Persona Analytics: Implementing Mouse-Tracking for an Interactive Persona System

Soon-Gyo Jung
Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
sjung@hbku.edu.qa

Joni Salminen
Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar; Turku School of Economics, University of Turku, Turku, Finland
jsalminen@hbku.edu.qa

Bernard J. Jansen
Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
jjansen@acm.org

ABSTRACT

Observing user interactions with interactive persona systems offers important insights for the design and application of such systems. Using an interactive persona system, user behavior and interaction with personas can be tracked with high precision, addressing the scarcity of behavioral persona user studies. In this research, we introduce and evaluate an implementation of persona analytics based on mouse tracking, which offers researchers new possibilities for conducting persona user studies, especially during times when in-person user studies are challenging to carry out.

CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI).

KEYWORDS

Personas, Analytics, Persona Analytics, Mouse-tracking, User Studies

ACM Reference Format:
Soon-Gyo Jung, Joni Salminen, and Bernard J. Jansen. 2021. Persona Analytics: Implementing Mouse-Tracking for an Interactive Persona System. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 08–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3411763.3451773

1 INTRODUCTION

Personas are fictitious user representations [12], and are frequently used in Human-Computer Interaction (HCI) and User-Centered Design (UCD) processes. Personas capture essential user needs and requirements for products [6], help designers avoid the self-referential bias [26], and make a mental model of end-users available when real users are absent [32]. Personas can also give faces to user analytics data [19], humanize a large number of market segments [9], help compare user types [23], provide design inspiration [30], help justify and prioritize design choices [33]. Research has also shown personas to be an effective communication tool when discussing user needs and pain points in a product team [8]. Personas are typically presented as profiles containing key information about the users (see Figure 1).

To advance the theory and practice of personas, studies on persona behavior (i.e., 'how end users interact with personas') are in great need. This is because understanding persona user behavior helps inform all aspects of the persona lifecycle [1], from their creation and evaluation to their successful adoption and use in organizations. Some of the several fundamental areas of inquiry that are currently unaddressed include: (1) *How do users interact with personas?* (2) *What persona information do users pay attention to?* (3) *What information causes users to change/reinforce their attitudes?* (4) *What information influences users' decision making?* (5) *How do users choose a persona for their task?* Empirical questions such as these are largely unaddressed in HCI literature, hindering the fundamental progress in persona design and development towards questions like "What kind of personas should we design?", "What features should a persona system have?", and so on. To address these fundamental questions, empirical persona user research is required.

While persona use can be measured in many ways, such as observations [15], ethnography [8], and eye-tracking [18, 44], mouse-tracking [40] has shown to be a useful technique in understanding how users interact with personas. First, such functionality can be integrated directly into interactive persona systems [24, 25], where it logs every mouse movement and click of the user, thus providing a rich dataset of how the system is used, and how the personas are interacted with. Second, mouse-tracking is an unobtrusive form of measuring user behavior [4], as it does not interfere with the user's natural behavior, and it requires no specific calibration from the user. Finally, mouse-tracking enables carrying out remote user studies that are required in exceptional times of a global pandemic when physical user studies are not possible.

These properties make mouse-tracking a feasible form of persona analytics (PA). Therefore, contributing to the more effective measurement of personas for both remote and in-person user studies, we implement a fully functional PA subsystem within an interactive persona system. We demonstrate the capabilities of the PA system and discuss ways of using it for persona user studies.

2 RELATED LITERATURE

The phrase "data-driven personas" originates from McGinn and Kotamraju [27], although quantitative personas were proposed earlier in software requirements engineering [5, 6]. The idea of

CHI '21 Extended Abstracts, May 08–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8095-9/21/05…$15.00
https://doi.org/10.1145/3411763.3451773

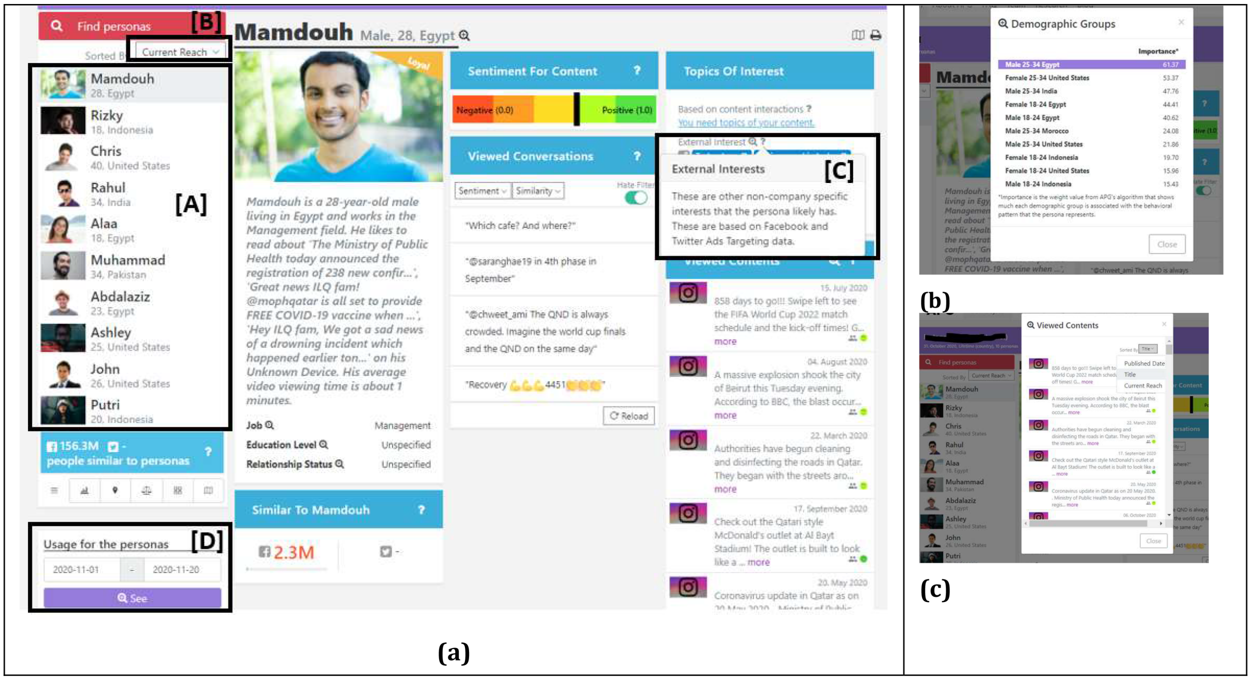

Figure 1: A data-driven persona profile. The left sidebar [A] shows the available personas on a given dataset. Users' mouse interactions (hovering, clicking) with the information elements ([B] [C] [D]) in the profile are tracked by Persona Analytics. The data is downloadable using the option in bottom-left corner, only available for system administrators.

using "data" for personas can be traced back to Cooper [12] who proposed that personas should be based on real user goals. The strive for "data" has consistently been repeated in persona literature [9, 10, 16, 17, 32, 50, 51]. However, until recently, the application of data-driven personas was limited.

According to Salminen et al. [34], three "availabilities" have caused a proliferation of data-driven personas: (1) availability of online user data from social media and online analytics platforms via their APIs, (2) availability of data science algorithms and libraries that can seamlessly be integrated into the data-driven persona creation process, and (3) availability of web technologies that surpass the limitations of paper as a medium of choice for personas. The latter removes the limitation of creating and presenting only a handful of personas, while offering the possibility to update the personas once they have been created, interaction techniques to drill down to the persona information and the option to make quantitative predictions [3]. Overall, these trends have been described transformational [28], signifying a transition from static "flat file" personas to "full-stack personas" that are traceable back to the unit-level data they were created from [19].

From data-driven personas, the next logical step of evolution is interactive persona system s [3, 28, 34], defined as interactive user interfaces (UI) that display persona profiles. This UI can, but not necessarily always, be accessed via web browsers [21, 22, 24, 25]. The benefits of web technologies are their broad applicability and accessibility. Personas served via the web can be accessed virtually from anywhere using any device that supports web browsing. Supporting technologies, such as user account management, can be integrated with relative ease using standard libraries and best practices. Interactivity enables persona users to perform actions on the personas, such as analyzing information on gender distributions, refreshing the persona quotes, filtering the quotes by sentiment and topic [42], and predicting a persona's interest for a given topic [2, 3].

Following these developments in data-driven personas and interactive persona systems, multiple opportunities can be envisioned. For example, interaction techniques and multimedia (e.g., persona chat/dialogue systems [11], video, AI agents . . .) could be incorporated into persona systems to serve various end-user needs [37]. New features for comparing personas by goal metrics [36] could be launched. Personas could be integrated into external system, such as recommendation, content management, and customer relationship management systems, and online advertising platforms [41]. However, the unifying factor behind these possibilities is the need for understanding the persona user behavior, which requires measurement.

3 PERSONA ANALYTICS

We define 'persona analytics' (PA) as *the systematic measurement of behaviors and interactions of persona users engaged with interactive persona systems*. When personas are provided through a web browser, PA takes place via mouse-tracking that records the persona users' mouse movements and clicks on the provided persona profiles, although other methods of analytics generation can be employed. Here, we present one possible implementation while

Table 1: Envisaged use cases (analytical questions) for Persona Analytics.

	Analytical Question	Informative for...
AQ01	What personas were most/least viewed?	Persona Creation
AQ02	What information was most/least viewed?	Persona Information Design
AQ03	What were the most/least common transitions between the personas?	Persona System Design
AQ04	What were the most/least common transitions between the information elements?	Persona Information Design
AQ05	What was the average number of persona visits per user?	Persona User Behavior
AQ06	What was the average persona system usage time per user?	Persona User Behavior
AQ07	What was the average number of visits per persona?	Persona User Behavior
AQ08	What was the average duration of visits per persona?	Persona User Behavior
AQ09	How many times (or how long) User A viewed Persona X?	Persona User Behavior
AQ10	*Comparisons*: How did different user groups (male/female; mature/young; software developers/marketers) differ in their usage behavior?	Persona User Segments

setting forth common motivations and metrics for future PA implementations.

3.1 Design Requirements

We defined specific analytical questions (AQ) for the PA to guide the analytics implementation (see Table 1). These use cases involve analytical questions that we deem important, as similar questions have been posed in empirical persona user studies [18, 35, 38, 39, 44, 46]. These AQs matter for two roles: researchers and analysts. Both the role types want to understand persona user behavior, but for different purposes. An academic researcher wants to understand fundamental patterns in persona user behavior. An analyst, in turn, wants specific answers to a given set of personas (e.g., "why was this persona considered interesting?").

For the PA system to be able to address these analytical questions, several requirements need to be satisfied. There is a need to capture (a) time spent per persona (based on the persona profile being active on the screen), (b) time spent in each persona information element (based on the mouse hovering over the element), and the (c) transition from one persona to another (based on clicks on the persona listing; see left sidebar in Figure 1). These metrics are inspired by previous persona user studies using a system and/or behavioral measurement [35, 38–40, 43, 44], deploying similar metrics that can be captured via mouse-tracking, including dwell time and sequence of persona information viewed. In the PA system, dwell time is defined as the time duration for which a user's mouse is placed over a screen element.

3.2 Implementation

In total, the PA system we developed tracks 223 UI elements. Seventy-six (34.1%) of these are click events. Other element types are section (n=120, 53.8%), tooltip (n=24, 10.8%), and input (n=3, 1.3%). Section items track the persona information elements that a user's mouse moves over. For example, [section:Left menu] tracks the user's interaction with the persona listing (see **[A]** in Figure 1a). Click events are recorded when the user clicks the persona profile. For example, [click:Persona > Headline > Demographic > More > Open] tracks the user's clicks for opening the additional demographic information of the persona (see Figure 1b).

Personas can be sorted by their audience size, and this behavior is tracked by [click:Left menu > Sorting] (see **[B]** in Figure 1a). PA also tracks the use of tooltips. Tooltips inform users on various persona information by explaining how the information was retrieved, providing transparency and explainability [42, 45]. For example, when users access the topics tooltip [tooltip:Persona > Viewed Contents > More > Popup > Content > Topic > Info], the system records this action (see **[C]** Figure 1a). Data exports of PA system (see **[D]** in Figure 1a) are available for system administrators. This functionality triggers a pop-up screen to download the mouse-tracking data. Finally, input elements track how users sort the persona information when accessing data breakdowns. Figure 1c shows the behavior of [input:Contents > Sorting].

Furthermore, PA tracks the user's interaction with each persona profile. Figure 2 shows the tracked information elements, which are: **A:** Headline (name, gender, age, country); **B:** About (picture, text description, job, education level, relationship status); **C:** Audience Size; **D:** Sentiment; **E:** Viewed Conversations; **F:** Topics of Interest; and **G:** Viewed Contents.

The content of these information elements is explained in related work [2, 3]. Each information element has child elements. For example, "About" (B in Figure 2) has the child elements of picture, text description, job, education level, and relationship status. The PA system tracks both the parent and child elements. The PA reports include a column "Hierarchy" with values "Parent" and "Child". Using this column, an analyst can choose to examine parent or child elements separately, and thus avoid duplicated dwell times.

Analysts can access the reports from the persona system's UI and by downloading the reports in a CSV file. Five report types are automatically generated in the data export:

1. **Persona report (grouped)** = contains data on what personas were visited, for how long, and how many times
2. **Persona report (ungrouped)** = contains data on the sequence of personas visited, with dwell times
3. **Information report (grouped)** = contains data on what persona information was visited, for how long, and how many times
4. **Information report (ungrouped)** = contains data on the sequence of persona information visited, with dwell times

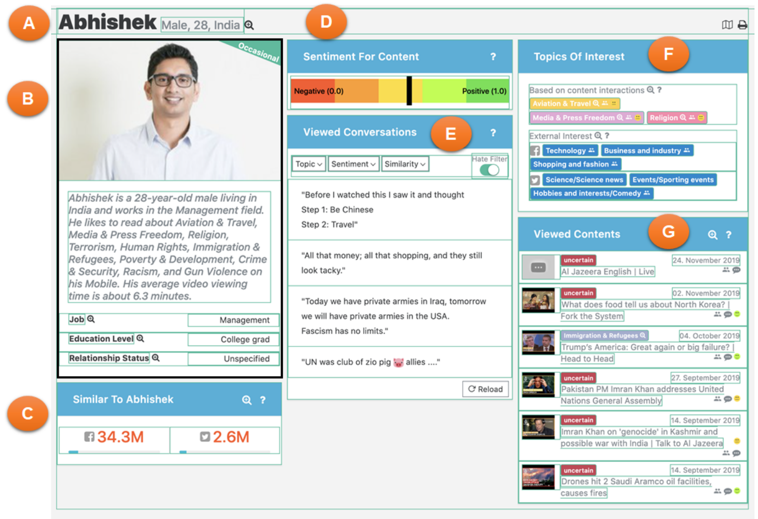

Figure 2: Visualization of the captured information elements. The borders indicate the elements being captured by the mouse tracking. As can be seen, the entire persona profile has been encoded with tracked elements.

Table 2: Validation tasks to validate the performance of PA.

	Task	Purpose
VT01	Visit (by clicking) all personas except the fifth persona in the listing	test of event logging
VT02	Visit (by clicking) the personas in order from top to bottom	test of transition logging
VT03	View all main information elements (defined in Fig. 2) in the third persona profile ("Rahul") by placing your mouse cursor on top of the information element while reading it	test of profile information logging

5. **Raw data** = all recorded mouse interactions with the specified elements and features of the persona system in chronological order. Analysts can use the raw logs for complicated analyses that require minute data.

The reports also include User IDs, Session IDs, Persona IDs, and precise timestamps of each logged event. Sample reports are provided in the Supplementary Material for the reader's convenience.

4 EVALUATION

In this section, we present results of a study carried out to evaluate the PA system's functionality. Two forms of evaluation are carried out: (a) internal technical evaluation that serves the purpose of controlled checks to ensure data is recorded appropriately, and (b) external user evaluation, where a user located in a different country carries out a task independently and the user's behavior is examined.

4.1 Internal Validation

For internal validation checks, ten personas were generated from the publicly available Instagram data of a mobile telephone operator (the method of persona generation is reported and validated in previous work [2, 3]). Three validation tasks (VT) were created to test different aspects of the PA system (see Table 2).

Table 3: Event logging results (VT01). In the test, the fifth persona was not visited, but all other personas were. The logs correctly show zero recorded events for the fifth persona and a varying number of events for the rest.

Persona	Persona profile events
Mamdouh	14
Abderaouf	5
Rahul	104
Abdalaziz	5
Muhammad	0
Bander	18
Ashley	6
Allaa	11
Mehmet	15
Hisham	20

One of the researchers carried out the three VTs. Results (Tables 3-5) show that all the tests were passed.

4.2 User Evaluation

We recruited one user to study how he reacts when using a persona analytics system. Although we are aware that having a single

Table 4: Testing the order of persona visits (VT02). The 'expected' column indicates the expected result based on the task requirements, and the 'observed' column indicates the observed result. The conditions match 100%.

Expected order	Observed order	Result
Mamdouh	Mamdouh	OK
Abderaouf	Abderaouf	OK
Rahul	Rahul	OK
Abdalaziz	Abdalaziz	OK
Bander	Bander	OK
Ashley	Ashley	OK
Allaa	Allaa	OK
Mehmet	Mehmet	OK
Hisham	Hisham	OK

participant can be considered as a drawback, we argue that one user is enough to produce enough electronic data to conduct an illustrative analysis of persona-user interaction via PA. The individual recruited was a Finnish male between the age of 40-50. He exemplifies a business user with pre-existing knowledge and experience in the use of personas (but no previous experience in using the persona system).

A realistic work-task scenario was crafted through discussions with the user, and the finalized scenario focused on the tourism sector because this context was of interest to the user. The persona system was used to generate personas using publicly available data from a specific Instagram channel of interest to the user. Ten personas were generated, as this corresponds to the system default, and a user account was created for the test user to access the system. The user's consent was acquired to record the mouse movement. The user then browsed the personas and provided written feedback on them and on the system. This evaluation focuses on analyzing the behavioral data of the user, and we leave the written feedback for future work.

Overall, the user spent 58.5 minutes browsing the personas. The user interacted with all the personas (see Figure 4), although the time spent per persona and the number of visits per persona varied greatly. Moreover, the personas were not visited in the order of presentation (from top to bottom), but instead in a non-linear fashion. Kendall tau Rank Correlation (τ = 0.244, p = 0.37) indicates a lack of dependence between the order of personas in the system and the observed order of viewing. The following chain ("persona n-gram") indicates the user's order of visiting different personas:

Jenni → Tiina → Jenni → Jennifer → Jenni → Jennifer → Marjatta → Ashley → Jennifer → Chris → Jenni → Ashley → Jennifer → Chris → Mikko → Henna → Mikko → Jenni → Chris → Jennifer → Ashley → Jennifer → Chris → Jennifer → Ashley → Tiina → Minna → Leena → Marjatta

Using such persona n-grams, a state transition matrix can be created and Marko Chain analyses can be deployed (cf. [14, 31, 47, 49]). This opens a major linkage to tools for analyzing persona user behavior. For example, *which personas are most frequently visited? Which are most often compared against one another?*

Altogether, the user made 29 transitions from one persona to another. From the data, we observed that the last visit was clearly an outlier by dwell time (12,137.9 seconds, which is more than 100 times higher than the average dwell time per persona when excluding this persona). In this case, the user most likely left the computer for a long time and returned later to log-out (thus triggering the exit event for logging the persona dwell time). This highlights the difficulty of tracking the last persona's dwell time accurately and the importance of manually inspecting the data. To remove the outliers' impact on results, we carry out imputation to replace with the mean dwell time per persona, calculated from the other personas. All aggregate dwell times reported are based on this outlier-corrected imputation.

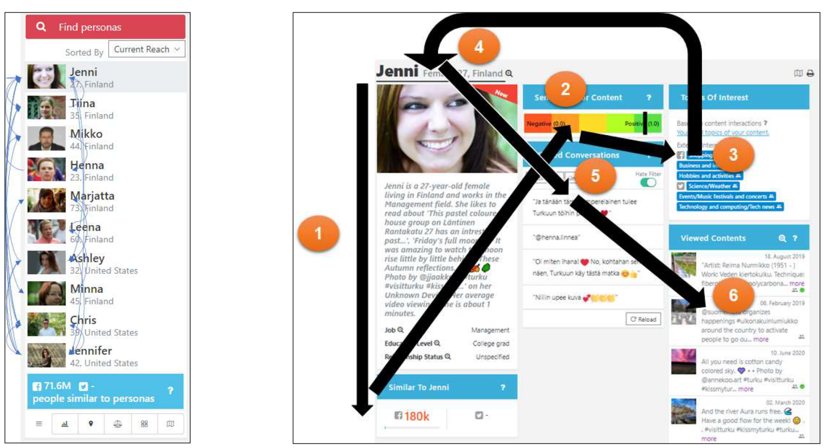

Figure 3: Left side: Reconstruction of the user's persona transition path, indicating a non-linear viewing behavior – the user does not click the personas in order from top to bottom but instead switches between them in a highly sporadic fashion. Right side: Reconstruction of the user's mousepath based on the logged data. Arrows indicate the direction of mouse movement. Numbers indicate the typical order of moving the mouse from one element to another.

Table 5: Information logging results (VT03). All the information elements reported at least one visit.

Persona	Persona information	Frequency of visits	Result (OK: >0)
Rahul	Headline	3	OK
Rahul	About	27	OK
Rahul	Audience Size	3	OK
Rahul	Sentiment	1	OK
Rahul	Viewed Conversations	41	OK
Rahul	Topics of Interest	9	OK
Rahul	Viewed Contents	14	OK

Table 6: User's mouse movement over different persona information.

Persona Information	Dwell time	Frequency of visits
About	848.212	10
Audience size	141.476	3
Topics of Interest	97.457	4
Viewed Conversations	50.715	4
Viewed Contents	23.877	2
Sentiment	2.344	3
Headline	0.735	3

The descriptive statistics of the user's dwell time per persona (M=121.1s, SD=204.1s) indicate strong variation, as the standard deviation is considerably higher than the mean. In other words, the user's attention is divided unevenly among the personas. The same is true for dwell time of specific persona information within the profile (M=7.5s, SD=61.7s), with details shown in Table 6

5 DISCUSSION

We have demonstrated a comprehensive solution to tracking user interactions with an interactive persona system. This solution enables precise measurement of user behavior for remote user studies (important for special circumstances such as during global pandemics), and opens several avenues for persona user studies that can contribute to our understanding of some fundamental questions in HCI and UCD: *How are personas actually used? What information do users most interact with? How do users browse and select personas?*

PA offers several advantages over ready-made solutions like Google Analytics (GA): (1) data ownership (no need to send data to a third-party server), (2) complete customizability of the reports (e.g., creation of persona n-grams), and (3) access to raw clickstream data if needed for more detailed analyses (not available in standard implementation of GA). Finally, (4) persona n-grams can be readily created from the PA's output, which would not be the case for the outputs of a standard GA implementation.

6 FUTURE DEVELOPMENT AND RESEARCH

One challenge is that the user's attention may not correlate with mouse movement [7, 29]. Nonetheless, implementing mouse dwell time already prepares the database structure and data processing for the possible deployment of eye-tracking in PA via a webcam in the future. An advantage of mouse-tracking is that it is an unobtrusive form of tracking natural user behavior. Other advantages are that (unlike eye-tracking) it requires no calibration, while giving accurate numbers of duration and sequence of visiting specific personas, which is something that eye-tracking can only give after manual annotation.

Another development item is showing elaborate reports directly in the persona system's UI – currently, users are required to download the reports and analyze them in a spreadsheet software. However, charts and tables summarizing the persona users' behavior could be presented directly in the system.

Exciting research opportunities are available with the introduction of PA. Calculating time spent per persona enables us to analyze how long a user investigated a certain persona profile. Behavioral topics such as order effects [13], revisit frequency, persona comparisons, satisficing behavior [48], and choice can be investigated deploying the persona state-transition matrix and Markov Chain techniques [20]. Persona information design can be informed by dwell time analyses, and typical persona viewing patterns and information viewing patterns can be deduced in interactive persona user studies using a live system. As such, the PA system opens up substantial avenues for future research.

REFERENCES

[1] Tamara Adlin and John Pruitt. 2010. *The Essential Persona Lifecycle: Your Guide to Building and Using Personas* (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[2] Jisun An, Haewoon Kwak, Soon-gyo Jung, Joni Salminen, and Bernard J. Jansen. 2018. Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. *Soc. Netw. Anal. Min.* 8, 1 (December 2018), 54. DOI: https://doi.org/10.1007/s13278-018-0531-0

[3] Jisun An, Haewoon Kwak, Joni Salminen, Soon-gyo Jung, and Bernard J. Jansen. 2018. Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data. *ACM Transactions on the Web (TWEB)* 12, 4 (2018), Article No. 27. DOI: https://doi.org/10.1145/3265986

[4] Bonnie Brinton Anderson, Anthony Vance, C. Brock Kirwan, Jeffrey L. Jenkins, and David Eargle. 2016. From warning to wallpaper: Why the brain habituates

to security warnings and what can be done about it. *Journal of Management Information Systems* 33, 3 (2016), 713–743.

[5] M. Aoyama. 2005. Persona-and-scenario based requirements engineering for software embedded in digital consumer products. In *Proceedings of the 13th IEEE International Conference on Requirements Engineering (RE'05)*, Washington, DC, USA, 85–94. DOI: https://doi.org/10.1109/RE.2005.50

[6] M. Aoyama. 2007. Persona-Scenario-Goal Methodology for User-Centered Requirements Engineering. In *Proceedings of the 15th IEEE International Requirements Engineering Conference (RE 2007)*, Delhi, India, 185–194. DOI: https://doi.org/10.1109/RE.2007.50

[7] Ernesto Arroyo, Ted Selker, and Willy Wei. 2006. Usability tool for analysis of web designs using mouse tracks. In *CHI'06 extended abstracts on Human factors in computing systems*, 484–489.

[8] Asa Blomquist and Mattias Arvola. 2002. Personas in action: ethnography in an interaction design team. In *Proceedings of the second Nordic conference on Human-computer interaction*, ACM, Aarhus, Denmark, 197–200. Retrieved May 28, 2017 from http://dl.acm.org/citation.cfm?id=572044

[9] Christopher N. Chapman, Edwin Love, Russell P. Milham, Paul ElRif, and James L. Alford. 2008. Quantitative Evaluation of Personas as Information. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1107–1111. DOI: https://doi.org/10.1177/154193120805201602

[10] Christopher N. Chapman and Russell P. Milham. 2006. The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 634–636. DOI: https://doi.org/10.1177/154193120605000503

[11] Eric Chu, Prashanth Vijayaraghavan, and Deb Roy. 2018. Learning Personas from Dialogue with Attentive Memory Networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Brussels, Belgium, 2638–2646. Retrieved June 20, 2019 from https://www.aclweb.org/anthology/D18-1284

[12] Alan Cooper. 1999. *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity* (1 edition ed.). Sams - Pearson Education, Indianapolis, IN.

[13] Hermann Ebbinghaus. 2013. Memory: A contribution to experimental psychology. *Annals of neurosciences* 20, 4 (2013), 155.

[14] Stephen Fitchett and Andy Cockburn. 2012. Accessrank: predicting what users will do next. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2239–2242.

[15] Erin Friess. 2012. Personas and Decision Making in the Design Process: An Ethnographic Case Study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), ACM, New York, NY, USA, 1209–1218. DOI: https://doi.org/10.1145/2207676.2208572

[16] Jonathan Grudin. 2006. Why Personas Work: The Psychological Evidence. In *The Persona Lifecycle*, John Pruitt and Tamara Adlin (eds.). Elsevier, 642–663. DOI: https://doi.org/10.1016/B978-012566251-2/50013-7

[17] Jonathan Grudin and John Pruitt. 2002. Personas, participatory design and product development: An infrastructure for engagement. In *Proceedings of Participation and Design Conference (PDC2002)*, Sweden, 144–161.

[18] Charles G. Hill, Maren Haag, Alannah Oleson, Chris Mendez, Nicola Marsden, Anita Sarma, and Margaret Burnett. 2017. Gender-Inclusiveness Personas vs. Stereotyping: Can We Have it Both Ways? In *Proceedings of the 2017 CHI Conference*, ACM Press, Denver, Colorado, USA, 6658–6671. DOI: https://doi.org/10.1145/3025453.3025609

[19] Bernard J. Jansen, Joni O. Salminen, and Soon-gyo Jung. 2020. Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics. *Data and Information Management* 4, 1 (2020). Retrieved from https://content.sciendo.com/view/journals/dim/4/1/article-p1.xml

[20] Jinyuan Jia, Binghui Wang, Le Zhang, and Neil Zhenqiang Gong. 2017. AttriInfer: Inferring User Attributes in Online Social Networks Using Markov Random Fields. In *Proceedings of the 26th International Conference on World Wide Web* (WWW '17), International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 1561–1569. DOI: https://doi.org/10.1145/3038912.3052695

[21] Soon-gyo Jung, Jisun An, Haewoon Kwak, Moeed Ahmad, Lene Nielsen, and Bernard J. Jansen. 2017. Persona Generation from Aggregated Social Media Data. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '17), ACM, Denver, Colorado, USA, 1748–1755.

[22] Soon-gyo Jung, Joni Salminen, Jisun An, Haewoon Kwak, and B. J. Jansen. 2018. Automatically Conceptualizing Social Media Analytics Data via Personas. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2018)*, San Francisco, California, USA.

[23] Soon-gyo Jung, Joni Salminen, and Bernard J. Jansen. 2019. Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI EA '19), ACM, Glasgow, UK, LBW2714:1–LBW2714:6. DOI: https://doi.org/10.1145/3290607.3312955

[24] Soon-Gyo Jung, Joni Salminen, and Bernard J. Jansen. 2020. Giving Faces to Data: Creating Data-Driven Personas from Personified Big Data. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion* (IUI '20), Association for Computing Machinery, Cagliari, Italy, 132–133. DOI: https://doi.org/10.1145/3379336.3381465

[25] Soon-gyo Jung, Joni Salminen, Haewoon Kwak, Jisun An, and Bernard J. Jansen. 2018. Automatic Persona Generation (APG): A Rationale and Demonstration. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, ACM, New Brunswick, NJ, USA, 321–324. DOI: https://doi.org/10.1145/3176349.3176893

[26] Nicola Marsden, Monika Pröbster, Mirza Ehsanul Haque, and Julia Hermann. 2017. Cognitive styles and personas: designing for users who are different from me. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, ACM, Brisbane, Queensland, Australia, 452–456.

[27] Jennifer Jen McGinn and Nalini Kotamraju. 2008. Data-driven persona development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Florence, Italy, 1521–1524. DOI: https://doi.org/10.1145/1357054.1357292

[28] T. Mijač, M. Jadrić, and M. Ćukušić. 2018. The potential and issues in data-driven development of web personas. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1237–1242. DOI: https://doi.org/10.23919/MIPRO.2018.8400224

[29] Vidhya Navalpakkam and Elizabeth Churchill. 2012. Mouse tracking: measuring and predicting users' experience of web-based content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2963–2972.

[30] Lene Nielsen. 2019. *Personas - User Focused Design* (2nd ed. 2019 edition ed.). Springer, New York, NY, USA.

[31] Kazuhiro Otsuka, Junji Yamato, Yoshinao Takemae, and Hiroshi Murase. 2006. Quantifying interpersonal influence in face-to-face conversations based on visual attention patterns. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, 1175–1180.

[32] John Pruitt and Jonathan Grudin. 2003. Personas: Practice and Theory. In *Proceedings of the 2003 Conference on Designing for User Experiences* (DUX '03), ACM, San Francisco, California, USA, 1–15. DOI: https://doi.org/10.1145/997078.997089

[33] Kari Rönkkö. 2005. An Empirical Study Demonstrating How Different Design Constraints, Project Organization and Contexts Limited the Utility of Personas. In *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences - Volume 08* (HICSS '05), IEEE Computer Society, Washington, DC, USA. DOI: https://doi.org/10.1109/HICSS.2005.85

[34] Joni Salminen, Kathleen Guan, Soon-gyo Jung, Shammur Absar Chowdhury, and Bernard J. Jansen. 2020. A Literature Review of Quantitative Persona Creation. In *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu, Hawaii, USA, 1–14. DOI: https://doi.org/10.1145/3313831.3376502

[35] Joni Salminen, Bernard J. Jansen, Jisun An, Soon-gyo Jung, Lene Nielsen, and Haewoon Kwak. 2018. Fixation and Confusion – Investigating Eye-tracking Participants' Exposure to Information in Personas. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2018)*, ACM, New Jersey, USA, 110–119. DOI: https://doi.org/10.1145/3176349.3176391

[36] Joni Salminen, Bernard J. Jansen, Jisun An, Haewoon Kwak, and Soon-gyo Jung. 2018. Are personas done? Evaluating their usefulness in the age of digital analytics. *Persona Studies* 4, 2 (November 2018), 47–65. DOI: https://doi.org/10.21153/psj2018vol4no2art737

[37] Joni Salminen, Bernard J. Jansen, Jisun An, Haewoon Kwak, and Soon-gyo Jung. 2019. Automatic Persona Generation for Online Content Creators: Conceptual Rationale and a Research Agenda. In *Personas - User Focused Design* (2nd ed.), Lene Nielsen (ed.). Springer London, London, 135–160. DOI: https://doi.org/10.1007/978-1-4471-7427-1_8

[38] Joni Salminen, Soon-gyo Jung, Jisun An, Haewoon Kwak, and Bernard J. Jansen. 2018. Findings of a User Study of Automatically Generated Personas. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI EA '18), ACM, Montréal, Canada, LBW097:1–LBW097:6. DOI: https://doi.org/10.1145/3170427.3188470

[39] Joni Salminen, Soon-gyo Jung, Jisun An, Haewoon Kwak, Lene Nielsen, and Bernard J. Jansen. 2019. Confusion and information triggered by photos in persona profiles. *International Journal of Human-Computer Studies* 129, (September 2019), 1–14. DOI: https://doi.org/10.1016/j.ijhcs.2019.03.005

[40] Joni Salminen, Soon-gyo Jung, Shammur Absar Chowdhury, Sercan Sengün, and Bernard J Jansen. 2020. Personas and Analytics: A Comparative User Study of Efficiency and Effectiveness for a User Identification Task. In *Proceedings of the ACM Conference of Human Factors in Computing Systems (CHI'20)*, ACM, Honolulu, Hawaii, USA. DOI: https://doi.org/10.1145/3313831.3376770

[41] Joni Salminen, Soon-gyo Jung, and Bernard J. Jansen. 2019. The future of data-driven personas: A marriage of online analytics numbers and human attributes. In *ICEIS 2019 - Proceedings of the 21st International Conference on Enterprise Information Systems*, SciTePress, Heraklion, Greece, 596–603. Retrieved August 22, 2019 from https://pennstate.pure.elsevier.com/en/publications/the-future-of-data-driven-personas-a-marriage-of-online-analytics

[42] Joni Salminen, Soon-gyo Jung, and Bernard J. Jansen. 2020. Explaining Data-Driven Personas. In *Proceedings of the Workshop on Explainable Smart Systems for*

Algorithmic Transparency in Emerging Technologies co-located with 25th International Conference on Intelligent User Interfaces (IUI 2020) (Vol-2582), CEUR Workshop Proceedings, Cagliari, Italy, 7. DOI: https://doi.org/urn:nbn:de:0074-2582-4

[43] Joni Salminen, Ying-Hsang Liu, Sercan Sengun, João M. Santos, Soon-gyo Jung, and Bernard J. Jansen. 2020. The Effect of Numerical and Textual Information on Visual Engagement and Perceptions of AI-Driven Persona Interfaces. In *IUI '20: Proceedings of the 25th International Conference on Intelligent User Interfaces*, ACM, Cagliary, Italy, 357–368. DOI: https://doi.org/10.1145/3377325.3377492

[44] Joni Salminen, Lene Nielsen, Soon-gyo Jung, Jisun An, Haewoon Kwak, and Bernard J Jansen. 2018. "Is More Better?": Impact of Multiple Photos on Perception of Persona Profiles. In *Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI2018)*, ACM, Montréal, Canada. DOI: https://doi.org/10.1145/3173574.3173891

[45] Joni Salminen, Joao M. Santos, Soon-gyo Jung, Motahhare Eslami, and Bernard J. Jansen. 2019. Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas. *International Journal of Human–Computer Interaction* 0, 0 (November 2019), 1–13. DOI: https://doi.org/10.1080/10447318.2019.1688946

[46] Joni Salminen, Sercan Sengun, Soon-gyo Jung, and Bernard J Jansen. 2019. Design Issues in Automatically Generated Persona Profiles: A Qualitative Analysis from 38 Think-Aloud Transcripts. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR)*, ACM, Glasgow, UK, 225–229. DOI: https://doi.org/10.1145/3295750.3298942

[47] Abhraneel Sarma and Matthew Kay. 2020. Prior Setting In Practice: Strategies and rationales used in choosing prior distributions for Bayesian analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12.

[48] H. A. Simon. 1956. Rational choice and the structure of the environment. *Psychological Review* 63, 2 (1956), 129–138. DOI: https://doi.org/10.1037/h0042769

[49] Gang Wang, Xinyi Zhang, Shiliang Tang, Haitao Zheng, and Ben Y. Zhao. 2016. Unsupervised clickstream clustering for user behavior analysis. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 225–236.

[50] Xiang Zhang, Hans-Frederick Brown, and Anil Shankar. 2016. Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), ACM, San Jose, California, USA, 5350–5359.

[51] Haining Zhu, Hongjian Wang, and John M. Carroll. 2019. Creating Persona Skeletons from Imbalanced Datasets - A Case Study using U.S. Older Adults' Health Data. In *Proceedings of the 2019 on Designing Interactive Systems Conference - DIS '19*, ACM Press, San Diego, CA, USA, 61–70. DOI: https://doi.org/10.1145/3322276.3322285