Towards PopBots: A Suite of Conversational Agents for Daily Stress

Pablo Paredes
Stanford School of Medicine
Stanford, CA 94305, USA
pparedes@stanford.edu

Nikhil Gowda
Alliance Innovation Lab
Silicon Valley, CA 95051, USA
nikhil.gowda@renault.com

Nantanick Tantivasadakarn
Symbolic Systems Department
Stanford, CA 94305, USA
nantanic@stanford.edu

Marco Mora-Mendoza
Computer Science Department
Stanford, CA 94305, USA
marcom3@stanford.edu

Grace Hon
Emmanuel Thierry Lincoln
Stanford School of Medicine
Stanford, CA 94305, USA
gracehon@stanford.edu
lincolnt@stanford.edu

Matthew Louis Mauriello
Stanford School of Medicine
Stanford, CA 94305, USA
mattm401@stanford.edu

If you wish to discuss the research, please contact the Pervasive Wellbeing Technology Lab at Stanford University by emailing: pparedes@stanford.edu

Workshop 33: Conversational Agents for Health and Wellbeing
CHI'20, April 25–30, 2020, Honolulu, HI, USA

Abstract
Mobile users are increasingly using chatbots to access services. Chatbots that understand user problems and emotions could be effective public health tools for stress management for those without professional support. However, stress management applications (including chatbots) have thus far been met with low adoption and high abandonment. In this workshop paper, we explore if short interactions with multiple chatbots can have wellness benefits and propose the creation of a suite of shallow chatbots—called *Popbots*—aimed at providing *in-situ* support for daily stressors by being quick, readily available, and engaging. To evaluate the feasibility of our approach and gather preliminary user feedback, we conducted an exploratory Wizard of Oz study (N=14), developed a prototype system, and evaluated this prototype in an online pilot (N=14) with university students. Based on data from 81 conversations, we note that users most often discussed issues related to "work" and self-reported that the system helped alleviate stress.

Author Keywords
Conversational agents, Virtual agents, Chatbots, Therapy, Stress relief, Stress management

CCS Concepts
•**Human-centered computing** $\rightarrow$ **Human computer interaction (HCI);** User studies;

Chatbot ("-bot")
Doom: Worst Case Scenario
Sherlock: Problem Solving
Glass-half-full: Positive thinking
Sir Laughs-a-: Humor
Treat-yourself-: Self love
Dunno: Distraction
Checkin: Checking in

Table 1: Chatbots & methods.

Doom-bot
*- Tell me more details about **problem**?*
- I'm sorry to hear that. What are you most afraid might happen as a result?
- Alright, on a scale of 1 to 10, 1 being impossible, 10 being certain, how likely is this scenario?
- Alright, in the case that this happens, what could you do to get back on track?
- Cool, looks like you have a plan B. Just remember, even though you cannot control everything, there is a way to get back on your feet.

Table 2: Example chatbot script

Introduction

Around 60-80% of primary care visits are due to psychological stress, but only 3% receive stress management advice [12]. Moreover, there is limited infrastructure geared towards preventative health and stress management. This has spurred growth of mental health applications which currently account for 29% of the health application market [3]. Nevertheless, lack of empirical validity affects their adoption and adherence [9]. Toward maximizing adoption, massively used messaging services are being leveraged to build conversational interfaces, or "chatbots", that create scalable health solutions. However, two key issues limit the design of effective therapeutic chatbots: lack of conversational datasets and lack of time for stress management - users may not have time for lengthy exchanges [2].

Toward addressing these limitations, we explore if short conversational interactions (*i.e.,* 2 minutes) can have wellness benefits using a suite of shallow interactive chatbots—called *Popbots*. Each *Popbot* is designed to converse with users about daily stressors (*e.g.,* deadlines, social interactions) using different coping techniques. Prior work has shown that users receiving micro-interventions report higher self-awareness of stress, lower depression symptoms, and having learned new ways to deal with stress [13]. We build off this approach and implement chatbots based on techniques drawn from cognitive behavioral therapy (CBT) [5], positive psychology [15], and self-compassion [1], which teach people how to recognize sources of stress, change negative behavioral reactions, and reframe thoughts. The aim of this system is to mitigate *in-situ* stress and teach skills through short conversations that: (i) are engaging, (ii) scalable over most text-based mediums, and (iii) personalizable—allowing opportunities to explore different ways of solving stressors. As early work, we ask: *what challenges and benefits do users perceive about our approach?*

Feasibility Study

To examine the feasibility of our proposal, we conducted a Wizard of Oz (WoZ) [6] study with follow-up interviews. This study used a subset of chatbots from Table 1 and allowed us to explore: (i) preferences around using suites of chatbots versus singular chatbot apps, and (ii) the types of stressors, if any, users would be willing to talk to chatbots about. The study lasted 3-days with participants meeting an experimenter daily for in-person sessions. We recruited 14 participants (6 female, 1 preferred not to identify; age range 18-50). Each participant was randomly assigned to either a *Multiple* condition that had three chatbots (*i.e.,* Positive Thinking, Worst Case Scenario, and Problem Solving), or a *Single* condition that contained only the Problem Solving bot. Participants in the *variable* condition were matched with chatbots using latin squares randomization [8].

We used tlk.io [7], a web chat interface that allows the creation of open chat channels, to create the interface between participants and experimenter. Participants believed they were interacting with a chatbot, while in reality they were interacting with an experimenter who was following conversation scripts created beforehand (e.g., Table 2). The participant was instructed to type a greeting in channel which cued the experimenter to start following the script.

During each session, participants had a single conversation. At the end of the conversation they rated how *helpful* and *enjoyable* they found it. After 3 sessions, participants completed a post-study questionnaire about perceived efficacy in stress reduction, usability, and use cases. Four participants were contacted for semi-structured interviews—two from each experiment condition. In each pair, one individual evaluated the chatbot as effective and the other found the chatbot ineffective. They were queried on their perceptions about the *Popbots*, and completed a card sort-

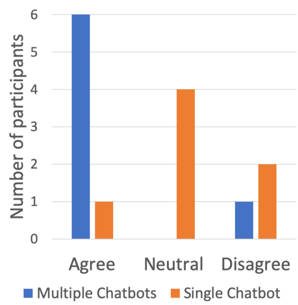

Figure 1: Perceived stress reduction for *Multiple* and *Single* chatbot conditions

Figure 2: Participant is welcomed by the *Popbots* and provides a stressor that is then interpreted by the system's state handler.

ing task placing a variety of stressors into buckets (*i.e.,* chatbots and their coping strategies).

WoZ Results
Preliminary analysis of the data shows a differences in self-reported stress between conditions (Figure 1). There was a higher perception of stress reduction among participants in the *variable* chatbot condition (blue, left is better), which helps motivate our suite approach to designing a chatbot system for daily stress. When requested to match stressors to chatbots among the *variable* condition users, we observed some general trends: (i) All participants (7/7) wanted to use the *Positive Thinking* chatbot for romantic breakup, (ii) all (7) wanted to use the *Problem Solving* chatbot for interpersonal conflicts with friends or coworkers, and (iii) most (6) wanted to use the *Worst Case scenario* chatbot for dealing with anxiety before an exam.

Follow-up Interview Results
Unexpectedly, most participants (3/4) were interested in using chatbots for coping with daily stressors even when support from humans was available. The objectivity, ease of use, and privacy chatbots offered was appealing for situations like: illness and injury, financial stress, interpersonal relationships, and social isolation. Participants believed that chatbots would provide a more effective solution because the *PopBots* provide quick therapy solutions on the spot. For example, one participant stated *"I'd rather talk about these [problems] in the void...and have a computer interact with me quickly."* All participants expected chatbots to have human-like characteristics like a typing delay despite being aware that chatbots can respond faster corroborating prior work on the mirroring of non-verbal [4], conversational cues [11], and personality traits [14]. Additionally, one participant described using multiple chatbots in sequence to help with finding appropriate solutions for complex stressors, e.g. use Positive Thinking to reduce anxiety first, and then Problem Solving to take care of the underlying problem.

Field Experiment
Based on the prior promising results, we implemented a system to be tested in the wild using Telegram™ [10]—a data-security compliant messaging platform, using a Python™ backend and a MongoDB® database to log interactions. When the user greets the chatbots, they receive a friendly message in response asking them to describe their current stressor and then a chatbot is recommended (Figures 2 - 3). User responses and data from past interactions are used to chose the chatbot. At the end the conversation the chatbot thanks the user and asks for feedback on whether the intervention helped reduce their stress.

System Evaluation. We recruited 14 students between the ages of 18-24 (79% Female, 21% Male). We collected: (i) daily surveys sent at 8pm which asked users to rate their daily stress levels, sleep quality, social interactions, and (ii) usability questions and chatbot preferences. We compensated participants for completing pre/post-study surveys, chatbot usage, and completion of the daily survey (M=$33.57, Mdn=$29.5, SD=$10.2). A total of 81 conversations were obtained.

Pilot Results.
Usage. Most participants used the *PopBots* at home (75%), due to it being the time they use their phones the most or have settled down. In addition, the bots were mostly used while participants were alone (67%). We observed two ways that users reported their stressors to the chatbots. First, a majority of participants (74%) tended to type out full sentences. For example, participants wrote sentences like *"Having to go to work tomorrow"*, *"My presentation that's coming up"*, and *"My friend being mad at me"*. An-

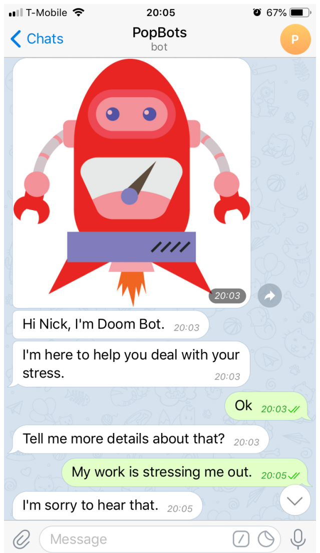

Figure 3: Participant interacts with a recommended bot: Doom-bot, the Worst Case Scenario chatbot

Stressor	# (%)
Work/School	30 (38%)
Relationships	18 (23%)
Physical Pain	4 (5%)
Travel	3 (4%)
Finance	2 (3%)
Decision making	2 (3%)
Self Doubt	2 (3%)
Other	4 (6%)
Non-stressors	16 (21%)

Table 3: Categories of Stressors based on conversations (N=81).

other common way (26%) participants reported stressors was to type out only keywords (*e.g., "money", "car", "family"*). Participants interacted with the bots for 1.95 minutes (*SD*=2.53) on average, which suggests that most users did not ponder over the conversations for very long. While many participants used the chatbots around 8 pm, numerous conversations occurred outside this time frame, suggesting the system was being used as intended—for coping with daily stressors that occurs any time throughout the day.

Effectiveness and Stressors. Most users found the chatbot conversations helpful or neutral (39.56% helpful, 26.92% neutral, 20.33% unhelpful, and 13.19% other). They also reported using our system's [switch] command to change bots and explore coping strategies. For example, one user wrote *"I wanted to try chatbots that I hadn't used before"*. Nine categories of stressors were identified (see Table 3) in the conversations. Stressors reported were mostly school or work related (*e.g.*, "Classes") followed by those of an interpersonal nature (*e.g.* "I'm worried that I've gotten myself into an awkward situation"). Non-stressors were submitted, which we attribute to the compensation scheme.

Design Challenges

In this workshop paper, we proposed a new approach to applying conversational interfaces to manage stress that involves a suite of shallow (*i.e.,* short duration) chatbots. We found evidence that users would use the system for certain stressors and that they would do so even if human support was available. These observations motivate continued development though there are many challenges including:

Parsing Stressors. Users prefer to use a chatbot if it is capable of understanding the underlying stressor versus providing a generic answer (*e.g.,* feeling sorry about their stress). Adding an intent detection system with machine learned models of conversations around daily stress would likely help to increase empathy and efficacy.

Bot matching. We expect users to explore across bots, but matching bots to stressors and early interactions matter. Users with a bad "first impression", for example, do not want to explore. We plan to generate better than random matches by creating stressor-intervention diad datasets using crowdsourcing methods.

Understanding Intents. User satisfaction with the pilot system was low due to technical issues with open text resulting in poorly matched or broken conversations. For example, the current implementation of the chatbots is only capable of detecting keywords or responses to simple yes/no questions. We plan to add intent recognition for certain parts of the conversation, and convert certain interactions to predetermined choice selectors. The key is to balance design simplicity with data-driven models.

Authoring for Novelty. To address the reduced interest in the system and habituation that resulted from the small number of available chatbots, we plan to allow health professionals and users to author chatbots to increase the number and variety in the suite. This means creating authoring tools that allow people from diverse backgrounds to create chatbots.

Triaging and escalation. An important challenge is creating an algorithm to detect whether the *PopBots* are capable of handling a stressor and escalating users to additional resources if needed (*e.g.* a human specialist, 9-1-1).

In the short term, we plan to: (i) further explore our data, (ii) evaluate broken conversation recovery, and (iii) develop a dataset of daily stressors and interventions. Ultimately, we aim to create an online learning recommendation system that pairs users to *PopBots* given stressor and context.

Acknowledgements

Work supported, in part, by the Stanford Institute for Human-Centered Artificial Intelligence and Groupe Renault.

REFERENCES

[1] Ashley Batts Allen and Mark R Leary. 2010. Self-Compassion, stress, and coping. *Social and personality psychology compass* 4, 2 (2010), 107–118.

[2] American Psychological Association. 2013. Are Teens Adopting Adults' Stress Habits? (2013).

[3] Emily Anthes. 2016. Pocket psychiatry: mobile mental-health apps have exploded onto the market, but few have been thoroughly tested. *Nature* 532, 7597 (2016), 20–24.

[4] Jeremy N Bailenson, Jeremy N Bailenson, Nick Yee, and Nick Yee. 2005. Digital Chameleons. *Society* 16, 10 (2005), 814–819. DOI:http://dx.doi.org/10.1111/j.1467-9280.2005.01619.x

[5] Aaron T. Beck. 1970. Cognitive therapy: Nature and relation to behavior therapy. *Behavior Therapy* 1, 2 (1970), 184–200. DOI:http://dx.doi.org/10.1016/S0005-7894(70)80030-2

[6] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies—why and how. *Knowledge-based systems* 6, 4 (1993), 258–266.

[7] Anthony Dmitriyev and Simon. 2019. tlk.io. (2019). https://tlk.io

[8] Lei Gao. 2005. Latin Squares in Experimental Design. *Michigan State University* (2005), 1–19. http://www.mth.msu.edu/

[9] IMS - Institute for Healthcare Informatics. 2015. Patient adoption of mHealth: Use, evidence and remaining barriers to mainstream acceptance. (2015).

[10] Telegram Messenger Inc. 2019. Telegram Messenger. (2019). https://telegram.org/

[11] François Mairesse and Marilyn A. Walker. 2008. Trainable Generation of Big-Five Personality Styles through Data-driven Parameter Estimation. Columbus.

[12] Aditi Nerurkar, Asaf Bitton, Roger B Davis, Russell S Phillips, and Gloria Yeh. 2013. When physicians counsel about stress: results of a national study. *JAMA internal medicine* (Jan 2013). DOI:http://dx.doi.org/10.1001/2013.jamainternmed.480

[13] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. 2014. PopTherapy: Coping with stress through pop-culture. In *Proceedings - PERVASIVEHEALTH 2014: 8th International Conference on Pervasive Computing Technologies for Healthcare*. 109–117. DOI:http://dx.doi.org/10.4108/icst.pervasivehealth.2014.255070

[14] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. 5370–5381. https://www.aclweb.org/anthology/P19-1534

[15] M. Seligman, M. E. P. Csikszentmihalyi. 2000. Positive psychology. *American Psychologist* 1 (2000), 5–14.