Chatbots for Experience Sampling – Initial Opportunities and Challenges

Florian Bemmann[a], Ramona Schoedel[a], Niels van Berkel[b] and Daniel Buschek[c]

[a]LMU Munich
[b]Department of Computer Science, Aalborg University
[c]Research Group HCI + AI, Department of Computer Science, University of Bayreuth

Abstract

The Experience Sampling Method is a widely used methodology for the collection of self-report data. These self-reports are typically collected through bespoke mobile applications or text messages. Recently, an increasing number of social messaging applications have introduced chatbots – automated services that operate inside existing chat applications. In this paper, we present an initial study on the use of chatbots for self-report studies. Furthermore, we outline three use cases in which the use of chatbots enables new research opportunities, namely personalised and empathic chatbots, psychometric instrument construction, and group-based assessments. We conclude with an overview of the opportunities and challenges that chatbots offer to researchers employing the Experience Sampling Method.

Keywords

Experience sampling, ecological momentary assessment, conversational user interface, chatbot

1. Introduction

The Experience Sampling Method (ESM) is a widely used data collection technique in which participants repeatedly answer a short number of questions throughout their daily lives [1, 2, 3]. Originally designed to be used in combination with analogue pen-and-paper and a beeper [2], the ESM has continued to evolve along with the development and uptake of novel technologies. Currently, the method is primarily relying on data collection through smartphones due to their widespread adoption and the opportunities provided by the collection of smartphone sensor data.

Recently, the use of chatbots, defined as "*chatterbot-like agents with simple, textual interfaces that allow users to access information, make use of services, or provide entertainment through online messaging platforms*" [4], has been identified as a potential future avenue for data collection in ESM studies [3]. Chatbots often operate within existing communication applications, reducing the barrier to entry for study participants when signing up for a study and integrating more naturally with typical smartphone usage behaviour. Early results with the use of chatbots as tool for ESM research appear promising. Chatbots instead of surveys have shown higher response quality and quantity [15], and participants are less likely to satisfice (*i.e.*, give lower quality responses motivated by reducing the cognitive effort of answering) [5].

Moreover, bots as 'experimenters' may pro-

Joint Proceedings of the ACM IUI 2021 Workshops, April 13-17, 2021, College Station, USA
florian.bemmann@ifi.lmu.de (F. Bemmann); ramona.schoedel@psy.lmu.de (R. Schoedel); nielsvanberkel@cs.aau.dk (N. van Berkel); daniel.buschek@uni-bayreuth.de (D. Buschek)

Type	Description	Further reading
Technical	Bot reuses existing frameworks, no need for dedicated UI.	[4]
Organisational, technical	Bot uses existing platforms for distribution (*e.g.*, bot store).	[3]
Organisational, technical	Bot uses existing apps that users already interact with (*e.g.*, chat app).	[4]
Technical	Bot runs on many platforms (*e.g.*, not OS specific).	[3]
Data quality/quantity, procedure	Conversation more interactive and engaging than questionnaire.	[5, 6, 7]
Data quality, procedure	Bot reacts to answers dynamically (cf. interviewing).	[3]
Data quality, procedure	Bot dynamically refers back to previously collected data.	[3, 8]
Data quality, procedure	Bot integrates qual. methods (*e.g.*, discussion) in quant. methods (*e.g.*, crowd task).	[9]
Data quality, procedure	Bot standardises procedure (*e.g.*, phrasing of questions), which may reduce experimenter effects.	[9]
Data quality, study experience	Bot may reduce social acceptability bias and increase willingness to disclose.	[9, 10, 11]
Data quantity, procedure	Bot serves as part of a group chat or as connection to peers of participants.	[12]
Data quality, study experience	Bot adapts to individual participant / answer / situation.	[13, 14]

Table 1
Overview of identified opportunities of using chatbots for ESM studies. Note that these are *potential* benefits. Some require further investigation and empirical evidence. The provided references propose or discuss these and related ideas.

vide further benefits, including [9, 11]: 1) enabling aspects of qualitative data collection in quantitative methods (*e.g.*, discussions among online participants), 2) limiting social acceptability biases and allowing people to share highly personal experiences (c.f. [10]), and 3) standardizing interview procedures and reducing experimenter effects. Benefits for response quality and participant engagement are also echoed in other related studies, covering contexts such as crowdtasks [6] and course evaluations [7].

Chatbots enable new approaches to interactive data collection. For example, intelligent chatbots powered by Natural Language Processing (NLP) can enable a more conversational interaction with the participant and adjust follow-up questions, or the conversational style of the chatbot can be adjusted to accomodate the research goal.

As messaging is a common activity in smartphone usage [16], the use of chatbots may overcome some of the challenges faced in typical ESM studies, such as a decrease in response rate over the duration of the study [3]. Yet, many questions facing the use of chat-

bots as an interface between researcher and participant remain unanswered. Similarly, the functionalities that chatbots may provide to researchers are not yet thoroughly explored. In this paper, we present an overview of the opportunities enabled by chatbots for use in ESM studies, building on initial work utilising chatbots for ESM [12, 8, 5], previously identified challenges in ESM-based research [17, 3], and functionalities offered by today's chatbots [18, 4]. We provide an overview of existing studies on chatbots for ESM and their identified opportunities in Table 1.

2. Qualitative User Study

We conducted a qualitative user study comparing app-based experience sampling (Figure 3) with a chatbot equivalent (Figure 1). Our aim was to get a first impression on the suitability of chatbots for experience sampling in general. The chatbot focused on the basic requirements: both bot and app prompted 2 to 4 questionnaires per day at a random schedule, asking ten exemplary single-choice items from psychological research (self-made items for valence, arousal and stress (see Figure 1), and situation [19]).

After introducing the bot with an opening message, the bot presented the items' answer options as buttons in the chat. Alternatively, free text responses were possible so that participants could convey what they think in a conversational manner. The bot's language was in a casual tone, deemed most appropriate by Kim et al. [5]. However, as the text of the question items itself is not to be changed, this only was applied in introductory and end messages. The naming, 'MyESBot' makes clear that one is chatting with a bot and not a real human [10]. We deliberately omitted any further features like an intelligent conversation flow, to focus on the concept of experience sampling as a bot and allow a fair comparison

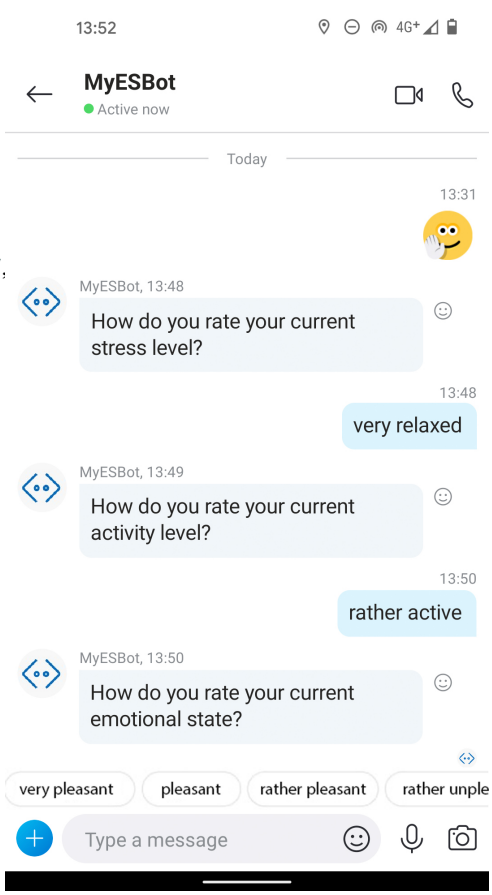

Figure 1: Screenshot of MyESBot, an Experience Sampling chatbot implemented in Skype.

with a traditional app-based ESM approach.

We deployed both approaches in a within-subject study (demographics are shown in Figure 2). The participants used both systems for 5 days each, starting with the app-based approach. For implementation of the app-based experience sampling we used the PhoneStudy app[1], a mobile sensing Android app. The chatbot was used via Skype, implemented using the Microsoft Bot Framework[2]. At the

[1] https://osf.io/ut42y/, *last accessed 6th January 2021.*
[2] https://dev.botframework.com/, *last accessed 6th*

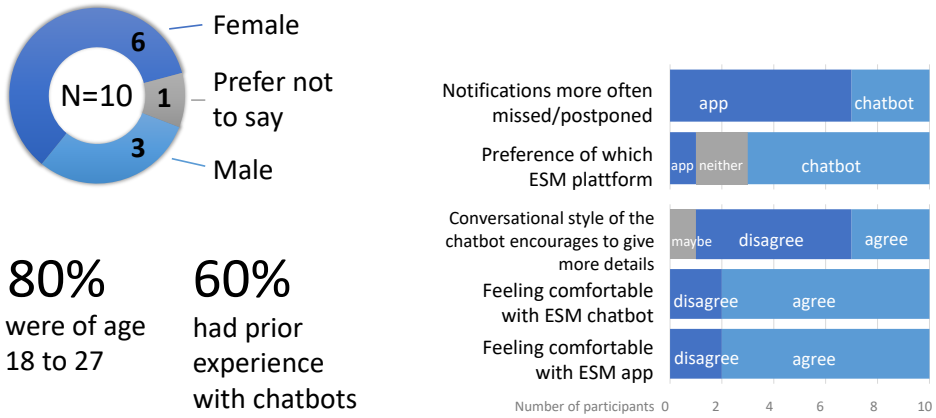

Figure 2: Descriptive statistics on the participants in our study (left) and overview of Likert-item responses (right).

end of the study we asked the participants to answer a survey, where we asked for prior experience with chatbots, usage behaviour of both our app and chatbots, preferences, and basic demographics.

Seven out of ten stated that they dismissed the app notifications more often than the bot's notifications. The perceived response rate was higher using the bot for most participants. An equal number, eight of ten, felt comfortable providing their information to the app and the bot. Seven participants stated that they preferred that the chatbot experience sampling does not require to install an additional app. (Figure 2)

Comments provided further insights into what people liked and disliked about the bot. P8 mentioned missing "*The feeling that you were talking to someone*" in the app in contrast to the bot, and also that "*The PhoneStudy felt more anonymous*". P7 stated "*I felt lonely while using the PhoneStudy app*".

January 2021.

3. Use cases

We present three use cases to discuss future possibilities of chatbots for Experience Sampling research.

3.1. Personalised and Empathic ESM Chatbots

A chatbot could be adapted to an individual participant with the intention of increasing likability, trust, and response quality – or even to better negotiate aspects of self-reflection and awareness which might be introduced by ESM itself. This might be particularly relevant for research questions on more sensitive and intimate personal experiences (*c.f.* [10]), or for research with vulnerable populations. Another potential foundation for individual adaptation is to provide the bot with a certain *personality* that fits the participant [20]. This impression of personality might be achieved, for example, by choice of words [21, 22], use of emojis [13, 23], or intonation (in the case of voice-based bots [22]).

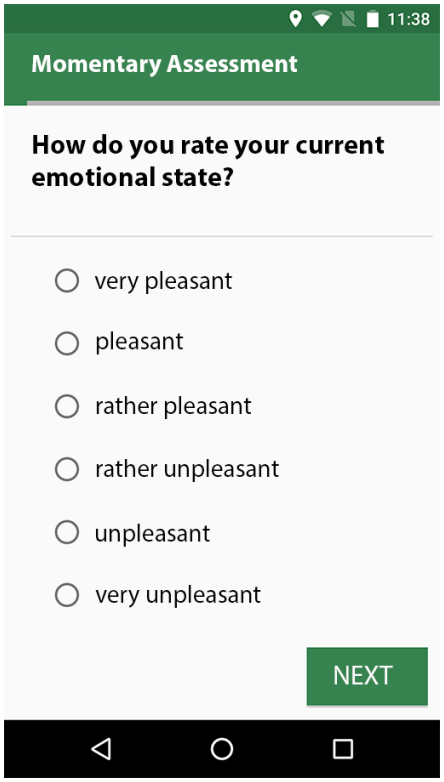

Figure 3: Screenshot of the user study's control condition: Experience Sampling in the PhoneStudy app.

3.2. Application in Psychometric Research

A further use case for the employment of ESM chatbots could be for constructing items suitable for regular experience sampling studies. The use of ESM chatbots could be a complementary approach to traditional ones. Combined with situation detection based on smartphone sensing data, conversations with the chatbot could be initialised in particular situations of interest and the chatbot could take over the part of the interviewer. This procedure would enable the qualitative collection of relevant behaviours and experiences *in situ* [24].

3.3. Group-based Experience Sampling

Chatbots, as integrated in group conversations, could enable researchers to more systematically study the development and changes of group-based interactions. This is especially valuable given the increasing use and frequency of digital communication tools.

The functionality of chat-based applications, *e.g.* responding with an emoji, can be used to automatically infer participants' opinions towards both the utterances by the bot and messages shared by other participants in the chat. Furthermore, chatbots could allow for both private assessment and public discussion. An ESM chatbot which operates simultaneously in a number of private chats as well as in a group-based chat would enable researchers to capture rich-insights into the overlap and differences among individuals in their private considerations versus their public expression.

Finally, maintaining (a long-term) response rate among participants has long been highlighted as a difficult challenge in the ESM literature [1, 2, 3]. Group-based ESM can alleviate this problem through two approaches. First, by being integrated in a group-chat, the 'ESM notification' will alert multiple people to respond to the incoming question. The response of each individual will subsequently inform all other participants (assuming the 'group chat' is not silenced), serving as a repeating reminder to participants to answer the presented question. Second, for some research questions the collection of a single response per notification could be considered as sufficient – by posing the question in a group chat the likelihood of a response being collected is increased.

4. Discussion: Opportunities and Challenges for Research

The presented literature and study indicate several relevant directions for further research on ESM chatbots, some of which we discussed in our Use Case section. In the following paragraphs, we chart a broader research agenda by listing challenges and potential research questions for future work.

Response UI: How might we design response input methods for ESM chatbots to elicit responses of high quality? For example, usual chat text entry, reply buttons, image capture, audio recording, etc. might all be more or less suitable and relevant, depending on question type, ESM context, and overarching research question. Similar to traditional ESM [3], such factors and their influence on response behaviour and quality need to be investigated for ESM chatbots.

Conversation flow: How might we design the conversation flow of ESM chatbots to elicit responses of high quality? Depending on the research interests behind an ESM study, chatbots may require different kinds of conversation flows. This includes design decisions such as "*When to ask what?*", "*When to ask follow-up questions?*", etc. Some of these decisions might be informed by insights from traditional ESM studies while others might arise specifically from the conversational context, in particular if a somewhat realistic conversational impression is considered important.

Personalisation: How might we design an ESM chatbot that gives a personalised and empathic impression? To achieve personalisation as motivated in the Use Case section, future research is needed which investigates when and how to personalise the bot and its responses, based on which information, and so on. Crucially, the impact of such personalisation efforts on responses and participants' study experience needs to be thoroughly studied as well.

Data analysis: How might we support researchers using ESM chatbots to gain insights from their new data collection opportunities? Current ESM methods typically collect easily quantifiable responses, such as ratings on Likert items. In contrast, chatbots may facilitate the collection of open text responses, for example due to the conversation style and the idea of a 'chatbot as an interviewer'. However, these open responses can be challenging to analyse, in particular on the scale of large ESM deployments. To address this, future work could investigate NLP methods for (semi-)automated text analysis to process such ESM chatbot conversations.

Evaluation: How might we evaluate ESM chatbots and their impact to inform future methodological decisions? This is a meta question for the method of ESM chatbots itself. Comparative investigations regarding response quality and quantity, the study experience for participants, and so on are important to inform methodological choices in the future: For example, when should researchers use a chatbot vs a traditional UI for their ESM studies? Which factors are relevant for making this decision?

Privacy and Trust: How might we design ESM studies with chatbots to protect participants' privacy? Researchers likely do not control the entire data collection pipeline, for example, if the bot runs via a third-party environment (*e.g.*, a Facebook bot). Addressing these issues requires a combination of efforts, including technical aspects (*e.g.*, encryption), organisational decisions (*e.g.*, where is final data stored?), participant information (*e.g.*, transparently communicating data flow), and UI (*e.g.*, in-app information, user controls). Furthermore, the use of a third-party channel might negatively impact trust of participants, compared to a dedicated ESM app released by a university.

Attention and control: How does the incorporation of ESM-questions into a chat application influence the user control over interruptions? Users are used to cope with messenger notifications, *e.g.* by muting specific channels or checking messages in batches. If ESM is integrated in a messenger application, the effects might be twofold: On the one hand the users might perceive less interruptions, as the users can apply their usual chat coping strategies. On the other hand response behaviour might be different and bias the collected data, as for example batch checking of messages leads to delayed response of the questionnaire.

5. Conclusion

Prior research has shown that chatbots are a promising tool for data assessment in the wild, *e.g.* for conducting questionnaires or experience sampling studies. However, the use of chatbots for experience sampling is still in its infancy. We presented a pilot study that points out the opportunities of new ways of interaction: a more personal relation towards the bot, fewer dismissed notification, and a probably less interruptive integration into daily life. Our use cases highlight opportunities which require further exploration, including empathic chatbots, bots as interviewers, and use of the ESM in group interaction. Furthermore, we propose novel applications for chatbots as a research tool: their application for item construction in psychometric research can avoid some disadvantages of current methods. In group conversations, chatbots could be used as a passive listener, or to prompt questions to a set of participants instead of just one at a time. We hope that our initial exploration on the use of chatbots for ESM research inspires and provokes the research community to further explore the use of chatbots as a research tool in future studies.

Acknowledgments

This project is partly funded by the Bavarian State Ministry of Science and the Arts and coordinated by the Bavarian Research Institute for Digital Transformation (bidt).

References

[1] M. Csikszentmihalyi, R. Larson, Validity and Reliability of the Experience-Sampling Method, in: M. Csikszentmihalyi (Ed.), Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi, 2014, pp. 35–54.

[2] R. Larson, M. Csikszentmihalyi, The Experience Sampling Method, 2014, pp. 21–34.

[3] N. van Berkel, D. Ferreira, V. Kostakos, The experience sampling method on mobile devices, ACM Computing Surveys 50 (2017) 93:1–93:40. doi:10.1145/3123988.

[4] L. C. Klopfenstein, S. Delpriori, S. Malatini, A. Bogliolo, The rise of bots: A survey of conversational interfaces, patterns, and paradigms, DIS '17, 2017, p. 555–565. URL: https://doi.org/10.1145/3064663.3064672. doi:10.1145/3064663.3064672.

[5] S. Kim, J. Lee, G. Gweon, Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality, CHI '19, 2019, p. 1–12. URL: https://doi.org/10.1145/3290605.3300316. doi:10.1145/3290605.3300316.

[6] S. Qiu, U. Gadiraju, A. Bozzon, Improving worker engagement through conversational microtask crowdsourcing, CHI '20, 2020, p. 1–12. URL: https://doi.org/10.1145/3313831.3376403. doi:10.1145/3313831.3376403.

[7] T. Wambsganss, R. Winkler, M. Söllner, J. M. Leimeister, A conversational agent to improve response quality in course evaluations, CHI EA '20, 2020, p. 1–9. URL: https://doi.org/10.1145/3334480.3382805. doi:10.1145/3334480.3382805.

[8] M. Fetter, T. Gross, Primiexperience: Experience sampling via instant messaging, CSCW '11, 2011, p. 629–632. URL: https://doi.org/10.1145/1958824.1958931. doi:10.1145/1958824.1958931.

[9] N. van Berkel, J. Goncalves, D. Hettiachchi, S. Wijenayake, R. M. Kelly, V. Kostakos, Crowdsourcing perceptions of fair predictors for machine learning: A recidivism case study, Proc. ACM Hum.-Comput. Interact. 3 (2019). URL: https://doi.org/10.1145/3359130. doi:10.1145/3359130.

[10] G. M. Lucas, J. Gratch, A. King, L.-P. Morency, It's only a computer: Virtual humans increase willingness to disclose, Computers in Human Behavior 37 (2014) 94–100. URL: http://www.sciencedirect.com/science/article/pii/S0747563214002647. doi:https://doi.org/10.1016/j.chb.2014.04.043.

[11] S. Wijenayake, N. van Berkel, J. Goncalves, Bots for research: Minimising the experimenter effect, CHI EA '20, 2020.

[12] Y.-L. Chang, Y.-J. Chang, C.-Y. Shen, She is in a bad mood now: Leveraging peers to increase data quantity via a chatbot-based esm, MobileHCI '19, 2019. URL: https://doi.org/10.1145/3338286.3344406. doi:10.1145/3338286.3344406.

[13] A. Ghandeharioun, D. McDuff, M. Czerwinski, K. Rowan, Towards understanding emotional intelligence for behavior change chatbots, ACII '19, 2019, pp. 8–14.

[14] A. Ghandeharioun, D. McDuff, M. Czerwinski, K. Rowan, Emma: An emotion-aware wellbeing chatbot, ACII '19, 2019, pp. 1–7.

[15] Z. Xiao, M. X. Zhou, Q. V. Liao, G. Mark, C. Chi, W. Chen, H. Yang, Tell me about yourself: Using an ai-powered chatbot to conduct conversational surveys, arXiv preprint arXiv:1905.10700 (2019).

[16] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, G. Bauer, Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage, MobileHCI '11, 2011, p. 47–56. URL: https://doi.org/10.1145/2037373.2037383. doi:10.1145/2037373.2037383.

[17] N. van Berkel, J. Goncalves, K. Wac, S. Hosio, A. L. Cox, Human accuracy in mobile data collection, International Journal of Human-Computer Studies 137 (2020) 1–4. doi:10.1016/j.ijhcs.2020.102396.

[18] M. Jain, P. Kumar, R. Kota, S. N. Patel, Evaluating and informing the design of chatbots, DIS '18, 2018, p. 895–906. URL: https://doi.org/10.1145/3196709.3196735. doi:10.1145/3196709.3196735.

[19] J. F. Rauthmann, D. Gallardo-Pujol, E. M. Guillaume, E. Todd, C. S. Nave, R. A. Sherman, M. Ziegler, A. B. Jones, D. C. Funder, The situational eight diamonds: A taxonomy of major dimensions of situation characteristics., Journal of Personality and Social Psychology 107 (2014) 677.

[20] S. T. Völkel, R. Schödel, D. Buschek, C. Stachl, V. Winterhalter, M. Bühner, H. Hussmann, Developing a personality model for speech-based conversational agents using the psycholexical approach, CHI

’20, 2020, p. 1–14. URL: https://doi.org/10.1145/3313831.3376210. doi:10.1145/3313831.3376210.

[21] B. Krenn, B. Endrass, F. Kistler, E. André, Effects of Language Variety on Personality Perception in Embodied Conversational Agents, in: M. Kurosu (Ed.), Human-Computer Interaction. Advanced Interaction Modalities and Techniques, Lecture Notes in Computer Science, 2014, pp. 429–439. doi:10.1007/978-3-319-07230-2_41.

[22] M. Braun, A. Mainz, R. Chadowitz, B. Pfleging, F. Alt, At your service: Designing voice assistant personalities to improve automotive user interfaces, CHI ’19, 2019, p. 1–11. URL: https://doi.org/10.1145/3290605.3300270. doi:10.1145/3290605.3300270.

[23] S. T. Völkel, D. Buschek, J. Pranjic, H. Hussmann, Understanding emoji interpretation through user personality and message context, MobileHCI ’19, 2019. URL: https://doi.org/10.1145/3338286.3340114. doi:10.1145/3338286.3340114.

[24] C. Wrzus, M. R. Mehl, Lab and/or field? measuring personality processes and their social consequences, European Journal of Personality 29 (2015) 250–271.