BMC Research Notes

RESEARCH NOTE

Open Access

Timing rather than user traits mediates mood sampling on smartphones

Beryl Noë[1*], Liam D. Turner[1], David E. J. Linden[2,3], Stuart M. Allen[1], Gregory R. Maio[4] and Roger M. Whitaker[1]

Abstract

Objective: Recent years have seen an increasing number of studies using smartphones to sample participants' mood states. Moods are usually collected by asking participants for their current mood or for a recollection of their mood states over a specific period of time. The current study investigates the reasons to favour collecting mood through current or daily mood surveys and outlines design recommendations for mood sampling using smartphones based on these findings. These recommendations are also relevant to more general smartphone sampling procedures.

Results: N=64 participants completed a series of surveys at the beginning and end of the study providing information such as gender, personality, or smartphone addiction score. Through a smartphone application, they reported their current mood 3 times and daily mood once per day for 8 weeks. We found that none of the examined intrinsic individual qualities had an effect on matches of current and daily mood reports. However timing played a significant role: the last followed by the first reported current mood of the day were more likely to match the daily mood. Current mood surveys should be preferred for a higher sampling accuracy, while daily mood surveys are more suitable if compliance is more important.

Keywords: Mood, Mood sampling, Study design, Smartphone, Smartphone study, Experience sampling methodology

Introduction

There are numerous approaches to assessing mood (e.g. using the PANAS [1], POMS [2], BMIS surveys [3], or the experience sampling method [4]), but only relatively recently have mood surveys migrated to the smartphone [5–7]. Sampling the mood of participants in this way requires a design choice to be made: either sampling current moods several times per day or collecting only once. This choice has different implications for the participant, representing a trade-off between interruption [8] and recall [9]. A single "daily" mood report requires the participant to be accurate with their reflection of the whole day, whereas "current" mood reporting samples a participant's mood at a particular time, but requires more frequent interruption of the user. As such, individual differences between participants and reporting

circumstances could influence responses. Delespaul [10] has already highlighted the importance of not exceeding six data collection points per day for experience sampling procedures. Given the burden the response requests place on the participant, especially when they are not interruptible, it is important to establish whether daily and current mood measures are interchangeable, resulting in recommendations for different data collection frequencies.

Individual differences may result in alternative response dispositions towards surveys [11–15]. Work in this area has found associations with big five personality traits but also need for cognition [16]. While these findings were collected from online surveys, participants might be differently inclined to smartphone-based surveys. Indeed, smartphone interruption to gain user attention and response is an emerging and already complex field in its own right [8].

Specifically in relation to mood surveys, individual differences in personality [17–19], impulsivity [20], and proneness to smartphone addiction [21, 22] could

*Correspondence: NoeB@cardiff.ac.uk
[1] School of Computer Science and Informatics, Cardiff University, The Parade 5, Cardiff CF24 3AA, UK
Full list of author information is available at the end of the article

Noë *et al. BMC Res Notes* (2017) 10:481

Page 2 of 5

contribute to a mismatch in the responses to current and daily reports. Also the intensity of current reported mood states and the amount of time between current and daily mood reports might have moderating effects, which can be predicted based on memory biases such as the recency effect [23], the primacy effect [23], and the peak-end rule [24].

Main text
Methods
Participants
Seventy-six participants were recruited through posters and online advertisement at Cardiff University UK, aged between 19 and 46 (M = 24.94, SD = 5.69). Thirty-nine participants were male, 36 female and 1 participant chose not to disclose their gender. Participants were selected on two aspects: they needed to have a smartphone running Android 4.4 or higher, and they had to have no history of mental illness.

The Android platform was both chosen for convenience (similar data collection on iOS is impeded by the operating system) and to reach a larger number of participants (at the start of the study, in May 2016, 46% of the British population uses Android and 43.39% iOS) [25, 26]. Participants were also selected on absence of mental illnesses. This was done so mental illnesses, especially those have affective symptoms would not become a confounding factor to this study.

Study design
All participants attended a briefing session where they downloaded our custom made application "Tymer", were given instructions on how to use the app and the distinctions between the different reporting options, and were asked to complete five surveys: SAS [27], PANAS [1], BFI [28], MCQ [29], and a demographics and smartphone use questionnaire. After 8 weeks of using Tymer, participants returned for a debriefing session where they retook the surveys and received monetary compensation.

The Tymer application prompted participants to report their current mood (CM) using a dartboard-shaped interface (as shown in Additional file 1: Figure S1 (left)), based on the circumplex model of affect [30], up to three times per day. Notifications requesting the user to complete CM reports only arrived while the smartphone was in use to maximise the likelihood of response. Additionally, participants were asked to select their daily mood (DM) (see Additional file 1: Figure S1 (right)), as part of an evening survey that was sent on the first screen unlock after 19:00 every day. Both type of reports could also be completed through the application interface. Notification expiration time was set to 10 min for CM prompts and at 23:55 for the DM survey. A typical day using the Tymer application is depicted in Additional file 1: Figure S2.

Data cleaning
While 76 participants were recruited, smartphone data was only obtained from 64 participants due to hardware problems and withdrawals from participation. The number of completed and uncompleted reports are shown for both types of surveys in Table 1. In 8 weeks, participants should have completed 56 DM surveys and 168 CM surveys. The mean participation rate considering these numbers was 79.8% for DM and 80.6% for CM surveys.

Pairs of CM and DM surveys undertaken on the same day were analysed. In case several DM surveys were completed for 1 day, only the first one was considered. This resulted in 7893 unique CM and 2667 unique DM surveys being analysed, resulting in 7893 pairs of current and daily mood surveys. In total, there were 1835 instances where a day had at least one CM-DM match, this represents 68.80% of the 2667 reported DM (also see Additional file 2). The BFI was mistakenly done twice by one participant at briefing; only the first submission was used.

Comparison of proportion of matches/non-matches to random
A binomial test was used to compare the proportion of matches and non-matches between CM and DM responses against the number of such matches that would occur in a random sample (1/9 = 11%). The proportion of matches was statistically greater than 11% ($p < .001$) with 2529 (32.04%) of the CM and DM survey pairs reporting the same mood.

Table 1 Frequencies of completion and source of CM and DM surveys

	CM		DM	
	Count	%	Count	%
All	14208	100	4416	100
Source				
Notification	13399	94.31	3851	93.56
App interface	809	5.69	265	6.44
Completion				
Expired	4468	31.45	154	3.74
Abandoned	64	0.45	178	4.32
Completed	8676	61.06	3266	79.35
Dismissed	1000	7.04	518	12.59
Blank	1	0.01	0	0

Noë *et al. BMC Res Notes* (2017) 10:481

Page 3 of 5

Data transformation

Since participation was voluntary, each participant had a varying number of data points. To summarise the data per participant, it was modified by either adopting the count (number of matching or non-matching CM and DM survey pairs), the median (time difference between CM and DM surveys, intensity of the current mood reports) or calculating a percentage (number of matching CM-DM pairs per day) of all instances concerning a participant. Spearman's correlation, Wilcoxon-Mann-Whitney test, and Wilcoxon signed ranks test can be applied due to these transformations.

Results

Effect of time on CM-DM report matches

The median time between evening and current surveys was significantly shorter for matches than non-matches ($Z = -3.103$, $p = .002$). For each participant, days in which matches in mood response occurred were categorised as follows: ALL, where all reported CM(s) of the day matched the reported DM, FIRST, LAST and MIDDLE where the reported CM(s) of the day that matched the reported DM were respectively the first, last, or neither first nor last (see Additional file 3). Since days that had both matches for the first and last reported CM would fall into both of these categories, they were split evenly between them (see Fig. 1). The resultant categorisation was therefore mutually exclusive. It should also be noted, that, since a day was defined as starting at 00:00 and ending at 23:59, some matches could have occurred after the evening survey was completed.

Wilcoxon's signed rank test was used to compare the count of all categories to one another. Matches in the LAST category were found to be significantly more frequent ($p < .01$) than in all other categories ($M = 8.24$, $SD = 5.44$); followed by matches in the FIRST category
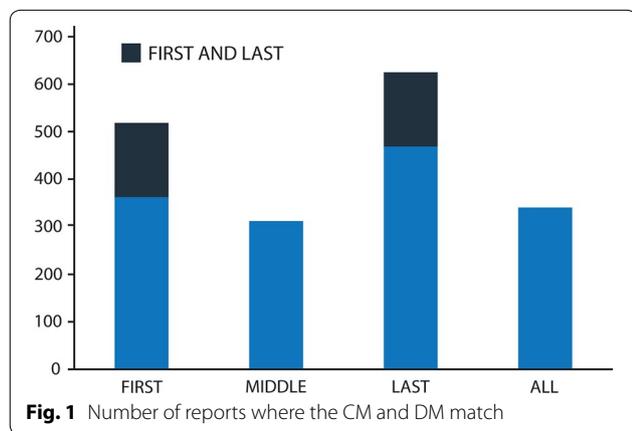

Fig. 1 Number of reports where the CM and DM match

Table 2 Z values and effect sizes for each category pair

Z values

	FIRST	ALL	MIDDLE	LAST
FIRST	–	−2.664**	−2.875**	−2.730**
ALL		–	−.140	−4.103***
MIDDLE			–	−4.748***
LAST				–

Effect sizes

	FIRST	ALL	MIDDLE	LAST
FIRST	–	.333	.359	.341
ALL		–		.513
MIDDLE			–	.594
LAST				–

*** p <.001, ** p <.01, * p <.05

($M = 6.57$, $SD = 4.49$), which were statistically greater ($p < .01$) than the ALL ($M = 5.31$, $SD = 4.40$) and MIDDLE ($M = 1.89$, $SD = 3.12$) categories. These results and their medium to high effect sizes are shown in Table 2 [31].

Additional results can be found in Additional file 4.

Discussion

This study has shown that there is evidence to suggest that CM and DM reports are interchangeable as a methodology to sample participant mood. Indeed 68.8% of the recorded DM matched a CM that was reported in the same day. None of the investigated intrinsic characteristics (gender, age, personality, etc.) had an effect on matches of current and daily moods, suggesting that a specific participant sample would not justify the choice of one over the other reporting method.

Further results show that time intervals between CM and the DM survey had a significant effect on CM-DM matches. This could imply that daily mood does not reflect as much the entirety of the day as intended. As predicted by memory biases, the last reported CM reports were more likely to match the DM report due to them being closer in time, while the first reported CM report came in second in terms of similarity. These findings are consistent with reports of the serial position effect [23], which shows a higher probability of recall for initial and final elements from a list, with lower probability for elements in-between, and with the final element having the highest probability overall. This implies that CM reports might be more accurate to sample current mood than DM reports are for collecting daily mood since memory biases come into play that slightly hinder the formation of an accurate daily summary. These

Noë *et al. BMC Res Notes* (2017) 10:481

Page 4 of 5

findings were supported by medium and high effect sizes (r > .333), which show that the sample size was sufficient to find these effects.

Daily mood surveys were also at a disadvantage considering the number of dismissed notifications (12.59% vs 7.04% for CM surveys), while its percentage of survey completions via the app interface is similar to that of CM surveys (6.44 and 5.69% respectively). However, participants might have not needed to dismiss notifications for CM as they expired more quickly. CM surveys were more invasive as participants were prompted at least three times per day, while DM surveys only happened once at set time. This is likely to have contributed to an overall higher completion rate for DM (79.35%) than CM (61.06%) reports.

Our average completion rates (about 80% for both types of surveys) were quite high considering the length of our study and are mostly higher than those reported in similar studies [32]. We believe the best way to increase compliance and accuracy of participants, would be to increase the incentives for good performance through feedback (e.g. higher rewards by providing visualisation of historic personal data or gamifying parts of the app). While feedback has been shown to increase compliance [33], the increased awareness could however influence the participant. While Downes-LeGuin and colleagues [34] have shown gamification to be ineffective to increase engagement even though it increased satisfaction, other studies do report heightened engagement [35].

Additional discussion points can be found in Additional file 5.

Conclusions

Whether current or daily mood surveys should be used to collect affective data on participants highly depends on the requirements of the study, and whether related *in-situation* context or device usage is important. One also needs to consider what exactly needs to collected: momentary mood fluctuations, or prevailing mood of the day. However our results indicate that both approaches can be used with confidence, albeit noting specific implications for each.

If participant compliance is of high importance, daily surveys should be favoured as participants are more likely to dismiss notifications if they are frequent or come at inopportune moments.

We note that while the investigated intrinsic characteristics did not affect the two surveys differently, effects for time did come into play. Current mood surveys are more accurate as the participant is directly asked for the mood state they are in at that instant, while a daily mood survey requires the participant to provide a summary of the

moods they have felt during the day and this cognitive task is vulnerable to memory biases.

Limitations

This study had a few limitations:

- Only Android users were selected. This has consequences on the generalisability of our results since previous literature has shown that Android and iPhone user groups may be quite distinct [36].
- CM and DM were collected simultaneously and could have influenced each other.
- Since the mood measures were all self-reported, given responses could be dishonest or not well-estimated. Misclicks can also occur.
- Smartphone data was missing from 12 participants.

Additional files

Additional file 1. Study design. Additional information on demographics and smartphone use questionnaire and two additional figures.

Additional file 2. Additional data. One additional table and three additional figures.

Additional file 3. Classification example. Additional figure.

Additional file 4. Additional results.

Additional file 5. Discussion of further applications of the Tymer app.

Author details
¹ School of Computer Science and Informatics, Cardiff University, The Parade 5, Cardiff CF24 3AA, UK. ² School of Psychology, Cardiff University Brain Research Imaging Centre, Maindy Road, Cardiff CF24 4HQ, UK. ³ School of Medicine, MRC Centre for Neuropsychiatric Genetics and Genomics, Maindy Road, Cardiff CF24 4HQ, UK. ⁴ Department of Psychology, University of Bath, 10 West, Bath BA2 7AY, UK.

Noë *et al. BMC Res Notes* (2017) 10:481

Page 5 of 5

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Watson D, Clark LA. The PANAS-X manual for the positive and negative affect schedule–expanded form. Order J Theor Ordered Sets Appl. 1994;277(6):1–27. doi:10.1111/j.1742-4658.2010.07754.x. arXiv: NIHMS150003
2. Lin S, Hsiao Y-Y, Wang M. Test review: the profile of mood states 2nd edition. J Psychoeduc Assess. 2014;32(3):273–7. doi:10.1177/0734282913505995.
3. Mayer JD, Gaschke YN. The experience and meta-experience of mood. J Pers Soc Psychol. 1988;55(1):102–11.
4. Conner Christensen T, Feldman Barrett L, Bliss-moreau E, Lebo K, Kaschub C. A practical guide to experience-sampling procedures. Stone Shiffman J Happiness Stud. 2003;4:53–78.
5. LiKamWa R, Liu Y, Lane N, Zhong L. Can your smartphone infer your mood. In: PhoneSense workshop. 2011. p. 1–5.
6. LiKamWa R, Liu Y, Lane ND, Zhong L. Moodscope. In: Proceeding of the 11th annual international conference on Mobile systems, applications, and services–MobiSys' 13. 2013. p. 389. doi:10.1145/2462456.2464449.
7. Khue LM, Jarzabek S. Demonstration paper: mood self–assessment on smartphones. In: Proceedings of the conference on wireless health. ACM, New York, NY, USA. 2015. doi:10.1145/2811780.2811921.
8. Turner LD, Allen SM, Whitaker RM. Interruptibility prediction for ubiquitous systems : conventions and new directions from a growing field. In: Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing. 2015. p. 801–812.
9. Gorin AA, Stone AA. Recall biases and cognitive errors in retrospective self-reports: a call for momentary assessments. In: Baum A, Revenson TA, Singer JE, editors. Handbook of health psychology. Mahwah: Lawrence Erlbaum Associates; 2001. p. 405–13.
10. Delespaul PAEG. Technical note: devices and time-sampling procedures. In: de Vries MW, editor. The experience of psychopathology: investigating mental disorders in their natural settings. New York: Cambridge Univesrity Press; 1992. p. 363–73.
11. Fan W, Yan Z. Factors affecting response rates of the web survey : a systematic review. Comput Hum Behav. 2010;26(2):132–9. doi:10.1016/j.chb.2009.10.015.
12. Marcus B, Schütz A. Who are the people reluctant to participate in research? Personality correlates of four different types of nonresponse as inferred from self- and observer ratings. J Pers. 2005;73(4):959–84. doi:10.1111/j.1467-6494.2005.00335.x.
13. Rogelberg SG, Conway JM, Sederburg ME, Spitzmüller C, Aziz S, Knight WE. Profiling active and passive nonrespondents to an organizational survey. J Appl Psychol. 2003;88(6):1104–14. doi:10.1037/0021-9010.88.6.1104.
14. Tuten TL, Bosnjak M. Understanding differences in web usage: the role of need for cognition and the five factor model of personality. Soc Behav Personal. 2001;29(4):391–8. doi:10.2224/sbp.2001.29.4.391.
15. Galesic M, Bosnjak M. Personality traits and participation in an online access panel. In: Presentation at the German online research conference, Bielefeld, Germany. 2006.
16. Cacioppo JT, Petty RE, Kao CF. The efficient assessment of need for cognition. J Personal Assess. 1984;48(3):306–7.
17. Chittaranjan G, Jan B, Gatica-Perez D. Who's who with big-five: analyzing and classifying personality traits with smartphones. In: Proceedings-international symposium on wearable computers, ISWC. 2011. p. 29–36. doi:10.1109/ISWC.2011.29.
18. Phillips JG, Butt S, Blaszczynski A. Personality and self-reported use of mobile phones for games. Cyberpsychol Behav. 2006;9(6):753–8. doi:10.1089/cpb.2006.9.753.
19. Ehrenberg A, Juckes S, White KM, Walsh SP. Personality and self-esteem as predictors of young people's technology use. Cyberpsychol Behav. 2008;11(6):739–41. doi:10.1089/cpb.2008.0030.
20. Whiteside SP, Lynam DR. The five factor model and impulsivity: using a structural model of personality to understand impulsivity. Personal Individ Differ. 2001;30:669–89.
21. Black DW, Belsare G, Schlosser S. Clinical features, psychiatric comorbidity, and health-related quality of life in persons reporting compulsive computer use behavior. J Clin Psychiatry. 1999;60(12):839–44.
22. Chou C, Hsiao M-C. Internet addiction, usage, gratification, and pleasure experience: the Taiwan college students' case. Comput Educ. 2000;35(1):65–80. doi:10.1016/S0360-1315(00)00019-1.
23. Murdoch BB. The serial position effect of free recall. J Exp Psychol. 1962;64(5):482–8.
24. Fredrickson BL, Kahneman D. Duration neglect in retrospective evaluations of affective episodes. J Personal Soc Psychol. 1993;65(1):45–55.
25. Statista: market share of Android in the United Kingdom (UK) from July 2011 to May 2017. 2017. https://www.statista.com/statistics/271240/android-market-share-in-the-united-kingdom-uk/. Accessed 18 July 2017.
26. Statista: market share of Apple iOS in the United Kingdom (UK) from December 2011 to May 2017. 2017. https://www.statista.com/statistics/271195/apple-ios-market-share-in-the-united-kingdom-uk/.
27. Kwon M, Lee JY, Won WY, Park JW, Min JA, Hahn C, Gu X, Choi JH, Kim DJ. Development and validation of a smartphone addiction scale (SAS). PLoS ONE. 2013;8(2):e56936. doi:10.1371/journal.pone.0056936.
28. John OP, Srivastava S. Big five inventory (Bfi). Handb Personal Theor Res. 1999;2:102–38. doi:10.1525/fq.1998.51.4.04a00260.
29. Kirby KN, Petry NM. Heroin and cocaine abusers have higher discount rates for delayed rewards than alcoholics or non-drug-using controls. Addiction. 2004;99(4):461–71. doi:10.1111/j.1360-0443.2003.00669.x.
30. Russell JA. A circumplex model of affect. J Personal Soc Psychol. 1980;39(6):1161–78.
31. Rosenthal R. Parametric measures of effect size. In: Cooper HM, editor. The handbook of research synthesis. New York: Russell Sage Foundation; 1994. p. 231–44.
32. Dubad M, Winsper C, Meyer C, Livanou M, Marwaha S. A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. Psychol Med. 2017. p. 1–21. doi:10.1017/S0033291717001659.
33. Hsieh G, Li I, Dey A, Forlizzi J, Hudson SE. Using visualizations to increase compliance in experience sampling. In: UbiComp'08. 2008. p. 164–167.
34. Downes-Le Guin T, Baker R, Mechling J, Ruyle E. Myths and realities of respondent engagement in online surveys. Int J Market Res. 2012;54(5):613–33.
35. Li W, Grossman T, Fitzmaurice G. GamiCAD: a gamified tutorial system for first time AutoCAD users. In: UIST'12. 2012. p. 103–112.
36. Shaw H, Ellis DA, Kendrick L-R, Ziegler F, Wiseman R. Predicting smartphone operating system from personality and individual differences. Cyberpsychol Behav Soc Netw. 2016;19(12):727–32.