"Moon Phrases": A Social Media Faciliated Tool for Emotional Reflection and Wellness

Munmun De Choudhury Michael Gamon Aaron Hoff Asta Roseway
Microsoft Research, Redmond WA 98052
{munmund, mgamon, aaronho, astar}@microsoft.com

***Abstract*—We propose an early prototype: a web-based tool "Moon Phrases" that leverages an individual's social activity online for promoting emotional reflection and wellness. Specifically, Moon Phrases tracks the emotion and linguistic expression of individuals as manifested on the social media Twitter, and presents a novel visualization—an analogy to the phases of the moon, to reveal their longitudinal trends. Social media platforms, including Twitter and Facebook provide a window onto the thoughts and feelings of individuals around small and big happenings in their lives. Motivated from research in psychology and HCI, we hypothesize that identifying the changes in language, emotion, and social activity on social media would enable individuals to reflect on their own behavior over time and in a fine-grained manner, which are otherwise known to be difficult to keep track of. We believe Moon Phrases thus bears the potential to act as a self-narrative or "behavioral fingerprint", and thereby serve as an unobtrusive mechanism to facilitate emotional wellness in individuals.**

I. INTRODUCTION

Emotional wellness impacts several aspects of our lives. For instance, it introduces self-empathy, giving an individual greater awareness and acceptance of their feelings. It also improves one's self-esteem and resilience, allowing them to bounce back with ease, from poor emotional health, and physical stress and difficulty. In a professional and work context, emotional wellness has been known to help individuals find better alignment with their work and sense of purpose, and work, family, and social life balance.

However emotional wellness is one the most important and overlooked parts of overall health (World Health Organization, 2010). Unhealthy mental processes are often accompanied by negative self-image; pessimism, skepticism, cynicism regarding self, others, life, and the world, sometimes leading to depression; residual anger; chronic fear/anxiety; and even emotional volatility. In fact mental illness, at times triggered as a consequence of such unhealthy mental processes, is a leading cause of disability worldwide [5]. Global provisions and services for identifying, supporting, and treating mental illness have been considered by WHO as insufficient. Additionally, there is no reliable laboratory test for diagnosing most forms of mental illness; typically, the diagnosis is based on the patient's self-reported experiences, behaviors reported by relatives or friends, and a mental status examination. However in such behavioral surveys responses are prompted by the experimenter and typically comprise recollection of (sometimes subjective) health facts. As a result they are vulnerable to memory bias or experimenter demand effects. Furthermore, because of the typically long temporal gaps (months, sometimes years) in which they are taken, an individual might lose the context that may be associated with behavioral disorders. Given these circumstances, implementing effective emotional wellness programs, interventions, and schemes become difficult.

In the context of these challenges, in this paper, we propose a tool called "Moon Phrases". This early prototype allows tracking of a user's *social activity online* and thereafter enables behavioral reflection—with the potential to act as a "fuzzy" diagnostic and with the goal of enhancing emotional wellness in their lives. In order to do so, Moon Phrases mines the historical postings of a user over time on social media (Twitter in this case), to quantify his/her behavior through affective expression and language use manifested in the posts.

People are increasingly using social media platforms, such as Twitter and Facebook, to share their thoughts and opinions with their contacts. Postings on these sites are made in a naturalistic setting and in the course of daily activities and happenings. As such, social media provides a means for capturing behavioral attributes, in fine temporal granularity, that are relevant to an individual's thinking, mood, communication, activities, and socialization. The affect and language used in social media postings may indicate feelings of worthlessness, guilt, helplessness, and self-hatred that characterize unhealthy mental processes [16]. Additionally, when individuals show withdrawal from social situations and activities, such changes might be salient with changes in volume of posting on social media [13].

In the design of Moon Phrases, we therefore pursue the hypothesis that language, affect, and social activity may be leveraged together to enable individuals reflect on their own behavior over time and in a fine-grained manner, e.g., as a self-narrative or "behavioral fingerprint", and thereby serve as an unobtrusive mechanism to facilitate emotional wellness.

II. BACKGROUND LITERATURE

A. Social, Psycholgical Environment and Emotional Wellness

Social networks and attributes relating to the environment of individuals have consistently been used to study behavioral health concerns and wellbeing. Kawachi et al. [5] explored the role of social ties and social capital in the maintenance of psychological wellbeing and treatment of behavioral health concerns. This prior research provides evidence that individuals' social environments contain information useful for understanding and intervening on emotional wellness. Further, psychological attributes also impact emotional

978-1-936968-80-0 © 2013 ICST

DOI 10.4108/icst.pervasivehealth.2013.252106

wellbeing. Rude et al., [16] found support for the claim that negative processing biases, particularly (cognitive) biases in resolving ambiguous verbal information can be reflective of unhealthy mental processes. Researchers in psycholinguistics have explored how expressions of language can be used to better understand human intentions, moods, and disorders. Computerized analysis of language has revealed cues about emotional closeness, neurotic tendencies, and stress [13,17].

However, research on harnessing social media for reflection on emotional wellness and psychological concerns is still in its infancy [9]. Kotikalapudi et al., [6] analyzed patterns of web activity of college students that could signal emotional concerns. Park et al. [14] found initial evidence that people *do* post about their affective concerns and even their treatment on Twitter. In another work [2], authors examined linguistic and emotional correlates for postnatal course of new mothers as manifested in Twitter. This early work points to the potential of social media as a signal to discover emotional concerns, and thereby help promote wellness.

B. Tools for Emotional Wellness in HCI

There is a rich body of work utilizing signals derived from a variety of sensors and modalities to promote emotional wellbeing in people [12]. Reflection on one's internal emotional state has been found to better enable us understand self, and others' responses to stress [7,20,21]. El Kaliouby et al. [3] discussed an emotional prosthetic using wearable sensors that helped understand and interpret emotion. McDuff et al. [8] extended it, and proposed an emotional prosthetic that allowed users to reflect on their emotional states over long periods of time. Close to our work is the work by Ståhl et al. [19]. They proposed Affective Diary: A tool that provided diary entry creation by incorporating mood, together with data derived from a user's phone, and their hand-written entries.

Interest has been growing in opportunities to employ social media and networking and other Internet data to encourage and promote healthy behavior and well-being [4,18]. Munson et al. [10] presented an application for Facebook that promotes health interventions. The breadth of work aimed at the use of sensors, or social media to promote physical and psychosocial health highlights the spectrum of opportunities for creating applications that can support and encourage health-related awareness and self-reflection.

III. THE DESIGN PROCESS OF MOON PHRASES

Our design process involved user-centered design, wherein we iteratively sought feedback from six fluent social media users. We produced low-fidelity prototypes (i.e., paper and digital mockups) and showed these to six such users to gather their reactions and comments.

A core challenge of the design process was identifying *what* are the best social media cues to be visualized, based on an end user's activity that can promote emotional wellness. We focused on Twitter as the social media tool, because of its data availability within our organization; however our tool could be easily adapted to any other social platform, such as Facebook or email. Our interactions with the fluent social media participants made it clear that patterns and levels of activity of a user define their interactions and overall

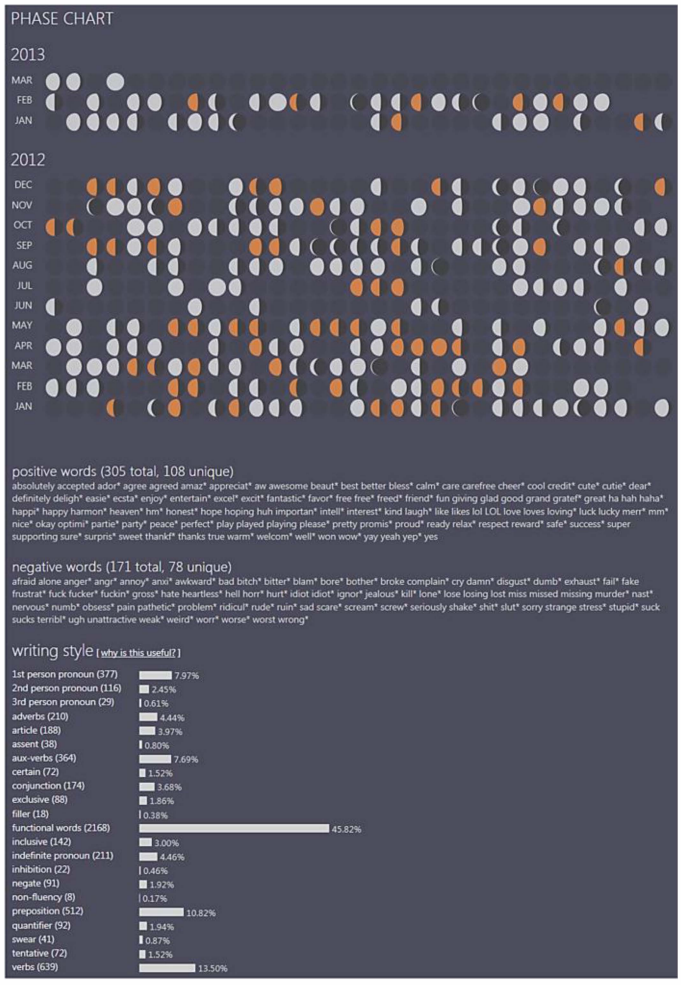

Figure 1. Screenshot of Moon Phrases for a particular Twitter user.

engagement with the social landscape. Therefore observing the volumes of postings on Twitter longitudinally was an important cue for reflection.

We also found evidence in conversations with our set of participants that, linguistic usage of various words was useful as psychological markers, since it conveys information about individuals' social surroundings, contexts and crises they are in. In fact this observation aligned with findings in the sociolinguistics literature [15]. For instance in [15], use of first person singular is known to be associated with negative affective states, because they indicate high self-attentional focus. Besides, pronouns and other function words may provide hints about the truthfulness of statements [17]. Furthermore, exclusion words, like "*but*", "*except*", "*without*", "*exclude*", that typically load with negations (*no, not, never*) are associated with greater cognitive complexity [16]. Hence, a key aspect of Moon Phrases was the ability of users to observe how they use language, and how the usage of various style categories reflected their behavioral characteristics.

Finally, our participants indicated the utility of being able to reflect on their emotional states via their postings on social media. Emotions are founded on interrelated patterns of cognitive processes, physiological arousal, and behavioral reactions. Since Twitter is used to broadcast updates on daily life, as well as on information of interest, our participants felt

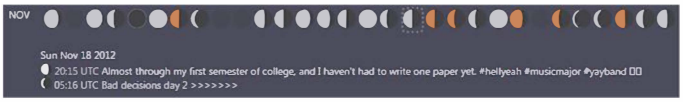

Figure 2. Interactivity of Moon Phrases: clicking on a "moon" a particular day reveals the Twitter posts made on that day.

that reflecting on their emotional expression historically over time might help them manage their feelings effectively, cope with stress and anxiety in a healthy way, and express emotions appropriately in the future.

Weaving these observations together, our design process indicated Moon Phrases to be a mechanism to show trends of people's social activity (i.e., volume of posting on Twitter), and affect over time; as well as usage of linguistic styles that might relate to people's social and psychological environment. We chose one day as the granularity to show these trends.

IV. Interface and Interaction Design

The central idea of Moon Phrases was to show daily trends of positive affect (PA) and negative affect (NA) as manifested in the Twitter postings of a particular user who is provided a mechanism to "log in" into the web tool. Given any valid Twitter username, we utilized Twitter's official API to collect their most recent 1000 posts. We archived these posts, and further, we expanded our historical archive of the same user whenever the web tool was queried for the same user.

A screenshot of the Moon Phrases interface is shown in Fig. 1. In order to effectively show daily affective trends based on these posts, we were motivated by an analogy to lunar phases. Corresponding to each day, we constructed a "moon", where the illuminated portion of the moon represented the degree of average of positivity over all postings in the same day. In other words, fuller moons (the light-filled area of the moon in Fig. 1) indicate greater positive affect. The moons that are white/lighter shade in color in the illuminated portion represented PA corresponding to 1-3 posts, whereas the orange/darker shade ones corresponded to more than three posts—thus an end user gets a sense of the general volume of posting that leads to certain measurement of affect. Hovering over a particular moon further reveals the exact number of posts on that day. Note that the phases of the moons do not have any connection to astronomical phenomena. Each of the "daily" moons is also interactive—on clicking on any moon on a certain day, a user could view the particular date and day of the week, as well as each posting made on that day, including the degree of positivity expressed in each, through a smaller-sized moon (Fig. 2).

As shown in Fig. 1, the postings are organized based on their timestamp of posting—most recent months are represented as rows at the top, and scrolling down one could browse the affect over previous timeframes. Below the phase chart showing affect and activity trends (Fig. 1), for intuitive understanding, we further provide two different lists of total and unique positive and negative words / regular expressions spanning the entire history of Twitter posts. For the purpose, we use a large psycholinguistic dictionary that has been validated to capture human affect, PA and NA (LIWC: http://www.liwc.net). Besides affect, as our design process indicated, we also demonstrate the linguistic style usage (as relative percentages) over all the postings of a user through a bar chart (bottom of Fig. 1). We utilize the popular psycholinguistic lexicon LIWC again, and focus on its 22 different linguistic style marker categories, e.g.: *articles, conjunctions, adverbs, personal pronouns, prepositions*.

V. Technical Feature Development

A central challenge in this work was the inference of the affective nature of Twitter postings. We infer positive and negative affect using machine learning models trained on Twitter collections to detect affect. For the purpose, we leveraged a study based on psychology literature where over 150 different mood hashtags that people use on Twitter were identified [1]. The notion here is that when mood hashtags are used *at the end of Twitter posts*, studies in [1] indicated that these hashtags often acted as a supervisory summary signal indicating a person's affective state in the context of the post. These hashtags were then mapped into positive and negative affect and used as a training signal to identify affect from Twitter posts, based on a text classifier—a maximum entropy classifier trained on unigrams and bigrams of post content. This classifier has been validated on Twitter datasets, with mean accuracy of more than 85% for the two classes of affect, making significant improvements over simple lexicon-driven approaches of matching positive and negative terms [1].

In our context, we first run the classifier on each Twitter post of a user. The output of the classifier is a distribution (probabilities) over PA and NA; thus PA and NA sum up to 1. That is, a post that has PA value of 0.7, would automatically imply that its NA value is 0.3. In order to render the moons per day, we use the PA value of each post on that day, thus derived, and compute the mean positivity over all posts from the user on the same day. Higher values of mean PA on a day were henceforth rendered using more illuminated moons.

Besides affect, we also compute a variety of linguistic styles in the Twitter postings of a user. Given all posts of the user, we follow a regular expression matching technique to determine the total number of words of a certain style occurring in them. We compute the proportion (relative percentages in the bar chart) of each style by dividing this count with the total number of words in all of the posts.

VI. Implications and Opportunities

Moon Phrases presents an early prototype in the design and deployment of effective and smart health intervention systems, derived from people's naturalistic social activity online, for emotional wellness. Automated assessment of behavior in this manner could serve as an early warning mechanism to individuals showing significant or unusual behavioral change. The strength of the tool lies in revealing longitudinal trends of affect and behavior, for instance identifying time periods of low positivity or high negativity—otherwise known to be challenging for individuals to keep track of [8]. Such feedback could be especially valuable to those who are not aware of their risk of unhealthy mental processes, such as depression, trauma, or even risk of suicide. In the case of individuals who might need help, Moon Phrases may also be adapted to log trends and serve as a diary-style data source to aid doctors or other trained professionals gain a deeper understanding of their patients (e.g., in [21]). Emotional markers identified by

Moon Phrases could thus enable adjuvant diagnosis of affective disorders, and serve as a complement to survey based approaches, such as the Center for Epidemiological Study Depression Scale, and help with diagnosis or early intervention by caregivers (e.g., via psychotherapy treatments) aimed at promoting improved health and wellness.

We also intend to incorporate, in future versions of Moon Phrases, a form of "risk score" of individuals to various affective disorders, based on *predictions* that can be made, ahead of time, about forthcoming extreme changes in their behavior and mood [2]. In operation, if inferred likelihoods of forthcoming extreme changes surpass a threshold, they could be warned or engaged, and information might be provided about professional assistance and/or the value of social and emotional support from friends and family.

A. Privacy and Ethical Considerations

Concerns regarding privacy and ethics may arise with tools analyzing social media behavior, as they ultimately leverage health-related information that may be considered sensitive—even if publicly available [11]. We note that Moon Phrases honors the privacy of the user with user-centric design that restricts the sharing of such information to the user herself and optionally to trusted members of friends and family, a trained medical practitioner or a support group. Nevertheless, this type of research, and consequently the nature of the findings it generates, needs to be considered with caution, and we encourage continued discussion of the topic by the research and practitioner communities.

B. Limitations and Future Directions

Although a tool like Moon Phrases enables the analysis of social media postings in fine-grained ways there are not feasible offline, there could be biases in terms of how accurately social media can *truly* reflect human affect and behavior. Additionally, we have little knowledge about people's idiosyncratic behavior "behind the scenes", socio-cultural environment, or aspects such as socio-economic status. Potentially, the limitations of Twitter may be tackled by adding complementary sources of behavioral information, in conjunction with health and wellness records. These opportunities remain ripe areas of future research. It is also important to note that, although the tool is intended for providing a temporally fine-grained narrative of affect and behavior manifested online, it will make the most sense for users who use social media extensively (i.e., post at least once a day), instead of ones who rarely use social media. We also recognize that our Moon-based representation of daily affect needs to be evaluated independently. Other alternatives that could be examined include: smiling/frowning faces; the Sun (sunny/cloudy); or other forms of trend-indicative infoviz.

Finally in this paper, we have not validated the Moon Phrases prototype in terms of its ability to act as an intervention mechanism for emotional wellness yet. Through our design was closely based on feedback from fluent social media users, in the future we intend to evaluate how Moon Phrases can *positively impact* their behavior, beyond revealing mere "signals" of their affect and behavior. We are considering a longitudinal study design in which a set of users may be asked to use the web tool over the course of three to four weeks; with self-reported behavioral and emotional reflection surveys undertaken at the end of every week. Aligning people's social media behavior with these self-reports of affect will hopefully reveal how Moon Phrases could be acting as an intervention mechanism. Certainly, investigating *what* could be the best mechanisms to intervene also constitutes an interesting future research opportunity.

REFERENCES

[1] M. De Choudhury, M. Gamon, and S. Counts. "Happy, Nervous or Surprised? Classification of Human Affective States in Social Media". In *Proc. ICWSM 2012*.

[2] M. De Choudhury, S. Counts, S., and E. Horvitz. "Major Life Changes and Behavioral Markers in Social Media: Case of Childbirth". In *Proc. CSCW 2013*.

[3] R. El Kaliouby, A. Teeters, and R. Picard. "An Exploratory Social-Emotional Prosthetic for Autism Spectrum Disorders". In BSN '06, (2006), IEEE.

[4] N. Kamal, S. Fels, and K. Ho. "Online social networks for personal informatics to promote positive health behavior". In *Proc. WSM '10.*

[5] L. Kawachi, and L. Berkman. "Social ties and mental health". Journal of Urban Health, 78(3), 458-467.

[6] R. Kotikalapudi, S. Chellappan, F. Montgomery, D. Wunsch, and K. Lutzen. "Associating depressive symptoms in college students with internet usage using real Internet data". IEEE Tech & Society Magazine.

[7] K. Liu. "A personal, mobile system for understanding stress and interruptions". MIT, 2004.

[8] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski. "AffectAura: an intelligent system for emotional memory." In *Proc* CHI 2012.

[9] M. Moreno, L. Jelenchick, K. Egan, E. Cox et al. "Feeling bad on Facebook: depression disclosures by college students on a social networking site". Depression and Anxiety 28(6):447–455.

[10] S. Munson, D. Lauterbach, M. Newman, and P. Resnick. "Happier Together: Integrating a Wellness Application Into a Social Network Site". In *Proc. of Persuasive 2010*, Springer. 27-39

[11] M. Newman, D. Lauterbach, S. Munson, P. Resnick, M. Morris. "It's not that I don't have problems, I'm just not putting them on Facebook: Challenges and Opportunities in Using Online Social Networks for Health". In *Proc. CSCW 2011*.

[12] M. Nicolaou, H. Gunes, and M. Pantic. "Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence-Arousal Space". IEEE T Affect Comput (2011).

[13] T. Oxman, S. Rosenberg, and G. Tucker. "The language of paranoia". *American J. Psychiatry* 139:275–82.

[14] M. Park, C. Cha, and M. Cha. "Depressive Moods of Users Captured in Twitter". In Proc. ACM SIGKDD Workshop on Healthcare Informatics.

[15] J. Pennebaker, M. Mehl, and K. Niederhoffer. "Pyschological aspects of natural language use: Our words, ourselves". *Annual Review of Psychology* 54: 547-477.

[16] S. Rude, C. Valdez, S. Odom, and A. Ebrahimi. "Negative cognitive biases predict subsequent depression". Cognitive Therapy and Research, 27(4), 415-429.

[17] S. Rude, E. Gortner, and J. Pennebaker. "Language use of depressed and depression-vulnerable college students". Cognition and Emotion, 1121-1133.

[18] S. Shyam Sundar, A. Oeldorf-Hirsch, J. Nussbaum, and R. Behr. "Retirees on Facebook: can online social networking enhance their health and wellness?" In *Proc. CHI EA 2011*. 2287-2292.

[19] A. Ståhl, K. Höök, M. Svensson, A. Taylora, and M. Combetto. "Experiencing the affective diary". Pers Ubiquit Comput (2009).

[20] J. Suls, P. Green, and S. Hillis. "Emotional reactivity to everyday problems, affective inertia, and neuroticism". Pers Soc Psychol B.

[21] A. Gaggioli, G. Pioggia, G. Tartarisco, G. Baldus et al. "A mobile data collection platform for mental health research." Personal and Uniquitous Computing, 17(2), 241-251.