA

[9] Doherty, G., Coyle, D., & Sharry, J. (2012, May). Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In Proc. of the 2012 ACM annual conference on Human Factors in Computing Systems (pp. 1421-1430). ACM.

[10] Eysenbach, G. (2005). The law of attrition. Journal of medical Internet research, 7(1).

[11] Kalia, M. (2002). Assessing the economic impact of stress [mdash] The modern day hidden epidemic. Metabolism, 51(6), 49-53.

[12] Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. Psychiatry Ann., 32(9), 1-7.

[13] Lazarus, R. S., & Folkman, S. (1984). Stress, appraisal, and coping. Springer Publishing Company.

[14] MacLean, D., Roseway, A., & Czerwinski, M. (2013, May). MoodWings: a wearable biofeedback device for real-time stress intervention. In Proc. of the 6th International Conference on Pervasive Technologies Related to Assistive Environments (p. 66). ACM.

[15] Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. Journal of marketing research, 45(6), 633-.

[16] McEwen, B. S., & Stellar, E. (1993). Stress and the individual: mechanisms leading to disease. Archives of internal medicine, 153(18), 2093.

[17] Moraveji, N., Olson, B., Nguyen, T., Saadat, M., Khalighi, Y., Pea, R., & Heer, J. (2011, October). Peripheral paced respiration: influencing user physiology during information work. In Proc. of the 24th annual ACM symposium on User interface software and technology. ACM.

[18] Muñoz, R. F., Lenert, L. L., Delucchi, K., Stoddard, J., Perez, J. E., Penilla, C., & Pérez-Stable, E. J. (2006). Toward evidence-based Internet interventions: A Spanish/English Web site for international smoking cessation trials. Nicotine & Tobacco Research, 8(1), 77-87.

[19] Olsen Jr, D. R. (2007, October). Evaluating user interface systems research. In Proc. of the 20th annual ACM symposium on User interface software and technology (pp. 251-258). ACM.

[20] Paredes, P. and Chan, M. (2011) CalmMeNow: exploratory research and design of stress mitigating mobile interventions. SIGCHI 2011 Extended Abstracts on Human Factors in Computing Systems. ACM

[21] Richardson, K. M., & Rothstein, H. R. (2008). Effects of occupational stress management intervention programs: a meta-analysis. Journal of occupational health psychology, 13(1), 69.

[22] Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39, 1161-1178

[23] Scheier, M. F., Weintraub, J. K., & Carver, C. S. (1986). Coping with stress: divergent strategies of optimists and pessimists. Journal of personality and social psychology, 51(6), 1257.

[24] Schueller, S. M. (2010). Preferences for positive psychology exercises. The Journal of Positive Psychology, 5(3), 192-203.

[25] Seligman, M. E., & Csikszentmihalyi, M. (2000). Positive psychology: an introduction. American psychologist, 55(1), 5.

[26] Watson, D., Clark, L. and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales." Journal of personality and social psychology 54.6

[27] https://moodgym.anu.edu.au/welcome

[28] http://www.fearfighter.com/

[29] http://www.beatingthebluesus.com/

[30] http://www.ptsd.va.gov/public/pages/PTSDCoach.asp

[31] http://www.alexa.com/topsites

[32] http://www.windowsphone.com/en-us/store/top-rated-apps

[33] http://www.windowsphone.com/en-us/store/top-rated-games

[34] http://www.windowsphone.com/en-us/store/top-rated-games