The Choice of a Persona: An Analysis of Why Stakeholders Choose a Given Persona for a Design Task

Joni Salminen (email: jonisalm@uwasa.fi)[1], Sercan Şengün[2], João M. Santos[3], Soon-gyo Jung[4], Lene Nielsen[5], and Bernard Jansen[4]

[1] School of Marketing and Communication, University of Vaasa, Vaasa, Finland
[2] College of Fine Arts, Illinois State University, IL, USA
[3] Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal
[4] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
[5] IT University of Copenhagen, Copenhagen, Denmark

Abstract. Although personas have been applied for two decades, not much is known about why a designer chooses a specific persona for a given design task. This question matters because if designers prefer one persona over another, then the needs and attributes of that persona would be favored in the design process, resulting in possible "blind spots" and bias in regards to other personas. To explore reasons and behaviors associated with the choice of a persona, we conduct an on-site user study with 37 participants in a workplace setting focused on a social media content creation task. Our findings show that factors affecting the choice of persona include age similarity between the persona and the designer, persona's looks, how many users the persona represents, time spent browsing the persona information, and whether the persona is (a)typical relative to other personas. Under different persona sets, these factors were correlated with the probability of a persona being chosen for a design task, and also supported by a qualitative analysis of the think-aloud records where the participants explained their persona choice. The findings provide implications for developing interaction techniques that support users' varying information needs and persona selection strategies, including recommenders that would increase the match between the designers' information needs and the available personas.

Keywords: Personas · User choice · User study · Human factors.

1 Introduction

Personas are 'customer segments with a face,' i.e., fictitious people that represent key customer segments [31]. Therefore, personas inform stakeholders (i.e., designers, marketing managers, software developers, etc.) about end-user aims, interests, and actions through depicting central user characteristics [14]. Personas embody behaviors and traits – such as pain points, needs, and wants – of end-users that designers seek to address. Personas were originally devised to address the prejudices and self-serving biases of designers, who may create systems

that relate to their own interests without clear guidance on who the user is [11]. Representing customers as personas intends to minimize reliance on personal viewpoints when trying to understand end-users' experiences, perceptions, and mental processes [31].

Cooper [7] identifies persona creation as a powerful tool for communication and interaction design. This is because personas represent real people by outlining goals and preferences that are directly extracted from the experiences of real end-users. Nielsen [30] highlights that personas can be understood as a means of communicating customer data by focusing on field observations rather than preconceived ideas, thus improving customer understanding and leading to higher customer-centricity. In the industry, personas are well-established, and they have been broadly applied in software engineering, human-computer interaction (HCI), health informatics, journalism, cybersecurity/privacy, video games, marketing, and other domains [7,25,30,9,16,4]. To support these efforts, several variants of personas have been developed, such as design personas, marketing personas, buyer personas, segment personas, patient personas, cybersecurity/privacy personas, player personas, and so on. For example, companies such as Spotify[6] and Microsoft[7] use personas to enhance their understanding of users and customers. The applicability of the persona method is thus broad and personas remain topical in industry and research.

Nevertheless, how stakeholders engage with personas is not well understood. *What makes a persona interesting? Why do users focus on one persona over another? Why do users select a specific persona for their task?* These are some of the open questions that motivate our research.

Stakeholders are people that use personas to make decisions about users or customers. Typically, after persona creation, these stakeholders are presented with several personas, referred to as a 'persona set' or a 'cast of personas' [30]. According to our experience from empirical persona user studies, for a given design task, the users tend to browse the available personas and then focus on one persona at a time for their task. Sometimes different users may choose the same persona, but often they choose different personas. The intriguing question is: *Why?* The answer may enhance the theoretical understanding of the mechanisms of how personas are used in practice. Answers can also inform persona design on what kind of personas to develop, as well as contribute to theory development regarding interaction between personas and their users. Overall, understanding persona choice can inform the HCI community on three vital aspects:

1. **Creation of personas** – empirically analyzing users' persona choice gives an idea of what information users pay attention to in the persona profile, which has direct implications for persona design.
2. **Use of personas** – gives an idea of the users' voiced reasoning of who fits the task they are going to do and why, which can enhance theoretical

[6] `https://spotify.design/article/the-story-of-spotify-personas`

[7] `https://techcommunity.microsoft.com/t5/driving-adoption/driving-user-adoption-user-personas-and-user-journey-maps/m-p/82058`

understanding of how people relate to personas and actually use them for design.
3. **Theory of personas** – through understanding persona choice, we can take steps towards generating a broader theory for persona–user interaction, which the HCI literature is currently missing.

Persona choice matters because —- similar to the theory of selective attention [23] —- the choice of whom to design dictates everything that follows! So, the selection mechanism is of interest - it deals with the designer contemplating, "I think this persona is worthwhile, I should focus on it." But why that persona and why not the next one? Contrary to this being a trivial matter, it is actually a fundamental matter for persona use. In turn, persona use dictates whether personas are valuable and for whom. Whose interests are product designers promoting?

The bottom line – to use a persona for decision making, you need to first choose one. We could, in some settings, argue that designers can choose and consider the interests of many personas, but from an experimental standpoint, the most straightforward way to study this is to ask them focus on one persona.

Even though persona choice is an interesting and important topic that can shed light on the interaction between personas and marketing stakeholders, there is little work on this topic. The consequence from this lack of attention is that users' strategies, behaviors, and given rationales of choosing one persona over another are poorly understood. Not much is known about the process of the users forming a connection with a specific persona, and what human factors contribute to such connections. So, there is a lack of understanding as to why a user chooses a certain persona over another when carrying out a design task, forming the research gap that our study addresses. We empirically investigate the choice of personas, formulating hypotheses and having professional users conduct tasks with different sets of personas in workplace setting. Our overarching research question (RQ) is: *Why do stakeholders choose a certain persona for a marketing task?* We investigate three aspects of this choice process:

RQ1: What characteristics of (a) the persona and (b) the stakeholder explain the stakeholders' persona choice? We address this question through quantitative analysis.
RQ2: What interaction aspects, including (a) persona presentation in the user interface (UI) and (b) stakeholders' interaction with personas (i.e., dwell time, frequency of visits), explain stakeholders' persona choice? We address this question through quantitative analysis.
RQ3: What strategies do stakeholders apply for their persona choice? We address this question through qualitative analysis.

To address these questions, we conduct a mixed-method study, using quantitative and qualitative methods. We carry out an empirical user study using an interactive persona system deployed in a large non-profit organization. *Interactive persona system* refers to a Web-based system that allows users to browse personas and their information freely using mouse navigation and interaction

techniques, such as selection of persona, scrolling the information, viewing data distributions, and so on [18]. The study applies two persona sets: one with less diversity, and another with more diversity, in terms of age, gender, and locations of personas. This is achieved using a data-driven persona system to generate these two sets. Data-driven persona generation refers to using algorithms and statistic techniques for automatic or semi-automatic segmentation and enrichment of digital user data [34]. We recruit thirty-seven participants from this organization, who each choose a persona for a design task (designing online content for the freely-chosen persona). We test six hypotheses that potentially explain the participants' choice of personas. Our results inform persona design by offering guidance on what factors matter for persona choice, as well as shedding light on users' reasoning and experiences when employing a persona for design tasks.

2 Literature Review

2.1 Cross-Disciplinary Perspectives to Persona Choice

Personas are fictitious people characterizing core or target customer or user groups [7]. A persona has a name, picture, and written description – it is an alternative to nameless, faceless group of people [18]. Personas are said to enhance the consideration of customer needs among design teams and other professionals engaged in customer-centric decision making [26]. They group similar customers under one archetype, facilitating the understanding of customers' needs and wants [30]. While it may not be practical to consider hundreds or thousands of individual customers when making business decisions, considering a core set of personas is manageable [17]. This concept of manageability is pervasive in persona theory, and can be generalized as follows: *There is a lot of information about customers. That information is summarized into a set of personas through a process of segmentation and personification. Then, among this persona set, stakeholders learn from one or more personas and make actionable decisions.*

Previous research in multiple fields presents potential reasons for stakeholders' focus on a given persona for a marketing task. Particularly relevant fields include HCI, information science (IS), and social psychology (SP). Their views can complement each other to form a more holistic picture of how stakeholders interact and engage with personas. In particular, HCI research highlights the importance of professionals' empathy for personas and the underlying customer base for accomplishing user-centered design goals. The psychological study of person perception (i.e., how people form impressions of others) can be extended to personas, even though personas are fictive in nature [30]. HCI and IS tend to speak of people as "users," whereas in SP, the notion of humans is closer to "individuals." Central in the psychological view of personas is the concept of person perception, which refers to the general tendency to form impressions of other people, which is a facet that also affects how stakeholders perceive personas [3]. From the IS field, we gain foundational insights on how stakeholders process information to support their decision making [38].

3 Hypothesis Development for RQ2 and RQ3

3.1 Hypotheses About Persona and User Characteristics

H1: Users Are More Likely to Choose Personas that are Similar to the User. In social psychology, evidence points out that similarity with another person is associated with a positive attitude towards that person [27]. This is referred to as *homophily* [20], defined as the "tendency of individuals to associate and bond with similar others" [28]. Here, we measure similarity in terms of demographic matching between all participant-persona pairs (cf. [40]). We focus on two traits: gender (H1a: Users are more likely to choose a persona from their own gender) and age (H1b: Users are more likely to choose a persona with an age similar to their own age). Initially, we also wanted to include the country for this analysis, but the personas did not have enough geographic variation to make this test possible.

H2: Users Are More Likely to Choose Attractive Personas. According to the *"what is beautiful is good"* effect, individuals perceive attractive people as having more desirable interpersonal traits. This results into higher willingness to form social bonds with these people relative to less attractive individuals. This concept is also known as the "physical attractiveness" bias, and it originated from a study by Dion et al. [8]. In the case of personas, we surmise that more attractive pictures increases the persona's probability of being chosen, so that the personas are chosen because they appear as more physically attractive than other personas.

H3: Users Are More Likely to Choose Personas that are Different from Others. The observed *salience* of an item is the state or quality by which it stands out from its neighbors [15]. The generated personas are all slightly different, but we label those are distinctly different from others as outliers. We then test if these outlier personas are more or less likely to be chosen by the participants. It is expected that, due to these personas being different, they are treated differently than other personas. Here, we classify a persona as an outlier based on their age, gender, nationality, and audience size relative to the entire persona set, with audience size referring to how many people the persona represents (see the method section for details).

H4: Users Are More Likely to Choose Personas with a High Segment Representation. The effect of *popularity*, arising from childhood where children learn that popularity is a desirable quality [22], drives social behavior from an early age. According to this notion, individuals are more willing to associate themselves with individuals that they perceive as popular or important than with individuals that lack these qualities [32]. In our case, the participants may presume that a persona with a large segment size is more popular or represents a larger group of people, and therefore should be chosen. The number of people the persona represents may be a predominant characteristic motivating the choice of a persona. This is because the user may presuppose that by choosing a persona that represents a large group of people, they are able to reach more people with their message because of the target's large audience rep-

resentation. In other words, this is a form of marketing logic [21], which may be relevant considering the type of task employed in the experiment (see the method section).

3.2 Hypotheses About System Features and User Behavior

H5: Users Are More Likely to Choose Personas that Appear either (a) Higher or (b) Higher and Lower in the Order of Presentation. In his work dating to 1885, Ebbinghaus [10] observed a relationship between recall and serial position, subsequently becoming a major benchmark for future studies. In his work of word list learning, Ebbinghaus proposed a U-shaped curve of recall, with the first and last items in a list being best remembered, referred to as primacy effect and recency effect, respectively.

In general, order effects refer to differences in individuals' responses that result from the order (e.g., first, second, third) in which the experimental materials are presented to them [39]. In the interactive persona system deployed in the current study, the personas are displayed in a user interface sorted by their representativeness of the data. We test two order effects: the *primacy effect* and the *serial-position effect*. The primacy effect is the tendency to remember the first piece of information better than information presented later on [33]. The personas are presented in a listing, so this effect suggests personas first in the listing are more likely to be chosen. According to this idea, a persona is chosen because participants see it sooner than the other personas. The participants either interpret the first seen information as more important, or simply remember it better than the subsequent information [33]. In turn, serial-position effect is the tendency of a person to recall the first and last items in a series best, and the middle items worst [29]. This effect suggests personas higher and lower in the listing are more likely to be chosen because, similarly to the primacy effect, these personas are more memorable due to their order of presentation.

H6: Users Are More Likely to Choose Personas They Engage Most with Based on (a) Dwell Time and (b) Number of Visits. The use of an interactive persona system allows us to measure user interaction (engagement) with the personas, which is an important aspect when personas are integrated into tools or systems that can be supervised [12].

From this idea, we surmise that the personas that are chosen are most often visited or viewed for a longer duration than other personas. The motivation for this hypothesis is given by the *uncertainty reduction theory* [5], which states that individuals reduce uncertainty about others by gaining information about them. We expect that the users engage more with the personas they end up choosing relative to other personas. According to this idea, users spend time viewing the personas' information, developing a preference for the personas they visit more often (or spend more time with), leading to positive affirmation and choice. Therefore, we surmise that more visits and higher dwell times for a persona increase its probability of being chosen. The difference of uncertainty reduction and mere-exposure effect [41] – which could also be seen as a relevant rationale for this hypothesis – is that the latter deals with how much a user would be exposed

to the persona is that in our research design, the participants were actively deciding the number and duration of persona visits, rather than being passively exposed to them. Therefore, we presume that more visits for a given persona took place because the participant was interested in that persona and wanted to learn more, which is compatible with the premise of uncertainty reduction.

4 Methodology

4.1 Research Site and Participants

Our data collection site was a major non-profit organization that advances important programs in education, research, and public health. We chose this organization because of their past experience of working with personas. The organization uses personas (a) to better understand their online audience, and (b) for strategic planning of online content design, which involves crafting communication policies and content to serve the various stakeholder groups of the organization. In total, there were 37 participants, of which 10 (27%) were females. The average age of the participants was 32.9 years (SD = 6.9). The participants held a variety of job positions within the organization, including data analysts, engineers, software developers, researchers, editors, social media managers, copywriters, project managers, and content specialists. The participant pool thus represents the myriad of positions dealing with creating end-user experiences in large organizations, involving people with varied backgrounds and expertise. The participants' earlier experience of personas was varied, including most having conceptual experience (i.e., knows what personas are but has not used them previously) (71%, n = 26), a little less than third (27%, n = 10) having some practical experience (has applied personas before, but not often), and one (3%, n = 1) having extensive experience (has frequently used personas in their job). Each participant was explained the foundational concept of personas regardless of their level of previous experience. Furthermore, it was clarified that the personas they were about to see were based on their organization's actual YouTube end-user data, i.e., that they were data-driven.

4.2 Persona Creation and Interactive Persona System

We generated the personas using a data-driven persona methodology reported and validated in previous work [1,2]. Several other persona experiments have applied this methodology [36,37,35], as it affords a standardized way for generating personas from real end-user data, enables users to interact with the personas, and records the interactions users have with personas in the system logs. The data-driven personas were automatically generated from the focal organization's YouTube end-user statistics using an algorithmic process and a system that has been validated by previous research [1,2,19]. The data for the persona creation comprised 1,473,275 view counts on 125 videos thru December 31, 2019 that were retrieved automatically via the YouTube Analytics API[8]. This data was

[8] `https://developers.google.com/youtube/analytics`

automatically organized in the form of an interaction matrix, and decomposed using non-negative matrix factorization [24]. Then, the system created the personas by automatically incorporating demographically tagged name, picture, and other information from a database. As the underlying analytics platform current defines only two genders, that is the number we used in this research, deferring other gender identities for future work.

The personas were provided to the participants using an interactive persona system, giving users the option to freely browse the personas and their information. The personas were shown to the participants using the interactive persona system, available at `https://persona.qcri.org`. The participants could freely interact with the generated personas, including switching between the personas, scrolling their information, learning about information definitions, reading the quotes, and so on. The persona information contained a (1) name, (2) demographics (age, gender, country), (3) picture, (4) text description, (5) sociographics (job, marital status, education), (6) sentiment, (7) topics of interest, (8) quotes, (9) most viewed content, and (10) audience size – i.e., the number of people the persona represents. For the experiment, two sets of personas were generated, as explained in the following section.

4.3 Experiment Design

In the experiment, each user went through two sessions of first using the interactive persona system and then carrying out the design task (also referred to as work task scenario). Because each user used the system twice with a different set of personas, the study design corresponds to within-subjects experiment – which is beneficial for mitigating the impact of individual user behaviors on the results. For both sessions, the participant was presented with a work task scenario (WTS) before being shown the system:

> *"Your task is to promote the [organization] as a workplace to a specific persona. A persona is a fictitious person that describes a real user segment. The personas you will see are created from the real audience data from [organization]'s YouTube channel. They represent [the organization]'s audience segments in YouTube."*

In the WTS, participants engaged with the interactive persona system to review personas and select the persona for which they were creating a YouTube video. This can be considered as a content design task in social media management, i.e., designing content for a specific target group. The organization suggested using this task because they perceived it natural for their intended use of personas. Thus, the task reflects a real use case of personas in an organization. For the within-subjects design, we used the persona generation system to automatically generate two persona sets of different number of personas. As a common practice, a persona set contains 3-7 personas [6]. We created such a set (**PS1** that contained 5 personas) first. However, due to the nature of our research question, we also needed a larger persona set to better capture variation among the persona attributes. Therefore, we created another set (**PS2**)

that contained 15 personas and had a higher degree of demographic diversity than the **PS1**. Both sets were automatically generated by the persona system from the same data collected from YouTube Analytics that represented the audience population watching the videos in the organization's YouTube channel. As we wanted the personas to remain data-driven, no other manipulations to the personas were made, apart from varying the number of personas generated, which afforded different personas for both sessions. Personas are illustrated in Figure 1 and Figure 2. The persona profiles with full information are shown in the Supplementary Material[9].

(a) Jafar (b) Ghada (c) Bassam (d) Faisal (e) Osamah

Fig. 1: Pictures of the personas in PS1. Jafar (a) and Bassam (c) were chosen the most often.

(a) Naif (b) Rami (c) Sami (d) Rajab (e) Imran

(f) Cambell (g) Jameela (h) Maryam (i) Michael (j) Faleh

(k) Nada (l) Alanood (m) Huda (n) Noura (o) Fatima

Fig. 2: Pictures of the personas in PS2. All personas, except Sami (c) and Alanood (l), were chosen at least once. Imran (e) was chosen more often than the other personas.

We counterbalanced the order of presentation, so that each participant was randomly assigned to either first seeing the **PS1** and then the **PS2**, or *vice versa*. For this, we manually created two different experimental sequences – in Sequence 1 the participant first saw the **PS1** and then the **PS2**; in Sequence 2, the order is the opposite. The purpose of this is simply to avoid the results from

[9] `https://www.dropbox.com/sh/gtuopopbqwgxjbw/AABPBK8KByeX2rvo3uOTsGNda?dl=0`

being biased by all the participants first seeing the set with a less diverse set of personas and then the more diverse set of personas. Instead, half will the see less diverse set first and the other half will see the more diverse set first. During the experiment, this was controlled by logging the user into a correct sequence and keeping record of which participant was allocated to which sequence.

4.4 Data Collection and Pre-processing

We gathered two main types of data: (1) system logs recorded the participants' engagement with the personas (the duration and number of visiting each persona), and (2) survey data was collected after task completion to collect user characteristics and their persona choices. After using the persona system, the participants selected a specific persona for their task. For statistical analysis, the data was structured so that each observation represents a participant session and a persona. For each such combination, two outcomes are possible – either the persona is chosen, or it is not. A 'Choice' variable was created, assuming a value of 1, if the participant chose the persona, or a 0, if the participant did not choose the persona. This was used as the dependent variable. The independent variables were:

- **Age delta** – the absolute difference between the persona and the participant's age for that combination;
- **Same Gender** – 1 if the participant and the persona are of the same gender;
- **Attractiveness** – a rating of attractiveness which was ascribed to the persona by an independent sample;
- **Outlier** – 1 if the persona is classified as an outlier;
- **Audience size** – indicates how many people are similar to the persona (according to data from Facebook Marketing API). The log of this variable was used to address its skewed distribution;
- **Order** – the relative presentation order of the persona (lower numbers indicating a higher position of the presentation order, i.e. 1 is the first, 2 is the second, etc.);
- **Visits** – the number of times the participant visited the persona's page; and
- **Dwell time** – the combined duration, in seconds, of all the participant's visits to the persona's page.

To isolate the effect of pictures, we recruited an independent sample of 50 people via the Prolific[10] survey platform to rate the attractiveness of the personas' facial pictures. The participants rated each picture using a Likert scale (1: Not at all – 5: Very much). Based on the responses, we assigned a mean attractiveness score for each persona. To address H6a, we calculated the timestamp differences between persona visits in the system log data to get the duration each persona was perused. We then compared that number with the ratio of each persona being chosen for the design task. From the log data, we also computed the number of visits (H6b) for each of persona-participant pair.

[10] `https://prolific.ac`

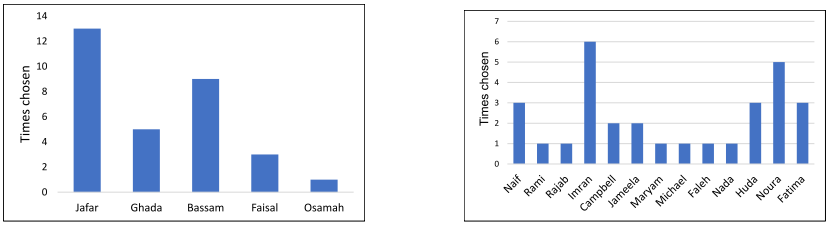
To test the outlier effect, we classified seven personas as outliers based on the following criteria. Three were in the **PS1** (Ghada because of gender; Faisal because of age; and Osamah because of country), and four were in the **PS2** (Cambell because of country, Michael because of country, Nada because of age (young), and Alanood because of age (elderly)).

To test the order effects, we computed the probability of a given persona being chosen, and correlated that probability with the persona's order using Spearman's correlation coefficient (ρ).

5 Exploratory Data Analysis

For the **PS1**, users' persona choices are strongly centered to two personas (see Figure 4). Together, Jafar and Bassam are chosen more than half of the time (59%, n=22). Osamah was chosen only once (3%). The tendency to select only a couple of personas did not appear in the **PS2**, but the choices were more evenly distributed. To quantify the variability of users' persona choices, we computed the *relative standard deviation (RSD)*, a standardized measure of variability of a distribution (in this case, the probabilities of each persona being chosen).

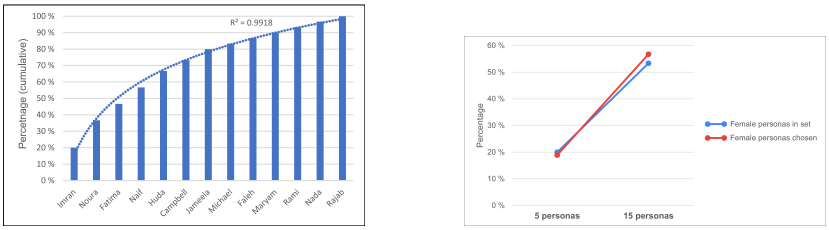
The values – obtained by dividing standard deviation of the probability distribution by its mean – indicate that the choice variation is higher for the **PS2** (RSD = 0.837) than for the **PS1** (RSD = 0.695). Therefore, more personas appear to increase the variability of users' persona choices. In the **PS1**, all personas were chosen at least once. In the **PS2**, 13 personas (86.7%) were chosen at least once. This indicates that increasing the number of personas results in users collectively making use of more personas – otherwise, there would be a stronger tendency of choosing just a few personas even with the **PS2**. The tendency of the users to make use of more personas is illustrated in Figure 4a.



(a) Persona Set 1 (b) Persona Set 2

Fig. 3: The number of times a persona was chosen in the two sets (x axis follows the order of personas shown in the system listing). Visually, the choices in the **PS1** seem to center around two personas, whereas in the **PS2** they are more evenly distributed.

We also contemplated that the gender of the persona would be an influential factor in the choice because, in the **PS1**, the most popular personas (Jafar and Bassam) are both male. However, the choices in the **PS2** refute this idea,



(a) (b)

Fig. 4: **(a)** Cumulative percentage of selected personas in the PS2. The dotted line describes the logarithmic function that is well matching the data ($R^2 = 0.99$), indicating a declining (but consistent) increase in participants' tendency to select different personas. This means more personas become a part of the set of chosen personas, but at a decreasing rate. **(b)** The ratio of female personas (i.e., the share of female personas in the generated persona set) and the share of participants choosing a female persona are almost perfectly correlated, which implies that the demographic attributes in the persona set strongly influence users' choice of a persona.

since there women are actually chosen more often (56.7% of the time) than men (43.3%). In fact, it appears that the gender ratio of personas shown to the user affects the gender ratio of their chosen personas (see Figure 4b). To test this effect, we conducted a chi-square test of independence for the likelihood of selecting a specific gender between the two persona sets. This association is statistically significant, such that increasing the number of personas increased the likelihood of selecting a female persona, X^2 (1, N = 74) = 11.26, p = .001.

Interestingly, more than half of the participants (59%, n = 22) chose a persona with a different gender in the second task. Out of those that changed, the vast majority (82%, n = 18) shifted from Male to Female persona. The remaining 18% (n = 4) changed from Female to Male persona. Thus, participants are 4.5 times more likely to first choose a Male persona and then Female persona than *vice versa*. These results imply that the consecutive use of different persona sets increases the diversity of the users' choice of personas.

Overall, the findings indicate an association between diversity – achieved by increasing the number of personas shown to users – and the users' persona choices. In terms of gender, the diversity of personas chosen mimics the diversity of personas shown, which is illustrated in Figure 4b.

6 RQ1 and RQ2: Hypothesis Testing

The hypotheses were tested using a logistic regression due to the binary outcome variable (apart from H5 that was tested using Spearman's correlation). Three regressions were conducted; one for the **PS1**, another for the **PS2**, and a third using the combined sets. We report the unstandardized coefficients (B), p-values, and odd ratios (OR) with the 95% confidence intervals (CI).

6.1 RQ1: What Characteristics of (a) the Persona and (b) the User Explain the Users' Persona Choice?

H1a: Users Are More Likely to Choose a Persona from Their Own Gender. There was no evidence of this effect in the **PS1**, $B = 0.123$, p = 0.806, OR = 1.131 [95% CI 0.422 – 3.024]; the **PS2**, $B = 0.144$, p = 0.806, OR = 1.131 [95% CI 0.423 – 3.024]; or the combined sets $B = 0.229$, p = 0.727, OR = 1.154 [95% CI 0.515 – 2.588]. Therefore, *H1a is not supported: there is no evidence that users are more likely to choose a persona from their own gender.*

H1b: Users Are More Likely to Choose a Persona with an Age Similar to Their Own Age. Age delta was found to have a significant effect for the **PS1**, B = -0.121, p ¡ 0.05, OR = 0.886 [95% CI 0.803 - 0.976]; and also in the combined sets, B = -0.067, p ¡ 0.05, OR = 0.934 [95% CI 0.889 - 0.982]. Each year of difference, in either direction, between the participant and the persona, reduces the likelihood of that persona being chosen. No evidence of this effect was found for the **PS2**, B = -0.025, p = 0.431, OR = 0.974 [95% CI 0.915 - 1.038]. Therefore, *H1b is partially supported: for the **PS1**, users were more likely to choose a persona with an age similar to their own age.*

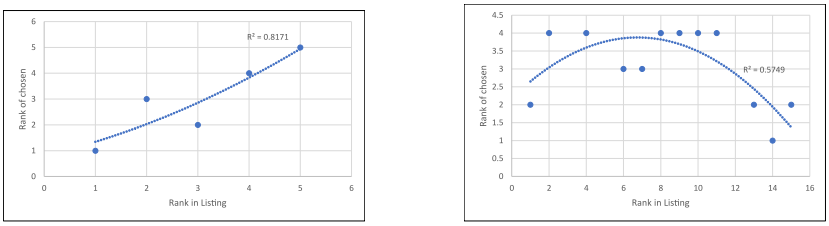
H2: Users Are More Likely to Choose Attractive Personas. No evidence of the 'what is beautiful is good' effect was found for the **PS1**, $B = 3.736$, p = 0.141, OR = 41.946 [95% CI 0.289 – 6067.582]; or the combined sets, B = 0.162, p = 0.547, OR = 1.175 [95% CI 0.693 – 1.992]. However, there is evidence of this occurring for the **PS2**, $B = 1.026$, p ¡ 0.05, OR = 2.789 [95% CI 1.083 – 7.180]. Therefore, *H2 is partially supported: for the **PS2**, users are more likely to choose attractive personas.*

H3: Users Are More Likely to Choose Personas that Are Different from Others. We found evidence of an outlier effect for the **PS1**, B = -1.718, p ¡ 0.01, OR = 0.179 [95% CI 0.053 – 0.607]; and the **PS2**, B = -2.261, p ¡ 0.05, OR = 0.104 [95% CI 0.016 – 0.660]. The effect was not statistically significant in the combined sets, despite the result being very close to significance, B = -0.638, p = 0.072, OR = 0.528 [95% CI 0.264 – 1.057]. Moreover, the detected effect *decreases* the persona's chance of being chosen, rather than increasing it (we will interpret this finding in the discussion). Therefore, *H3 is not supported: in both sets, personas that are different from others are less likely to be chosen.*

H4: Users Are More Likely to Choose Personas with a Higher Segment Representation. Higher audience representation was found to not influence the odds of a persona being chosen in the **PS1**, B = –0.707, p = 0.187, OR = 0.493 [95% CI 0.172 – 1.410]; or combined sets, $B = 0.017$, p = 0.697, OR = 1.017 [95% CI 0.846 - 1.222]. However, there was evidence of this occurring in the **PS2**, $B = 0.373$, p ¡ 0.05, OR = 1.451 [95% CI 1.050 – 2.005]. Therefore, *H4 is partially supported: in the PS2, users were more likely to choose personas with a higher segment representation.*

6.2 RQ2: What Interaction Aspects, Including (a) Persona Presentation and (b) Users' Interaction with Personas Explain Users' Persona Choice?

H5: Users Are More Likely to Choose Personas that Appear either (a) Higher or (b) Higher and Lower in the Order of Presentation. In the **PS1**, there is a very strong correlation between order and the probability of a persona being chosen (ρ = -0.900, p ¡ 0.05). Indeed, the persona which was first in order had a 41.93% probability of being chosen, while the second most likely persona to be chosen – with a 29.03% probability – was third in order. All others had even lower probabilities, the smallest of which was the last in order. This correlation was no longer present in the **PS2** (ρ = 0.174, p = 0.534), and likewise also absent in the combined sets (ρ = -0.192, p = 0.418). We also tested the quadratic and cubic fits for the **PS2**, but both fits were non-significant (Quadratic, p = 0.509; Cubic, p = 0.476). Therefore, *H5a is partially supported: for the **PS1**, personas appearing higher in the listing are more likely to be chosen.* However, *H5b is not supported: there is no evidence that personas appearing higher and lower in the listing are more likely to be chosen.* Figure 5 illustrates the order effects.



(a) Set 1 (b) Set 2

Fig. 5: Correlations between the ranking of persona being chosen (y axis) and ranking of the persona in system's listing (x axis). **(a)** shows close resemblance with the expected primacy effect. The inverse U-shaped curve in **(b)** is compatible with Ebbinghaus's U-shaped curve, where the first and last items in a list are best remembered. (Note that the curve is inverse here as lower position in y-axis implies higher probability of being chosen.) Yet, this effect is not statistically significant.

H6: Users Are More Likely to Choose Personas that (a) They Spend More Time Engaging with or (b) View More Frequently. For H6a, in the **PS1**, the longer a user spends on the persona's page, the greater the probability of that persona being chosen, B = 0.010, p ¡ 0.05; OR = 1.009 [95% CI 1.001 - 1.018]. For the **PS2**, no significant effects were detected for dwell time, B = 0.002; p = 0.067, OR = 1.002 [95% CI 0.999 - 1.004]. For the combined sets, dwell time is also significant, B = 0.003, p ¡ 0.05, OR = 1.003 [95% CI 1.000 - 1.005]. Therefore, *H6a is partially supported: for the **PS1** and the combined persona sets, users were more likely to choose personas they spend more time with.*

For H6b, no evidence was found in the **PS1**, B = -0.062, p = 0.588, OR = 0.939 [95% CI 0.750 - 1.177]; or the **PS2**, B = 0.087, p = 0.412, OR = 1.090 [95% CI 0.886 - 1.341]. However, in the combined sets, the number of times the participant visited the persona increased the likelihood of that persona being chosen. Each visit increases the odds of that persona being chosen by roughly 1.2 times, B = 0.164, p ¡ 0.05, OR = 1.178 [95% CI 1.039 - 1.335]. Therefore, *H6b is partially supported: for the combined persona set, users were more likely to choose personas they frequently view in the persona system.*

7 RQ3: What Strategies Do Users Apply for Their Persona Choice?

To explore the reasons given by the participants for their choice of a persona, we qualitatively coded their responses to the open-ended question: "Why did you choose this persona?". For this analysis, we combined the responses of the **PS1** and the **PS2** and analyzed them collectively. The analysis was done by categorizing each response according to the type of information or rationale the participant mentioned was a driving factor behind their choice.

7.1 Hypothesized Strategies

We defined a coding scheme based on the study hypotheses, and included five strategies for persona choice: (a) **similarity-seeking** (H1), (b) **looks or appearances** (H2), (c) **deviance-seeking** (H3), (d) **average-seeking** (H3), (e), and (f) **audience maximization** (H4). The coding scheme also included "other" which indicates a strategy that was not expressed in any of the other choices. In these cases, the coders were asked to write a note describing the strategy.

Two of the researchers coded the open answers independently for these strategies. The two researchers then discussed in order to reconcile their views on the disagreed instances (n = 36). Through this discussion, final labels were assigned to the disagreed instances, which meant that the two researchers now fully agreed on the coded instances. On two instances, it was agreed that both coders were correct, so both labels were assigned to the instances – *"He seemed distinctive in terms of age and I can identify more with him. He was also single."* (Participant B22) [similarity & deviance-seeking] and *"Because of the falcon (unique and interesting)" (B01) [looks or appearances & deviance-seeking].*

Table 1 shows the coding results. Out of the predefined strategies, the most common was similarity-seeking (n=11, 16.9%). Reasoning based on demographic similarity was made based on gender and age – e.g., *"Female and she is close to my age"* (A15) – but not country. While similarity was mentioned in terms of demographics seven times, it was also mentioned in terms of interests six times – e.g., *"I feel I share [with] him the interest in volunteering and giving to the community"* (A02). Two cases had both types of similarity – e.g., *"Most similar to me in age and interest"* (A04). This indicates that, when forming connections

Table 1: Seven strategies for persona selection. Sorted from most to least common. N = 68 instances.

Strategy	The user chooses the persona based on...	Frequency (%)
Fit-seeking	...the compatibility with the task; compares personas with an ideal profile.	29 (42.6%)
Similarity-seeking	...similarity in terms of interests, background, or demographics.	11 (16.2%)
Deviance-seeking	...something different that caught their attention.	9 (13.2%)
Audience maximization	...the audience representation.	8 (11.8%)
Looks or appearances	...appearances (i.e., the picture).	6 (8.8%)
Diversity-seeking	...increasing gender diversity in their sequential choices (e.g., first chooses a male, then a female persona).	3 (4.4%)
Average-seeking	...the persona being typical or representative of the persona set.	2 (2.9%)

with personas, participants are not only seeking for demographic similarity, but also for interest-based similarity.

All of the predefined strategies appeared in the open answers, lending support for the hypotheses. Strategies for similarity (esp. age), audience maximization and deviance-seeking support the hypotheses. However, even though the participants referred to looks or appearances , we did not find any case where the participant mentions that the persona looks good or attractive. Instead, the persona's looks were referred to as "authentic" (A10), "friendly" (B03), "smiling and happy" (B07), and "formal and [appropriate]" (B20). Thus, the visual information participants evoke deals more with aptitude and personality than attractiveness. This conclusion in supported by the fact that the most frequently selected persona among the **PS1** is Jafar (chosen 35% of the time, n=13), then Bassam (24%, n=9). If we consider only the 84% (n=31) of the participants that were able to name a persona, then the combined share of Jafar or Bassam goes up to 71% – 42% and 29%, respectively. Therefore, Jafar and Bassam were particularly selectable. Figure 1 shows that these personas appear as young, dynamic, and positive, which could contribute to their choice.

Furthermore, **average-seeking** refers to choosing the persona because it represents the group of personas – that is, is typical or representative of the personas being shown. Two participants (3.1%) expressed this strategy (e.g., *"Personal interests similar to the average, he feels positive in social media, his occupation is popular, the videos he watched were related to [the organization], his age is representative to the biggest group."* (B08)).

7.2 Strategies Emerging from the Qualitative Analysis

In addition to the predefined strategies, the coding led to the discovery of two new strategies. These strategies were identified inductively, i.e., through the analysis of the participants' responses. Conceptually, these did not match any of the expected strategies. Out of these, (6) **diversity-seeking** refers to choosing a persona with a different gender for the second task. Three participants (4.6% of the coded instances) made an explicit note of this (*"Trying to advertise [the organization's] content and [the organization] as a workplace to a more gender balanced audience"* (A01)). This strategy can be seen as an effect of the sequential nature of the study: the participants had to choose two personas, so some of

them considered that it would be justified to pick personas with different genders. This behavior was not especially encouraged by the researchers, but instead if reflects the general awareness of gender diversity among the participants.

Finally, (7) **fit-seeking** focused on the compatibility between the persona and the task. After reconciling the coding, fit emerged as the most prevalent category (n=28, 43.1% of total instances). The responses in the fit category indicated that several participants already had an ideal profile in their minds before interacting with the personas (*"I was looking for a middle age local female persona."* (A02); *"young, midway between western and Arabic culture, so both open and traditional as I see [the organization]"* (B04)) and were comparing personas against this ideal profile. The discussions among the two coders also clarified the conceptual boundaries of fit. Three aspects of fit were particularly striking: cultural, ideal target, and intersectional fit. *Cultural fit* implies that the participant focuses on the match between the organization's culture and the participant. *Ideal segment identification* implies that the participant has a predefined sense of the ideal persona for the task, defined typically by age, gender, or interests. *Intersectional fit* implies that the participant focuses on several features simultaneously, not highlighting one dominant reason for their choice – e.g., *"[I chose this persona] because of his age range, his positive sentiments, his interest in sports, startup and research which are key functional areas of [the organization] and his persona range of 299,300"* (B17).

Even though there may be some conceptual overlap between these subcategories, it became clear that what we termed as "fit" was a major driver in how the users approached the persona selection. For example, Nada appeared as a professionally promising person with the right interests for the task, as well as having cultural fit with the organization: *"As a young aspiring woman, she would make a perfect candidate as an employee or a higher education student."* (B12).

8 Discussion

8.1 Implications for Persona Theory

The findings revealed interesting insights about the use of personas. As mentioned in the beginning of this work, the reasons for applying personas in the design process is to help overcome the problem of self-referential design [30]. This generally requires that the persona description (a) can enable the persona user to have empathy for the persona, and (b) that the persona user has the ability to be empathetic. For the first point, taking a further look at the similarity seeking (*"I feel I share him the interest in volunteering and giving to the community"* (A02)) shows that the description enables the user to have empathy and perceive the persona profile as a realistic replacement of real users.

On the other hand, the fact that the persona's similarity to one-self was mentioned on several occasions in participants' voiced reasoning implies that self-referential information does play some role in persona use. This is also supported

by the significant findings regarding age match between the persona and the user, indicating that personas from the same age group are more likely to be chosen. As such, it appears that *persona users are not completely free from using self-referential information in their designs — users not only compare a persona to other personas, but also to their own attributes.* This finding highlights the role of social and human factors in persona use.

The appearances of the persona seem to influence the choice, but not necessarily in the terms of the persona being "beautiful" or attractive, even though evidence for this was found in the **PS1**. More commonly, the appearances of the persona seem to be intertwined with other factors, such as similarity – the persona looking like "me" or "my kind of person". Possibly, other traits in the pictures such as youth and dynamism may play a more important role, as these serve as information that relates to task. Overall, *physical attractiveness does not increase the persona's likelihood of being chosen for the task.*

The fact that outlier personas are *less* likely to be chosen suggests that going for average personas is typical behavior. As most users seem to choose majority personas, *marginalized groups may require extra support (e.g., visual saliency, higher rank of presentation) to increase their chances of being chosen for design tasks by users.* A corollary is that increasing the demographic diversity of the persona set shown to users seems to increase the diversity of the personas chosen by the users, which could be seen beneficial for inclusive design [13].

The persona's audience size also seems to matter, in that some users justify their decisions based on it. This may result from the nature of the task: a YouTube content creation task implies the goal of reaching a large audience. Other task types should be tested to understand persona choice in a variety of settings. Therefore, the observed popularity effect might in fact represent another facet of the fit theory, i.e., the idea that the persona is predominantly chosen based on its compatibility for the task, rather than its external qualities, such as demographics or attractiveness.

The primacy effect was only significant for the **PS1**, where there was a strong linear relationship. Thus, *the order of presentation and the layout, in this case the persona listing, seems to have an impact especially on the smaller number of personas, where the faster overview makes the persona user pick the top persona instead of one that serves the task best.*

In terms of users' engagement with the persona profile increasing the persona's chance of being selected, the results indicate that engagement with a specific persona profile does not predict that the persona will be chosen, but, out of the tested interaction metrics, *dwell time is a stronger predictor for persona choice than the number of times the persona was visited.*

8.2 Design Implications

The design implications from this work are as follows.

First, interactive persona systems afford flexibility for users to find personas that match a wide range of strategies for a given task. As users can choose and browse personas that fit with their choice criteria for completing tasks, they can

ultimately find the personas deemed appropriate for the task at hand. Therefore, *persona developers should consider presenting the personas via digital media, such as Web browsers, using functionality that enables browsing, searching, and filtering based on persona attributes (e.g., demographics, audience size).*

Second, the fact that the increase in the proportions of the shown and chosen female personas are almost perfectly correlated implies that users' choices are influenced by gender proportions in the persona set. If persona developers show more female personas, more female personas are chosen for the design task. Therefore, *when the overall goal is to increase the persona user's knowledge of the demographic diversity of the end-user base, it is advisable to show personas in balanced demographic proportions.*

Third, the fact that personas' order can affect how they are used and chosen implies that persona systems should give tools for persona developers to change the order of personas when this is considered important. For example, minority groups could be prioritized to increase their visual saliency in the UI. *The choice of directing users' attention towards default personas versus letting them discover personas freely remains an important design choice for interactive persona systems.*

8.3 Limitations and Future Work Avenues

The results imply that either the diversity or the number of personas affects users' persona choice, but our analysis does not disentangle these two explanations. Some users explicitly mentioned the number of personas when using the system (e.g., "(...) if you have like fewer number of personas, you will have maybe more diverse kind of group and segments, and as you increase the number of personas, you know, if your [data] has like more fine-grained groups, that might do the same [...] just because you're increasing, then you're creating things that are kind of similar." (B19)). Even though we leave the detailed analysis for future work, our exploratory findings imply that, as diversity and the number of personas are strongly correlated, persona creators can increase the number and diversity of personas in order to target more inclusive design outcomes, as designers choosing more diverse personas would logically result in consideration for the needs of more diverse groups of people (all else being equal).

Using a data-driven persona creation methodology can be considered both as a strength and as a weakness. It is a strength because by using authentic personas that we researchers did not manipulate gives the users an authentic purpose as opposed to us purposefully making choices regarding the personas' gender, age, or any other attribute. On the other hand, by not making such choices, we lose the control over these factors; so the results are exploratory rather than definitive. We opted for the data-driven approach because we wanted to keep the personas realistic for the participants, so that personas actually represent the YouTube viewers. Not using data-driven personas and claiming that the personas were based on real data would have been deceptive.

In terms of research design, the study design was not optimal for testing the order effects. Ideally, we would need to present the order randomly to a number of

different participants and see if the probabilities of the personas remain similar regardless of position. These limitations originate, on the one hand, from the limited sample size (it is difficult to recruit enough participants to satisfy sample size requirements for complex experimental designs) and, on the other hand, from the use of the data-driven persona system that by default shows the personas in a specific order for all users. Future studies should more rigorously test the order effects associated with persona presentation.

Our coding of persona information does not necessarily correspond with the participants' understanding of the information. For example, we did not encode the persona with a falcon in the picture as an outlier (see Figure 1d), but this was nonetheless observed as an outlier by one of the participants because falcons have a particular meaning in the persona's country. Because such tacit cultural nuances can be easily overlooked by persona developers when choosing persona information, validation of the personas for cultural sensibilities would be advisable, although it is likely impossible to perfectly control the subjective interpretation of personas by the participants.

A particularly interesting aspect is the role of the algorithm – when generating more personas, the persona set becomes more varied. This can be interpreted by the mechanism by which increasing the number of personas allows for more "demographic slots" to be filled. This interpretation encourages future studies with persona sets containing more personas. The conventional rationale of keeping persona sets small, typically within the range of 3-7 [6], is that a small number of personas is cognitively more manageable (e.g., considering printing out 100 persona profiles in paper sheets – presenting them to end users in an efficient way would be difficult if not impossible). However, given the search and filtering functions of an interactive persona system, the cognitive burden can be decreased, thus enabling a host of future studies with a higher number (and a higher degree of demographic diversity) of personas.

9 Conclusion

Exploratory support explaining business professionals' persona choice was found for age similarity, persona appearances, number of people the persona represents, dwell time with the persona, and whether or not the persona is an outlier. The qualitative analysis revealed that participants applied seven different strategies for persona selection: similarity-seeking, looks or appearances, deviance-seeking, diversity-seeking, average-seeking, fit-seeking, and audience maximization. The process of choosing personas appears to be influenced by the diversity of the personas available to the user. Persona choice and related interaction between stakeholders and personas requires more empirical research.

References

1. An, J., Kwak, H., Jung, S.g., Salminen, J., Jansen, B.J.: Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. Social Network Analysis and Mining **8**(1)

(2018). https://doi.org/10.1007/s13278-018-0531-0, https://www.readcube.com/articles/10.1007/s13278-018-0531-0
2. An, J., Kwak, H., Salminen, J., Jung, S.g., Jansen, B.J.: Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data. ACM Transactions on the Web (TWEB) **12**(3) (2018)
3. Anvari, F., Richards, D., Hitchens, M., Babar, M.A., Tran, H.M.T., Busch, P.: An empirical investigation of the influence of persona with personality traits on conceptual design. Journal of Systems and Software **134**, 324–339 (Dec 2017). https://doi.org/10.1016/j.jss.2017.09.020, http://www.sciencedirect.com/science/article/pii/S0164121217302078
4. Atzeni, A., Cameroni, C., Faily, S., Lyle, J., Flechais, I.: Here's Johnny: A Methodology for Developing Attacker Personas. In: 2011 Sixth International Conference on Availability, Reliability and Security. pp. 722–727. Vienna, Austria (Aug 2011). https://doi.org/10.1109/ARES.2011.115
5. Berger, C.R., Calabrese, R.J.: Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. Human communication research **1**(2), 99–112 (1974), publisher: Oxford University Press
6. Blomquist, A., Arvola, M.: Personas in action: ethnography in an interaction design team. In: Proceedings of the second Nordic conference on Human-computer interaction. pp. 197–200. ACM, Aarhus, Denmark (2002), http://dl.acm.org/citation.cfm?id=572044
7. Cooper, A.: The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity (2nd Edition). Pearson Higher Education (2004)
8. Dion, K., Berscheid, E., Walster, E.: What is beautiful is good. Journal of personality and social psychology **24**(3), 285 (1972), publisher: American Psychological Association
9. Duda, S.: Personas—Who Owns Them. In: von Gizycki, V., Elias, C.A. (eds.) Omnichannel Branding: Digitalisierung als Basis erlebnis- und beziehungsorientierter Markenführung, pp. 173–191. Springer Fachmedien Wiesbaden, Wiesbaden (2018). https://doi.org/10.1007/978-3-658-21450-0_8, https://doi.org/10.1007/978-3-658-21450-0_8
10. Ebbinghaus, H.: Memory: A contribution to experimental psychology. Annals of neurosciences **20**(4), 155 (2013), publisher: Karger Publishers
11. Faily, S., Flechais, I.: Persona cases: a technique for grounding personas. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 2267–2270 (2011)
12. Faily, S., Lyle, J.: Guidelines for integrating personas into software engineering tools. In: Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems - EICS '13. p. 69. ACM Press, London, United Kingdom (2013). https://doi.org/10.1145/2494603.2480318, http://dl.acm.org/citation.cfm?doid=2494603.2480318
13. Goodman-Deane, J., Waller, S., Demin, D., González-de Heredia, A., Bradley, M., Clarkson, J.P.: Evaluating Inclusivity using Quantitative Personas (Jun 2018). https://doi.org/10.21606/drs.2018.400, https://www.dropbox.com/sh/cbj04ez206dexrd/AADKHi9uwwDXvpwndnSWEAk1a?dl=0&preview=Goodman-Deane+Waller+Demin+Gonz%C3%A1lez-de-Heredia+Bradley+Clarkson+400.pdf
14. Grudin, J.: Why Personas Work: The Psychological Evidence. In: Pruitt, J., Adlin, T. (eds.) The Persona Lifecycle, pp. 642–663. Elsevier (2006). https://doi.org/10.1016/B978-012566251-2/50013-7, https://linkinghub.elsevier.com/retrieve/pii/B9780125662512500137

15. Hamilton, D.L., Fallot, R.D.: Information salience as a weighting factor in impression formation. Journal of Personality and Social Psychology **30**(4), 444–448 (1974). https://doi.org/10.1037/h0037033, place: US Publisher: American Psychological Association
16. Holmgard, C., Green, M.C., Liapis, A., Togelius, J.: Automated Playtesting with Procedural Personas with Evolved Heuristics. IEEE Transactions on Games **PP**(99), 1–1 (2018). https://doi.org/10.1109/TG.2018.2808198
17. Jansen, B.J., Jung, S.g., Salminen, J.: Creating Manageable Persona Sets from Large User Populations. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–6. ACM, Glasgow, United Kingdom (2019). https://doi.org/10.1145/3290607.3313006
18. Jansen, B.J., Salminen, J.O., Jung, S.G.: Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics. Data and Information Management **4**(1), 1–17 (Mar 2020). https://doi.org/10.2478/dim-2020-0005, `https://content.sciendo.com/view/journals/dim/4/1/article-p1.xml`, publisher: Sciendo Section: Data and Information Management
19. Jung, S.g., Salminen, J., Kwak, H., An, J., Jansen, B.J.: Automatic Persona Generation (APG): A Rationale and Demonstration. In: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval. pp. 321–324. ACM, New Brunswick, NJ, USA (2018). https://doi.org/10.1145/3176349.3176893
20. Kandel, D.B.: Homophily, selection, and socialization in adolescent friendships. American journal of Sociology **84**(2), 427–436 (1978), publisher: University of Chicago Press
21. Kotler, P., Levy, S.J.: Broadening the concept of marketing. Journal of marketing **33**(1), 10–15 (1969), publisher: SAGE Publications Sage CA: Los Angeles, CA
22. LaFontana, K.M., Cillessen, A.H.: Children's perceptions of popular and unpopular peers: A multimethod assessment. Developmental psychology **38**(5), 635 (2002), publisher: American Psychological Association
23. Lavie, N., Hirst, A., De Fockert, J.W., Viding, E.: Load theory of selective attention and cognitive control. Journal of experimental psychology: General **133**(3), 339 (2004), publisher: American Psychological Association
24. Lee, D.D., Seung, S.H.: Learning the Parts of Objects by Non-negative Matrix Factorization. Nature **401**(6755), 788–791 (1999)
25. LeRouge, C., Ma, J., Sneha, S., Tolle, K.: User profiles and personas in the design and development of consumer health technologies. International journal of medical informatics **82**(11), e251–e268 (2013)
26. Long, F.: Real or imaginary: The effectiveness of using personas in product design. In: Proceedings of the Irish Ergonomics Society Annual Conference. vol. 14. Irish Ergonomics Society Dublin (2009)
27. Lydon, J.E., Jamieson, D.W., Zanna, M.P.: Interpersonal similarity and the social and intellectual dimensions of first impressions. Social cognition **6**(4), 269–286 (1988), publisher: Guilford Press
28. Murase, Y., Jo, H.H., Török, J., Kertész, J., Kaski, K.: Structural transition in social networks: The role of homophily. Scientific Reports **9**(1), 4310 (Mar 2019). https://doi.org/10.1038/s41598-019-40990-z, `https://www.nature.com/articles/s41598-019-40990-z`, number: 1 Publisher: Nature Publishing Group
29. Murdock Jr., B.B.: The serial position effect of free recall. Journal of Experimental Psychology **64**(5), 482–488 (1962). https://doi.org/10.1037/h0045106, place: US Publisher: American Psychological Association

30. Nielsen, L.: Personas - User Focused Design. Springer, New York, NY, USA, 2nd ed. 2019 edition edn. (Feb 2019)
31. Nielsen, L., Nielsen, K.S., Stage, J., Billestrup, J.: Going Global with Personas. In: Proceedings of the INTERACT 2013 conference. pp. 350–357. Springer, Berlin, Heidelberg, Cape Town, South Africa (2013). https://doi.org/10.1007/978-3-642-40498-6$_2$7, https://link.springer.com/chapter/10.1007/978-3-642-40498-6_27
32. Parkhurst, J.T., Hopmeyer, A.: Sociometric Popularity and Peer-Perceived Popularity: Two Distinct Dimensions of Peer Status. The Journal of Early Adolescence **18**(2), 125–144 (May 1998). https://doi.org/10.1177/0272431698018002001, https://doi.org/10.1177/0272431698018002001, publisher: SAGE Publications Inc
33. Raffel, G.: Two Determinants of the Effect of Primacy. The American Journal of Psychology **48**(4), 654–657 (1936). https://doi.org/10.2307/1416518, https://www.jstor.org/stable/1416518, publisher: University of Illinois Press
34. Salminen, J., Guan, K., Jung, S.G., Chowdhury, S.A., Jansen, B.J.: A Literature Review of Quantitative Persona Creation. In: Proceedings of the ACM Conference of Human Factors in Computing Systems (CHI'20). ACM, Honolulu, Hawaii, USA (Apr 2020)
35. Salminen, J., Jung, S.g., Chowdhury, S.A., Sengün, S., Jansen, B.J.: 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 (Apr 2020). https://doi.org/https://doi.org/10.1145/3313831.3376770
36. Salminen, J., Jung, S.G., Santos, J.M., Jansen, B.J.: Does a Smile Matter if the Person Is Not Real?: The Effect of a Smile and Stock Photos on Persona Perceptions. International Journal of Human–Computer Interaction **0**(0), 1–23 (Sep 2019). https://doi.org/10.1080/10447318.2019.1664068, https://doi.org/10.1080/10447318.2019.1664068
37. Salminen, J., Santos, J.M., Jung, S.G., Eslami, M., Jansen, B.J.: Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas. International Journal of Human–Computer Interaction **0**(0), 1–13 (Nov 2019). https://doi.org/10.1080/10447318.2019.1688946, https://doi.org/10.1080/10447318.2019.1688946
38. Singh, V.: Personas and Scenarios as a Methodology for Information Sciences. Qualitative and Quantitative Methods in Libraries **7**(1), 123–134 (Feb 2019), http://qqml-journal.net/index.php/qqml/article/view/462, number: 1
39. Stewart, R.H.: Effect of continuous responding on the order effect in personality impression formation. Journal of Personality and Social Psychology **1**(2), 161 (1965), publisher: American Psychological Association
40. Westphal, J.D., Zajac, E.J.: Who shall govern? CEO/board power, demographic similarity, and new director selection. Administrative science quarterly pp. 60–83 (1995), publisher: JSTOR
41. Zajonc, R.B.: Attitudinal effects of mere exposure. Journal of personality and social psychology **9**(2p2), 1 (1968), publisher: American Psychological Association