

# ARE PERSONAS DONE? EVALUATING THE USEFULNESS OF PERSONAS IN THE AGE OF ONLINE ANALYTICS

JONI SALMINEN, BERNARD J. JANSEN, JISUN AN,  
HAEWOON KWAK AND SOON-GYO JUNG

## ABSTRACT

*In this research, we conceptually examine the use of personas in an age of large-scale online analytics data. Based on the criticism and benefits outlined in prior work and by practitioners working with online data, we formulate the major arguments for and against the use of personas given real-time online analytics data about customers, analyze these arguments, and demonstrate areas for the productive employment of data-driven personas by leveraging online analytics data in their creation. Our key tenet is that data-driven personas are located between aggregated and individual customer statistics. At their best, digital data-driven personas capture the coverage of the customer base attributed to aggregated data representations while retaining the interpretability of individual-level analytics; they benefit from powerful computational techniques and novel data sources. We discuss how digital data-driven personas can draw from technological advancements to remedy the notable concerns voiced by scholars and practitioners, including persona validation, inconsistency problem, and long development times. Finally, we outline areas of future research of personas in the context of online analytics. We argue that to survive in the rapidly developing online customer analytics industry, personas must evolve by adopting new practices.*

## KEY WORDS

Data-Driven Personas; Online Analytics; Customer Segmentation

## INTRODUCTION

The abundance of social media data has increased the difficulty of making sense of data (Järvinen 2016; Saggi & Jain 2018; Salminen, Milenković & Jansen 2017), while at the same time making it possible to automatically infer customers attributes from social media that were previously accessible only by survey research. Researchers have begun utilizing publicly available social media posts to infer customer attributes, including personality traits, political orientation, brand liking, and needs and wants (Ardehaly & Culotta 2015; Del Vecchio, Mele, Ndou, & Secundo 2017; Jung, An, Kwak, Salminen, & Jansen 2017; Volkova, Bachrach, & Durme 2016). Free expression in social media provides opportunities to learn about the needs and traits of groups and individuals (Owusu et al. 2016). Overall, the development of computational techniques and the availability of online data has resulted in an increased interest in data-driven personas: (a) to describe the content consumption patterns of diverse online audiences

(Salminen, Şengün, et al. 2018), and (b) to use online data, namely online analytics and social media posts, in persona generation (An, Kwak & Jansen 2017; Zhang, Brown & Shankar 2016).

However, at the same time, the usefulness of personas has been questioned. There is a plethora of alternative online analytics tools (e.g., Google Analytics, SimilarWeb), services (e.g., comScore, HitWise) and metrics (Boghrati et al. 2017; Clarke & Jansen 2017; Järvinen & Karjaluo 2015) that one can employ to understand customers. Moreover, companies have gained access to individual-level data that performs well for many marketing purposes, including customer relationship management (Zerbino et al. 2018), providing tailored experiences and recommendations (Ronen, Yom-Tov & Lavee 2016), one-to-one targeting of online ads (Miralles-Pechuán, Ponce & Martínez-Villaseñor 2018), and enabling the creation of sophisticated customer segments (Jansen et al. 2017; Rundle-Thiele, Dietrich & Kubacki 2017). Although prevalent in many fields, the use of automated techniques is especially pertinent in the field of digital marketing which is shifting toward a higher degree of personalization, micro-targeting, and one-to-one marketing (Bleier & Eisenbeiss 2015). Thus, there are concerns about personas providing real value in such an environment.

In this research, we evaluate the use and usefulness of personas given the easy availability of online individual customer data for digital marketing use cases. Traditionally, personas are used in replacement of actual one-to-one data about customers (Howard 2015), as mental models that help keeping customers in mind in the absence of having real customers available when making decisions about design (Nielsen 2013), software development (Pruitt & Grudin 2003), and marketing (Russell & Toklu 2011). However, with the widespread availability of online analytics data and numerous programs, techniques, and platforms to process the gathered data, a research question arises: *Are personas still valid as a marketing tool in the era of online analytics?*

We address this question through a conceptual inquiry and analysis, drawing from two streams of literature: (a) persona studies (for drawing the benefits and shortcomings of personas) and (b) digital marketing research (for drawing the use of digital marketing). We aim to conceptually combine these streams to better understand the role of personas in the era of online analytics and digital marketing. Therefore, this work is *conceptual research discussing the value of personas in the era of online analytics*. Our purpose is to review the related literature for better positioning personas given the context of online analytics, which is currently lacking in the extant persona literature. With our inquiry, we aim to avoid the advocacy issue mentioned by Matthews, Judge & Whittaker (2012, p. 1220): “[persona] literature [...] generally takes a position of advocacy and lacks objectivity.” That is, we explore both the possible strengths and the shortcomings of personas in an objective manner.

First, we identify the benefits to which personas are associated in the literature. Second, we summarize the traditional criticisms of personas. Third, we formulate new critical arguments against the use of personas given the widespread availability of online analytics data. We then address these arguments, examining their relevance to traditional and digital data-driven personas, and discuss the ability of technology to solve the major shortcomings of personas. Finally, we conclude by presenting ideas for the future of personas in the era of online analytics.

Our key contributions are twofold. First, we present novel criticism relating to the use of personas for customer-related decision making, addressing concerns from both digital analytics practitioners and academic scholars. Second, we reposition personas in the age of online analytics by conceptually distinguishing between traditional data-driven personas and digital data-driven personas, and analyzing the implications of both for solving persona challenges. We,

therefore, establish and discuss the role of personas in the context of online analytics, which has not previously been done in the persona literature.

## **METHODOLOGICAL APPROACH**

Being a conceptual and survey work, this manuscript relies on prior literature as a source of evidence. To this end, we perform a literature review to identify the key benefits and criticisms of personas in the academic literature. To find the works of this domain, we conducted searches on Google Scholar and Science Direct with the base keyword “personas” and then scanned the abstracts of the articles for cue words indicating weaknesses, downsides, or critiques of personas. In this literature query, we focused on finding critical papers, as most persona studies list the benefits but more rarely include criticism (Matthews, Judge & Whittaker 2012). The literature on persona benefits is, therefore, easier to find. In total, we identify 32 relevant papers, many of which are conference articles in the field of computer science (sub-fields e.g. human-computer interaction, design, and software development). We read these articles in detail and use them to discover related articles discussing the strengths and weaknesses of personas (a technique referred to as snowball sampling (Provan & Milward 1995).

## **WHY ARE PERSONAS USEFUL?**

Cooper (1999) introduced personas as a design technique for understanding and communicating the goals and needs of different user types. Since then, personas have been employed by designers, software developers, and marketers, among other decision-maker groups (Nielsen 2013; Pruitt & Adlin 2006; Mulder & Yaar 2006). Personas crystallize a specific user type, often focusing on core users in the absence of an immediate contact to the end user (Floyd, Jones & Twidale 2008). A user can refer to a user of a software system, such as a website or mobile application, a player of a game, customer of a product or service, audience of online content, target segment for marketing campaigns, a patient of public health services, and so on (Pruitt & Grudin 2003; LeRouge et al. 2013; Ma & LeRouge 2007; Scott 2007; Dong, Kelkar & Braun 2007; Nacke, Drachen & Göbel 2010; Vahlo & Koponen 2018). In the remainder of this work, we use the concept of ‘customers’ when referring to these groups. In this research, we examine personas specifically in the context of online analytics, adopting the perspective that personas are created to provide real value for their end users (Cooper 1999; Gudjónsdóttir 2010), such as achieving more user-friendly designs or more empathetic advertising texts. Evaluating personas, therefore, has to take place in relation to their *value in use* (Kartemo, Akaka & Vargo 2017).

The benefits of personas according to the literature are summarized in Table 1. The communicational benefits arise from summarizing customer information into an intuitive format of representation that can be communicated with little effort (Holtzblatt, Wendell & Wood 2005) within organizations, teams, departments, and with external stakeholders (Matthews, Judge & Whittaker 2012). In theory, personas provide an engaging description of the end users’ needs and wants, in the form of another human being that is more memorable than numbers (Goodwin 2009; Hill et al. 2017). At their best, personas become shared mental models that individuals rely upon when making decisions (Nielsen 2013) that concern the specific user type (Cooper 1999). This enables the decision makers to discuss experiences and backgrounds different from their own and realize that the customer preferences may deviate from their own preferences (Miaskiewicz & Kozar 2011).

Table 1: Benefits associated with the use of personas

Category	Description
<b>Communication</b>	Personas facilitate user-oriented communication within and between teams in the organization.
<b>Psychology</b>	Personas enhance the immersion required for designing ‘for a person’ instead of fuzzy and complex target groups.
<b>Transformation</b>	Personas challenge existing assumptions about customers and orientate trade-off decisions when customers have conflicting needs.
<b>Focus</b>	Personas help focus design decisions on user goals and needs rather than on system attributes and features.

The psychological benefits are rooted in identification with the personas (Miaskiewicz, Sumner & Kozar 2008), whereby decision makers can obtain an empathic understanding of users, immersing themselves in real situations experienced by others. Decision makers can use this ability to predict customer behavior under different circumstances (Pruitt & Grudin 2003). This mental modeling relies on human beings’ innate ability of empathy and immersion (Krashen 1984), and is, therefore, a powerful agent for enhanced motivation and purpose. At best, personas can give a higher sense of meaning to one’s work. Consider the psychological difference between creating a software product to the nameless target group of *24–35 year-old women*, compared to creating a mobile application for *Jane, a stressed single mum who wants to better manage her time*.

Transformational benefits relate to challenging the established perceptions about the users within the organization (Miaskiewicz & Kozar 2011). Because the creation of personas is based on gathering real evidence of the users (Pruitt & Grudin 2003), the results can deviate from the existing preconceptions and truisms within the organization. When there is friction between the perceived and real goals of the users, accurate persona representations can reduce this perceptual gap by conveying factual information about users’ needs and wants (Pruitt & Adlin 2006). If the organization and the decision makers are open to re-aligning their perceptions, the persona exercise can prevent and rectify false conceptions of the end users (Matthews, Judge & Whittaker 2012).

Finally, personas can facilitate focusing on the most important audiences (Miaskiewicz & Kozar 2011). This helps decision makers with strategic agenda to prioritize certain customers over others, and thus resolve conflicting needs and wants among the customer base. For example, if Persona A wants *Feature set x*, while Persona B wants *Feature set y*; by considering the overall strategy of which users the organization wants to serve (a practice known as customer portfolio management) (Johnson & Selnes 2004), we can define the optimal product features to focus on (Cooper 1999; Ma & LeRouge 2007). Thus, personas help to prioritize product requirements and help determine if the right problems are being solved while curbing the self-centering bias of the decision makers (Matthews, Judge & Whittaker 2012).

## CRITICISM OF PERSONAS

The literature includes a lot of substantial criticism of personas. To better understand and dissect this criticism, we have categorized it into three sections that are roughly compatible with the typical lifecycle of a persona project: first personas are created (*creation of personas*), then assessed by decision makers (*evaluation of personas*), and finally applied in real scenarios and use cases (*use of personas*). Various critical arguments arise in the literature at each of these stages. These arguments are summarized in Table 2 and discussed afterward.

Table 2: Established criticism of personas from prior literature.

Category	Key issues	Authors
Creation of personas	a) persona creation takes a long time b) personas are expensive to create c) personas can be biased by their creators d) personas are based on non-representative data	Pruitt and Grudin (2003); Vincent and Blandford (2014); Hill et al. (2017)
Evaluation of personas	e) personas lack credibility f) personas are not accurate or verifiable g) the information in personas is not relevant for decision makers h) personas are inconsistent	Chapman and Milham (2006); Bødker et al. (2012); Matthews et al. (2012)
Use of personas	i) not using the created personas j) using personas for politics and power play k) using personas to justify preconceptions l) personas change in time	Rönkkö et al. (2004); Rönkkö (2005); Chapman and Milham (2006)

Hill et al. (2017) point out that (a) *creating quality personas takes considerable time and effort*. According to Vincent and Blandford (2014), persona creation can take months. In a similar vein, Pruitt and Grudin (2003) advocate an in-depth research effort for persona creation, typically lasting months. It is seen that for personas to be accurate, considerable investigative work is required. This tends to result in (b) *persona projects being expensive*, in the range of tens of thousands of American dollars (Marsden & Haag 2016; Miaskiewicz, Sumner & Kozar 2008). Consequently, as Rönkkö (2005) found, the amount of effort may lead to questioning the return on investment of persona projects. Moreover, the high cost of persona creation tends to exclude them from the reach of small businesses and startups, as pointed out by Salminen, Jansen, An, Kwak, and Jung (2018).

There are also concerns about the validity of personas, or how well the personas match the reality. A commonly mentioned validity concern is that (d) *personas are based on insufficient or non-representative data* (Chapman & Milham 2006). Personas that are built based on relatively few qualitative interviews may not represent the underlying user groups in a statistically valid manner. For example, the resulting personas may suffer from bias when the interviewed subjects are chosen based on availability rather than representativeness of the entire customer base. Overall, (c) *personas risk inheriting organizational tensions and individual biases*, including political and strategic ambitions of their creators (Hill et al. 2017; Massanari 2010; Rönkkö 2005). Vincent and Blandford (2014, p. 1098) argue that “[persona creation] has

been adapted, depending on what people want to accomplish and why.” Since persona representations are not easily verifiable, a deliberate selection bias can take place. This bias can also be involuntary so that the creators of the personas are projecting their own prejudices unconsciously in the persona description.

Additionally, since persona creation work is typically qualitative, (e) *personas lack the credibility of numbers* and are, to some, interpretative and subjective instead of rigorous and believable (Chapman & Milham 2006). Even when using the best practices of qualitative inquiry, number-oriented decision makers may consider personas as ‘nice narratives’ instead of serious decision-making instruments, resulting in resistance for the use of personas (Massanari 2010). For example, the participants in the study of Matthews et al. (2012) found personas misleading, abstract, and unrealistic. In a similar vein, Bødker et al. (2012, p. 93) report that “as soon as the project started [...] it became a concern that the 12 personas seemed distant from actual citizens, very general and difficult to activate.” There can be many reasons for such interpretations. For example, the subjective experiences and impressions of decision makers may conflict with personas (Marsden & Haag 2016). Furthermore, other information about users, such as direct customer feedback, can conflict with persona information, because the full complexity and range of the customer base deviates from the idealized personas (Chapman & Milham 2006). In such cases, decision-makers need to consider the credibility of personas against other sources of data. This may result in a willingness to hold on to one’s existing beliefs (Delfabbro 2004), trusting one’s own observations and insights instead of more abstract personas.

Moreover, there is no objectively right or wrong answer on which information to include in the persona profile (Bødker et al. 2012; Chapman & Milham 2006). Some decision-makers, for example, might prefer data relating to customer journey while others are more interested in psychographics. Therefore, (g) *information selection for personas is arbitrary* and may not be of use in a given scenario or use case. The information should be based on the information needs of the end users of personas (Sinha 2003). These needs vary across industries and use cases, even between job roles within the same organization. For example, “marketing personas” would include information such as consumption patterns and consumer motivations, goals, and likes and dislikes (Thoma & Williams 2009), whereas online content producers would prefer information on content consumption patterns (Nielsen et al. 2017).

Additionally, (h) *personas are said to be inconsistent*, meaning that they are created by combining information from several unrelated data sources, without ensuring that the individual pieces of information are commensurable (Matthews, Judge & Whittaker 2012). Bødker et al. (2012) refer to personas as “Frankenstein’s monsters”, postulating that they can be patched up from any information available. Due to the above reasons, (f) *the accuracy of personas is difficult to validate*. If decision makers in an organization are aware of this risk, they will not trust the persona representations and will downplay their use in real decision-making situations (Chapman & Milham 2006). The lack of trust is aggravated when decision makers do not personally participate in the persona creation (Long 2009), and, thus, the lack the psychological ownership of the persona artifacts (Bødker et al. 2012). Overall, these adverse dynamics assert strain on the credibility of personas. If the attitude of the decision makers is not favorable to the use of personas, the potential benefits remain unachieved.

Finally, we identified five threats for the actual use of personas. First, there are situations where (i) *personas are fully developed but then left without meaningful use*. For example, Rönkkö et al. (2004) report a case where a considerable amount of time was used to develop personas that were never implemented. In a similar vein, Matthews, Judge and Whittaker (2012) found that personas had little impact on the actual design work. Friess (2012)

conducted an ethnographic study among designers and found a serious imbalance between the creation and use of personas in real decision-making situations. Chapman and Milham (2006) see the lack of validation resulting in political conflicts where (j) *choosing a persona is a question of opinion instead of a fact*. For example, the created personas may be rejected in favor of the manager's pre-existing beliefs about the customers (Chapman & Milham 2006; Pruitt & Grudin 2003), to frame the strategic and operational discussions in the organization (Rönkkö 2005), or (k) *interpreted to confirm existing beliefs* rather than to seek new explanations (Salminen, Jung, et al. 2018).

Rönkkö (2005) also found that personas were used as a form of after-the-fact justification, so design choices based on other inputs were later communicated to other team members as if they were based on the personas. Rönkkö et al. (2004, p. 115) describe a case where the persona evolved in time and was eventually seen differently by stakeholders of the company: "The [...] persona who originated as a middle-age businessman ended up as a less clearly definable figure, e.g. a younger careerist of both the male and the female sex, and a diversity of different professions whose common characteristics was mobility, e.g. salesman, plumber, nurse, policeman, veterinary." This practice in effect nulls the acclaimed benefit of aligning user understandings within the organization. Finally, (l) *use of personas is hindered by changing customer behavior*; as the customer behavior changes, personas should be updated to reflect these changes (Jung et al. 2017). However, being that data collection is typically expensive, the updating may not be possible, and personas risk expiring rapidly in real use.

Overall, the above challenges risk creating situations where the persona benefits remain largely theoretical and do not materialize in real use cases (Friess 2012; Marsden & Haag 2016). In addition, there are novel concerns arising from the use of personas in comparison to utilizing online analytics data to understand customers.

## NEW CRITICISM

We now move toward introducing the context of online analytics. In addition to the established criticism laid out in the previous section, there have been newly-found criticisms of personas, arising from the availability of online analytics data, metrics, and techniques for business purposes. To demonstrate the logic of this new criticism, we present the following quotations retrieved from recent online writings by practitioners dealing with online customer data:

*"Whereas personas were once a good starting point to identify 'buckets' of customers, the limitations of persona-based marketing have become apparent as the consumer decision-making journey veered from its predictable linear path and increased in complexity."*<sup>i</sup>

*"Personas tend to be exhaustive where it's not needed (demographics, names, pictures are not necessary most of the times), while they fail to summarise the complex variety of needs and usage scenarios that real users express in real life situations."*<sup>ii</sup>

*"The idea of a persona or an average customer was the typical way that marketers would think about their customer base. But now with advancements in technology, with modeling, with more available skill sets, they are able to understand and predict future behavior at more granular levels, and it's a dramatic shift that's happening."*<sup>iii</sup>

*"You're still clinging to generic user personas in the age of Big Data? LOL."*<sup>iv</sup>

The above critical statements are chosen to illustrate the anti-persona sentiment taking place among some of the data-oriented business professionals. It is not a comprehensive sample, nor do we argue that all business professionals proficient with online analytics data and tools would perceive personas as useless. However, there are several recent blog posts that dispute the usefulness of personas and therefore it is worthwhile to bring this criticism to scholarly attention and analyze it objectively in the proper frame of context. The crux of the criticism can be summed up in three categories:

**increased complexity** (Quotes 1 and 4) – *personas are not able to capture the diversity and nuances of the increasingly large online audiences. Typically, one would create only a few personas (less than 10) to describe the core users which may not be enough in the era of online analytics and fragmented consumer behavior.*

**redundant information** (Quote 2) – *personas are overly focused on superficial demographic information instead of focusing real needs and wants of the customer base. Also, the information presented tends to be static and not dynamic.*

**lack of prediction** (Quote 3) – *personas are descriptive and not predictive; they cannot be used for prediction, unlike other analytics tools.*

To understand this criticism, we must bear in mind that personas are inherently connected to decision making about customers. In this sense, they are analytical tools and fall under the scope of other analytics solutions when applied in practical use cases. As such, practitioners using personas are questioning them in comparison to other analytical tools. For example, it is now commonplace to target individual users within digital marketing. Online advertising platforms, such as Facebook Ads and Google AdWords, have constructed social graphs and knowledge graphs (Venkataramani et al. 2012) with each user a node with descriptive properties that can be used for advertisement targeting. Moreover, there have been substantial algorithmic advances that enhance targeting and optimization of advertising (Graepel et al. 2010; Wang & Yuan 2015). Techniques such as multi-armed bandits define the search space and find the best matches given an overall target group or population (Chatwin 2013). Given that users can be targeted and analyzed individually, what purpose is there for aggregated data representations, such as personas? To answer this question, we examine the ability of data-driven personas.

## **EVALUATING DATA-DRIVEN PERSONAS IN THE LIGHT OF CRITICISM: DISTINGUISHING BETWEEN TRADITIONAL AND DIGITAL DATA-DRIVEN PERSONAS**

Cooper's (1999) initial idea was for personas to be data-driven, i.e., based on real insights about the users. Later, other scholars working with personas have confirmed this view of personas originating from comprehensive investigative data collection among real customers (Chapman et al. 2008; Howard 2015; Pruitt & Grudin 2003). However, the conceptual difference between manual and digital data-driven personas has been poorly established in the prior literature, even though this distinction is central for understanding the role of personas amidst digital data. Therefore, we separate between traditional data-driven personas (TDDPs) and digital data-driven personas (DDDPs) and focus on analyzing their strengths for solving the criticism proposed by scholars and practitioners.

What are DDDPs, then? The primary descriptor of DDDPs is that they *bridge persona creation between quantitative data and computational techniques*. Several examples can be



found in the literature. For example, Chapman et al. (2008) use conjoint analysis to reduce a data-set to persona-like representations. Zhang, Brown, and Shankar (2016) use click-stream data to generate personas of website visitors. (An, Kwak & Jansen 2017) develop a methodology and system for completely automatic persona generation using YouTube audience statistics. These novel approaches illustrate ways to combine online analytics data and personas in a way that draws from the benefits of large-scale online data and retains the core benefits of personas in showing the data as another human being.

Online analytics data and computational techniques for processing the data provide at least four major advantages for persona creation (Salminen et al. 2017): (1) the possibility to automatically collect large volumes of data through application programming interfaces (APIs), (2) the availability of behavioral data (not only survey and interview responses), providing better grounds for statistical methods (Chapman et al. 2008), (3) scalability, meaning that data analysis algorithms and automatic systems can process millions of user interactions from millions of content pieces, (4) near real-time responsiveness, enabling customer insights to change as the underlying data changes. These features also make DDDPs different from mere “quantitative personas” suggested in the literature (Mesgari, Okoli & de Guinea 2018). The concept of quantitative persona captures the statistical aspect of data analysis, but it does not correctly capture the aspects of automation and large-scale data analysis associated with the use of online analytics data characterized by high volume, velocity, veracity, and variety (Storey & Song 2017).

At its best, automatic data collection and analysis is cost-efficient and behaviorally accurate across the whole user base, providing excellent foundations for the creation of data-driven personas. Additionally, DDDPs do not necessarily need to be either quantitative or qualitative, but they can draw from both types of data, as demonstrated by Salminen, Şengün, et al. (2018) with their hybrid personas created using quantitative online analytics data and qualitative insights.

How are DDDPs, then, able to solve persona challenges? And how do they compare against TDDPs? We perform a conceptual analysis evaluating these questions for each point of criticism laid out in the previous sections. After this, we discuss the findings and evaluate the benefits. Table 3 provides a comparison of the ability of TDDPs and DDDPs to address the criticism.

If we contrast DDDPs against the criticism of personas laid out in the previous sections, we find that they have the potential to solve many acute problems. First, automation enables rapid persona creation. Whereas the creation of personas using manual methods, such as ethnography and surveys, can take several months, digital persona generation systems are able to run the required calculations in the matter of a few hours (Jung et al. 2017). Second, the personas can be generated by inferring latent patterns of users’ behavior, e.g. video viewing or website browsing (An, Kwak & Jansen 2017). This technique is robust against personal biases of human creators and produces personas that are based on behavioral data instead of self-stated data that has been shown vulnerable to respondent bias (Fisher 1993). Third, representativeness of the sample is not an issue, when the persona generation is based on the whole user base. For example, Salminen, Şengün, et al. (2017) generated personas from YouTube data of a major online news media company consisting of millions of viewers. Using online analytics data potentially solves the trade-off of relying on either qualitatively rich but non-verifiable data or using numbers that are accurate but lose the immersion of another human being.

Table 3: Evaluation of TDDPs and DDDPs against persona criticism.

		<i>Applies to</i>	
Category	Key issues	TDDPs	DDDPs
Creation of personas	a lot of effort and time needed to create quality personas	Yes, manual data collection and analysis	No, data collection and analysis can be automated
	personas are expensive to create	Yes, because they require manual labor	No, because persona generation can be automated and replicated
	personas can be biased by their creators	Yes, because information selection is made subjectively	No, because algorithms decide the information shown
	personas are based on non-representative data	Yes, because sampling is limited	No, because one can sample the whole customer base
Evaluation of personas	personas lack credibility	Potentially yes	Potentially yes
	personas are not accurate or verifiable	Yes, qualitative analysis is difficult to replicate systematically	No, the results can be statistically evaluated
	the information in personas is not relevant for decision makers	Potentially yes	Potentially yes
	personas are inconsistent	Potentially yes	Potentially yes
Use of personas	not using the created personas	Potentially yes	Potentially yes
	using the personas for politics and power play	Potentially yes	Potentially yes
	using personas to justify one's preconceived notions	Potentially yes	Potentially yes
	personas change in time	Yes, because data collection and analysis would need to be repeated	No, because personas can be updated to reflect changes in data
Personas in online analytics context	increased complexity	Yes, because the number of personas is limited	No, because the personas can capture a large number of patterns
	redundant information	Potentially yes	Potentially yes
	lack of prediction	Potentially yes	No, because the underlying data can be used for prediction

Fourth, the “file drawer effect” (Rönkkö et al. 2004) becomes an aggravated issue for TDDPs as time goes by because of the underlying customer behavior changes over time but the manually collected data does not change to reflect these changes. DDDPs are responsive to those changes because the personas can be periodically updated to reflect the most recent behaviors. If the data is directly obtained from the APIs of online platforms, this update process can be completely automated, as demonstrated by (An, Kwak & Jansen 2017). Regarding the DDDPs in the context of online analytics, several studies support our argument that DDDPs can handle the increasing complexity of audiences and customer bases. For example, (Liapis et al. 2015) are able to infer dozens of behavioral gaming patterns from their dataset. Salminen, Şengün, et al. (2018) summarize the audience of a large social media channel, consisting of viewers from more than 200 countries, to five representative personas. (Kwak, An & Jansen 2017) identify hundreds of unique content consumption patterns among channel users that they use to generate personas.

An example of prediction is demonstrated by Jung, Salminen, An, Kwak, and Jansen (2018). They use online analytics data for persona generation, and then predict the interest of a persona to a given video using an underlying topic matching algorithm. Even though some prediction tasks could be done using TDDPs (e.g., “Would Martin like our content?”), the method of doing so inevitably involves a prominent level of subjectivity, whereas using DDDPs the algorithm treats the prediction as a numerical problem.

## LIMITATIONS OF DDDPs

From the evaluation in Table 3, we see that DDDPs have the greatest potential for solving persona creation and digital context problems. Some of the evaluation issues can also be addressed. In contrast, they do not seem to provide a considerable advantage for use of personas, as the application is subject to organizational and individual biases. Moreover, consistency has been found an issue also in DDDPs (Salminen, Nielsen, et al. 2018) as in TDDPs (Bødker et al. 2012). In fact, inconsistency might be even heightened in automatic persona generation, as there is a subjective safeguard for making sure that the information pieces selected by the machine are topically consistent (An, Kwak & Jansen 2017). Moreover, even in the context of DDDPs, information selection remains a challenge, as Salminen, Jung, et al. (2018) observed in their user study.

Despite the progress made in developing DDDPs, there remain many open challenges, such as reliance on current audience data, lack of depth, and a discerning lack of basic attributes in many of the approaches (i.e., not generating persona profiles but behavioral archetypes). Another major limitation of the current DDDPs methodologies is that none of them include deeper information and insights about the users, such as customer pain points, motivations, needs and wants that are essential for the depiction of full, rounded personas (Nielsen 2013). Understanding the deeper motivations of customers is an essential question for marketers (Dichter 1964). Furthermore, while the existing DDDPs may be efficient in modeling current audiences, the decision makers might be interested in potential customers (Thoma & Williams 2009). This interest can be explained by the expansive goals of a typical marketing organization; that is, marketers are pressured to find novel audiences and markets. However, when the DDDPs are generated from current audience data (Jung et al. 2018), such information is not readily available. Thus, we conjecture that the greater the need for reaching new audiences, the riskier it is to use DDDPs for decision making. Moreover, the use of existing online analytics data can lead to confirmation bias. For example, a decision maker may only target *Women, age 25–34* with his efforts, so when DDDPs reveal to him that the group is indeed his core customers, he

carries on saying “I was right,” maintaining his targeting and never, in fact, trying out other target groups. In other words, DDDPs must be interpreted carefully.

Furthermore, a major limitation of DDDPs is the lack of participation of the team using the personas. The process of persona creation has been found valuable *per se*, as participation increases the decision makers’ interest in and understanding of personas (Molenaar 2017). This aspect is lacking in DDDPs that are created by algorithms “in distance”, representing a potential threat for adoption and active use of the persona among decision makers (Matthews, Judge & Whittaker 2012). It is, however, possible to create so-called hybrid personas (Miaskiewicz, Sumner & Kozar 2008; Salminen et al. 2017) that combine quantitative and qualitative aspects. Finally, like for TDDPs, another major limitation for DDDPs is that the benefits postulated in the existing persona literature remain *potential*, depending on the decision makers’ actual willingness to use the personas. Table 4 evaluates the applicability of persona benefits to TDDPs and DDDPs.

Table 4: Applicability of persona benefits to TDDPs and DDDPs

		<i>Applies to</i>	
Category	Description	TDDPs	DDDPs
Communication	Personas facilitate user-oriented communication within and between teams in the organization.	Potentially yes	Potentially yes
Psychology	Personas enhance the immersion required for designing ‘for a person’ instead of fuzzy and complex target groups.	Potentially yes	Potentially yes
Transformation	Personas challenge existing assumptions about customers and orientate trade-off decisions when customers have conflicting needs.	Potentially yes	Potentially yes
Focus	Personas help focus design decisions on user goals and needs rather than on system attributes and features.	Potentially yes	Potentially yes

The benefits of personas are a question of value in use (Kaartemo, Akaka & Vargo 2017). That use varies by use case and user of personas. Generally, the same challenges in transcending the theoretical value of personas to practice apply to both TDDPs and DDDPs. However, as alternatives to numerical online analytics data, personas do have some distinct advantages. For example, dealing with numbers poses cognitive challenges for individuals who often cannot recall many numbers at a time (Miller 1956), whereas human attributes are more easily remembered (Mulken, André & Müller 1998). Therefore, DDDPs seem to provide an ample alternative for presenting numerical data, even though they are not perfect.

## DISCUSSION

In this research, we examined the role of personas given the widespread availability of online analytics data, and its perceived useful for business purposes. We reviewed scholarly criticism of personas and extended it with fresh perspectives from practitioners working with online analytics data. Although our inquiry is conceptual in its nature, these quotations served as a basis for understanding the challenges and disadvantages of personas in the modern customer analytics environment. To better address the question of personas' usefulness in this environment, we dissected the concept of personas into TDDPs and DDDPs. This separation was crucial to make the analysis because the two types of personas are different in their ability to address the challenges associated with personas in the context of online analytics.

The main contributions of this research include

- a) identifying new critical arguments against personas in the context of online analytics, previously not discussed in persona studies, and appending this criticism to the continuum of established persona criticism; and
- b) conceptually differentiating between traditional and digital data-driven personas and separately analyzing their ability to address the established and novel criticism of personas, with the finding that DDDPs possess considerable strengths in regard to both types of criticism.

Regarding the answer to our research question, namely can personas provide value in the age of online analytics, we answer that the real value of DDDPs is provided by *giving faces to data*, as an alternative way to present online analytics information. Ultimately, however, the usefulness of personas comes down to specific use cases (Cooper 1999). In general, personas are useful for tasks requiring a qualitative understanding of customers and numbers are useful for getting a general overview and, in the case of machine-based decision-making (i.e., marketing automation), making individual level optimization. To this end, we postulate that individual data is optimal for automated decision making, whereas aggregate data such as personas work best for human decision making, especially relating to decisions at the strategic level.

The claimed irrelevance of personas seems to be based on the confusion of their use in an age of online analytics data. It seems that the criticism presented by online analytics practitioners is based on understanding personas as TDDPs, while overlooking the potential of DDDPs. This insight further supports the purpose of conceptually separating these two types of personas, and clearly communicating their differences to end users of online analytics data.

Data-driven personas are not a novel idea, as the purpose of persona creation has always been to use real customer information to generate realistic user characterizations (Cooper 1999). However, using computational techniques, such as machine learning, provide tremendous opportunities toward this end (Salminen, Jansen, et al. 2018). At the same time, there are a plethora of open research questions to answer. Moreover, the persona research related to DDDPs tends to be fragmented, while it would make sense for researchers to collaborate and validate the works of one another to make evolutionary progress in this field.

The core benefits of using personas for design, system development, and marketing have not changed. However, for tasks such as targeting or recommendation engines, individual level analytics are likely to perform more efficiently. While these methods are likely to excel in those use cases, their application to other use cases, such as strategic decision making, is more limited. As with any analytical technique, the use of personas is relative to the problem one seeks an

answer to. In this light, the question “*Are personas useful?*” should be rephrased as “*When are personas useful?*” Finding the answer requires persistent conceptual and empirical research and exploring new contexts such as online analytics.

## CONCLUSION

In conclusion, personas remain a viable option even in the era of online analytics data. It is possible to combine automatic data collection and other computational techniques to create accurate persona profiles that can also be used for advanced purposes such as prediction. However, some of the challenges of TDDPs are inevitably inherited such as the end user relating to the actual use of personas and their perceived credibility. Therefore, more research and development work is needed to overcome these challenges and to show the tangible value of personas in actual use.

## END NOTES

<sup>i</sup> See <http://www.cmo.com/features/articles/2017/1/16/why-personas-dont-work-and-what-innovators-are-doing-differently.html#gs.zxQ5E9s>

<sup>ii</sup> See <https://www.humaneinterface.net/article/are-personas-really-useful>

<sup>iii</sup> See [http://knowledge.wharton.upenn.edu/article/160811b\\_kwradio\\_fader-mariychin-mp3-zodiac/](http://knowledge.wharton.upenn.edu/article/160811b_kwradio_fader-mariychin-mp3-zodiac/)

<sup>iv</sup> See <http://thecontextofthings.com/2016/08/25/user-personas/>

## WORKS CITED

- An, J, Kwak, H & Jansen, BJ 2017, ‘Personas for Content Creators via Decomposed Aggregate Audience Statistics’, in *Proceedings of Advances in Social Network Analysis and Mining* (ASONAM 2017), Sydney, Australia.
- Ardehaly, EM & Culotta, A 2015, ‘Inferring latent attributes of Twitter users with label regularization’, retrieved 29 October 2018 from <http://cs.iit.edu/~culotta/pubs/ehsan15inferring.pdf>.
- Bleier, A & Eisenbeiss, M 2015, ‘Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where’, *Marketing Science*, vol. 34, no. 5, retrieved 18 May 2016, from <http://pubsonline.informs.org/doi/abs/10.1287/mksc.2015.0930>.
- Bødker, S, Christiansen, E, Nyvang, T & Zander, P-O 2012, ‘Personas, people and participation: challenges from the trenches of local government’, *Proceedings of the 12th Participatory Design Conference: Research Papers - Volume 1* (PDC '12), K Halskov, H Winschiers-Theophilus, Y Lee, J Simonsen, and K Bødker (Eds.). ACM, New York, NY, USA, pp. 91-100. DOI=<http://dx.doi.org/10.1145/2347635.2347649>
- Boghrati, R, Hoover, J, Johnson, KM, Garten, J & Deghani, M 2017, ‘Conversation level syntax similarity metric’, *Behavior Research Methods*, vol. 50, no. 3, pp. 1055-1073, doi: 10.3758/s13428-017-0926-2.
- Chapman, CN, Love, E, Milham, RP, ElRif, P & Alford, JL 2008, ‘Quantitative Evaluation of Personas as Information’, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 52, no. 16, pp. 1107-1111.

- Chapman, CN & Milham, RP 2006, 'The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 5, pp. 634–636.
- Chatwin, RE 2013, 'An overview of computational challenges in online advertising', *American Control Conference (ACC) 2013, IEEE*, pp. 5990–6007, retrieved 29 October 2018 from <http://ieeexplore.ieee.org/abstract/document/6580778/>.
- Clarke, TB & Jansen, BJ 2017, 'Conversion potential: a metric for evaluating search engine advertising performance', *Journal of Research in Interactive Marketing*, vol. 11, no. 2, pp. 142–159.
- Cleary, PD 1997, 'Subjective and Objective Measures of Health: Which is Better When?', *Journal of Health Services Research*, vol. 2, no. 1, pp. 3–4.
- Cooper, A 1999, *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity*. Sams - Pearson Education, Indianapolis, IN.
- Del Vecchio, P, Mele, G, Ndou, V & Secundo, G 2017, 'Creating value from Social Big Data: Implications for Smart Tourism Destinations', *Information Processing & Management*, vol. 54, no. 5, pp. 847-860, doi: <https://doi.org/10.1016/j.ipm.2017.10.006>
- Delfabbro, P 2004, 'The stubborn logic of regular gamblers: Obstacles and dilemmas in cognitive gambling research', *Journal of Gambling Studies*, vol. 20, no. 1, pp. 1–21.
- Dichter, E 1964, *Handbook of consumer motivations: the psychology of the world of objects*. McGraw-Hill.
- Dong, J, Kelkar, K & Braun, K 2007, 'Getting the most out of personas for product usability enhancements', Usability and Internationalization. HCI and Culture. UI-HCII 2007. Lecture Notes in Computer Science, N Aykin (eds), vol 4559. Springer, Berlin, Heidelberg, doi: [https://doi.org/10.1007/978-3-540-73287-7\\_36](https://doi.org/10.1007/978-3-540-73287-7_36).
- Fisher, RJ 1993, 'Social Desirability Bias and the Validity of Indirect Questioning', *Journal of Consumer Research*, vol. 20, no. 2, pp. 303–315.
- Floyd, IR, Jones, MC & Twidale, MB 2008, 'Resolving Incommensurable Debates: A Preliminary Identification of Persona Kinds, Attributes, and Characteristics', *Artifact*, vol. 2, no. 1, pp. 12–26.
- Friess, E 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, pp. 1209–1218, retrieved from <http://doi.acm.org/10.1145/2207676.2208572>.
- Goodwin, K 2009, *Designing for the Digital Age: How to Create Human-Centered Products and Services*. Wiley, Indianapolis, IN.
- Graepel, T, Candela, JQ, Borchert, T & Herbrich, R 2010, 'Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine', in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 13–20, retrieved September 20, 2017, from [http://machinelearning.wustl.edu/mlpapers/paper\\_files/icml2010\\_GraepelCBH10.pdf](http://machinelearning.wustl.edu/mlpapers/paper_files/icml2010_GraepelCBH10.pdf).
- Gudjónsdóttir, R 2010, *Personas and scenarios in use, Skolan för datavetenskap och kommunikation, Kungliga Tekniska högskolan*, Stockholm.
- Hill, CG, Haag, M, Oleson, A, Mendez, C, Marsden, N, Sarma, A & Burnett, M 2017, 'Gender-Inclusiveness Personas vs. Stereotyping: Can We Have it Both Ways?', *CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing*, pp. 6658–6671, retrieved May 28, 2017, from <http://dl.acm.org/citation.cfm?doid=3025453.3025609>.
- Holtzblatt, K, Wendell, JB & Wood, S 2005, *Rapid Contextual Design: A How-to Guide to Key Techniques for User-centered Design*, Elsevier.
- Howard, TW 2015, 'Are Personas Really Usable?', *Communication Design Quarterly Review*, vol. 3, no. 2, pp. 20–26.
- Jansen, BJ, An, J, Kwak, H, Salminen, J & Jung, S-G 2017, 'Viewed by Too Many or Viewed Too Little: Using Information Dissemination for Audience Segmentation', *Proceedings of the*

- Association for Information Science and Technology*, vol. 54, no. 1, pp. 189–196, doi: <https://doi.org/10.1002/pra2.2017.14505401021>.
- Järvinen, J 2016, 'The use of digital analytics for measuring and optimizing digital marketing performance', *Jyväskylä Studies in Business and Economics* 170, retrieved September 8, 2017, from <https://jyx.jyu.fi/dspace/handle/123456789/51512>.
- Järvinen, J & Karjaluoto, H 2015, 'The use of Web analytics for digital marketing performance measurement', *Industrial Marketing Management*, vol. 50, pp. 117–127, <https://doi.org/10.1016/j.indmarman.2015.04.009>.
- Johnson, MD & Selnes, F 2004, 'Customer portfolio management: toward a dynamic theory of exchange relationships', *Journal of Marketing*, vol. 68, no. 2, pp. 1–17.
- Jung, S, Salminen, J, An, J, Kwak, H & Jansen, BJ 2018, 'Automatically Conceptualizing Social Media Analytics Data via Personas', *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)*, retrieved 29 October 2018 from <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17810/17089>.
- Jung, S-G, An, J, Kwak, H, Ahmad, M, Nielsen, L & Jansen, BJ 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*, pp. 1748–1755, retrieved 29 October 2018 from [http://www.bernardjansen.com/uploads/2/4/1/8/24188166/jansen\\_personas\\_chi2017.pdf](http://www.bernardjansen.com/uploads/2/4/1/8/24188166/jansen_personas_chi2017.pdf).
- Jung, S-G, An, J, Kwak, H, Salminen, J & Jansen, BJ 2017, 'Inferring social media users' demographics from profile pictures: A Face++ analysis on Twitter users', in *Proceedings of 17th International Conference on Electronic Business*, Dubai, retrieved 29 October 2018 from <https://aisel.aisnet.org/iceb2017/22/>.
- Kaartemo, V, Akaka, MA & Vargo, SL 2017, 'A Service-Ecosystem Perspective on Value Creation: Implications for International Business', in *Value Creation in International Business*, Palgrave Macmillan, Cham, pp. 131–149, retrieved 11 February 2018, from [https://link.springer.com/chapter/10.1007/978-3-319-39369-8\\_6](https://link.springer.com/chapter/10.1007/978-3-319-39369-8_6).
- Krashen, SD 1984, 'Immersion: Why it works and what it has taught us', *Language and Society*, vol. 12, no. 1, pp. 61–64.
- Kwak, H, An, J & Jansen, BJ 2017, 'Automatic Generation of Personas Using YouTube Social Media Data', *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 833–842, retrieved 29 October 2018 from [http://www.bernardjansen.com/uploads/2/4/1/8/24188166/jansen\\_personas\\_hciss2016.pdf](http://www.bernardjansen.com/uploads/2/4/1/8/24188166/jansen_personas_hciss2016.pdf).
- LeRouge, C, Ma, J, Sneha, S & Tolle, K 2013, 'User profiles and personas in the design and development of consumer health technologies', *International Journal of Medical Informatics*, vol. 82, no. 11, pp. e251–268.
- Liapis, A, Holmgård, C, Yannakakis, GN & Togelius, J 2015, 'Procedural Personas as Critics for Dungeon Generation', in AM Mora & G Squillero (eds), *Applications of Evolutionary Computation*, Springer International Publishing, Cham, pp. 331–343, retrieved 14 March 2018 from [http://link.springer.com/10.1007/978-3-319-16549-3\\_27](http://link.springer.com/10.1007/978-3-319-16549-3_27).
- Long, F 2009, 'Real or imaginary: The effectiveness of using personas in product design', in *Proceedings of the Irish Ergonomics Society Annual Conference*, Irish Ergonomics Society Dublin.
- Ma, J & LeRouge, C 2007, 'Introducing User Profiles and Personas into Information Systems Development', *AMCIS 2007 Proceedings*, retrieved 29 October 2018 from <http://aisel.aisnet.org/amcis2007/237>.
- Marsden, N & Haag, M 2016, 'Stereotypes and Politics: Reflections on Personas', in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, pp. 4017–4031, retrieved 18 March 2018, from <http://doi.acm.org/10.1145/2858036.2858151>.
- Massanari, AL 2010, 'Designing for imaginary friends: information architecture, personas and the politics of user-centered design', *New Media & Society*, vol. 12, no. 3, pp. 401–416.



- Matthews, T, Judge, T & Whittaker, S 2012, 'How Do Designers and User Experience Professionals Actually Perceive and Use Personas?', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pp. 1219–1228, retrieved 29 October 2018 from <http://doi.acm.org/10.1145/2207676.2208573>.
- Mesgari, M, Okoli, C & de Guinea, AO 2018, 'Creating Rich and Representative Personas by Discovering Affordances', *IEEE Transactions on Software Engineering*, doi: [10.1109/TSE.2018.2826537](https://doi.org/10.1109/TSE.2018.2826537)
- Miaskiewicz, T & Kozar, KA 2011, 'Personas and user-centered design: How can personas benefit product design processes?', *Design Studies*, vol. 32, no. 5, pp. 417–430.
- Miaskiewicz, T, Sumner, T & Kozar, KA 2008, 'A latent semantic analysis methodology for the identification and creation of personas', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1501–1510, retrieved from <http://dl.acm.org/citation.cfm?id=1357290>.
- Miller, GA 1956, 'The magical number seven plus or minus two: some limits on our capacity for processing information', *Psychological Review*, vol. 63, no. 2, pp. 81–97.
- Miralles-Pechuán, L, Ponce, H & Martínez-Villaseñor, L 2018, 'A novel methodology for optimizing display advertising campaigns using genetic algorithms', *Electronic Commerce Research and Applications*, vol. 27, pp. 39–51.
- Molenaar, L 2017, 'Data-driven personas: Generating consumer insights with the use of clustering analysis from big data', Thesis, retrieved 17 March 2018, from <http://resolver.tudelft.nl/uuid:12d7f261-20b4-4656-93d7-fed2b437aefb>.
- Mulder, S & Yaar, Z 2006, *The User is Always Right: A Practical Guide to Creating and Using Personas for the Web*, New Riders.
- Mulken, S van, André, E & Müller, J 1998, 'The Persona Effect: How Substantial Is It?', in *People and Computers XIII*, Springer, London, pp. 53–66, retrieved 28 February 2018, from [https://link.springer.com/chapter/10.1007/978-1-4471-3605-7\\_4](https://link.springer.com/chapter/10.1007/978-1-4471-3605-7_4).
- Nacke, LE, Drachen, A & Göbel, S 2010, 'Methods for evaluating gameplay experience in a serious gaming context', *International Journal of Computer Science in Sport*, vol. 9, no. 2, pp. 1–12.
- Nielsen, L 2013, *Personas - User Focused Design*, Springer Science & Business Media, Springer-Verlag London.
- Nielsen, L, Jung, S-G, An, J, Salminen, J, Kwak, H & Jansen, BJ 2017, 'Who Are Your Users?: Comparing Media Professionals' Preconception of Users to Data-driven Personas', in *Proceedings of the 29th Australian Conference on Computer-Human Interaction, OZCHI '17*, pp. 602–606, retrieved 28 January 2018, from <http://doi.acm.org/10.1145/3152771.3156178>.
- Owusu, RA, Mutshinda, CM, Antai, I, Dadzie, KQ & Winston, EM 2016, 'Which UGC features drive web purchase intent? A spike-and-slab Bayesian Variable Selection Approach', *Internet Research*, vol. 26, no. 1, pp. 22–37.
- Provan, KG & Milward, HB 1995, 'A preliminary theory of interorganizational network effectiveness: A comparative study of four community mental health systems', *Administrative Science Quarterly*, vol. 40, no. 1, pp. 1–33.
- Pruitt, J & Adlin, T 2006, *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. Morgan Kaufmann, Boston.
- Pruitt, J & Grudin, J 2003, 'Personas: Practice and Theory', in *Proceedings of the 2003 Conference on Designing for User Experiences, DUX '03*, pp. 1–15, retrieved from <http://doi.acm.org/10.1145/997078.997089>.
- Ronen, R, Yom-Tov, E & Lavee, G 2016, 'Recommendations meet web browsing: enhancing collaborative filtering using internet browsing logs', *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pp. 1230–1238, doi: [10.1109/ICDE.2016.7498327](https://doi.org/10.1109/ICDE.2016.7498327).
- Rönkkö, K 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, retrieved 29 October 2018 from <http://dx.doi.org/10.1109/HICSS.2005.85>.
- Rönkkö, K, Hellman, M, Kilander, B & Dittrich, Y 2004, 'Personas is Not Applicable: Local Remedies Interpreted in a Wider Context', in *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1, PDC 04*, pp. 112–120, retrieved 29 October 2018 from <http://doi.acm.org/10.1145/1011870.1011884>.
- Rundle-Thiele, S, Dietrich, T & Kubacki, K 2017, 'Why We Need Segmentation When Designing Social Marketing Programs', in *Segmentation in Social Marketing*, Springer, Singapore, pp. 197–214, retrieved 9 May 2018, from [https://link.springer.com/chapter/10.1007/978-981-10-1835-0\\_13](https://link.springer.com/chapter/10.1007/978-981-10-1835-0_13).
- Russell, S & Toklu, C 2011, 'Acquiring and Applying Market Knowledge for Large Software Purchases: Products, Personas, and Programs', in *Knowledge Management for Process, Organizational and Marketing Innovation: Tools and Methods*, IGI Global, pp. 195–215.
- Saggi, MK & Jain, S 2018, 'A survey towards an integration of big data analytics to big insights for value-creation', *Information Processing & Management*, vol. 54, no. 5, pp. 758–790.
- Salminen, J, Jansen, BJ, An, J, Kwak, H & Jung, S-G 2018, 'Research Roadmap for Automatic Persona Generation: Principles and Open Questions', Poster presented at *The Fifth Machine Learning and Data Analytics (MLDAS 2018) Symposium*, Doha, Qatar.
- Salminen, J, Jung, S-G, An, J, Kwak, H & Jansen, BJ 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*, p. LBW097:1–LBW097:6, retrieved 26 April 2018, from <http://doi.acm.org/10.1145/3170427.3188470>.
- Salminen, J, Milenković, M & Jansen, BJ 2017, 'Problems of Data Science in Organizations: An Explorative Qualitative Analysis of Business Professionals' Concerns', in *Proceedings of International Conference on Electronic Business (ICEB 2017)*, Dubai.
- Salminen, J, Nielsen, L, Jung, S-G, An, J, Kwak, H & Jansen, BJ 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)*, Montréal, Canada.
- Salminen, J, Şengün, S, Kwak, H, Jansen, BJ, An, J, Jung, S, Vieweg, S & Harrell, F 2017, 'Generating Cultural Personas from Social Data: A Perspective of Middle Eastern Users', in *Proceedings of The Fourth International Symposium on Social Networks Analysis, Management and Security (SNAMS-2017)*, Prague, Czech Republic.
- 2018, 'From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas', *First Monday*, vol. 23, no. 6, retrieved 3 June 2018, from <http://firstmonday.org/ojs/index.php/fm/article/view/8415>.
- Scott, DM 2007, *The New Rules of Marketing*, John Wiley and Sons, Hoboken, New Jersey.
- Sinha, R 2003, 'Persona Development for Information-rich Domains', in *Extended abstracts of the 2003 Conference on Human Factors in Computing Systems (CHI 2003)*, Ft. Lauderdale, Florida, USA.
- Storey, VC & Song, I-Y 2017, 'Big data technologies and management: What conceptual modeling can do', *Data & Knowledge Engineering*, vol. 108, March 2017, pp. 50–67.
- Thoma, V & Williams, B 2009, 'Developing and Validating Personas in e-Commerce: A Heuristic Approach', in *Human-Computer Interaction – INTERACT 2009, Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, pp. 524–527, retrieved 29 October 2018 from [https://link.springer.com/chapter/10.1007/978-3-642-03658-3\\_56](https://link.springer.com/chapter/10.1007/978-3-642-03658-3_56).
- Vahlo, J & Koponen, A 2018, 'Player Personas and Game Choice', in N Lee (ed), *Encyclopedia of Computer Graphics and Games*, Springer International Publishing, Cham, pp. 1–6, retrieved March 6, 2018, from [http://link.springer.com/10.1007/978-3-319-08234-9\\_149-1](http://link.springer.com/10.1007/978-3-319-08234-9_149-1).
- Venkataramani, V, Amsden, Z, Bronson, N, Cabrera III, G, Chakka, P, Dimov, P, Ding, H, Ferris, J, Giardullo, A, Hoon, J & others 2012, 'Tao: how Facebook serves the social graph', in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*,

- pp. 791–792, retrieved 29 October 2018 from <http://dl.acm.org/citation.cfm?id=2213957>.
- Vincent, CJ & Blandford, A 2014, 'The challenges of delivering validated personas for medical equipment design', *Applied Ergonomics*, vol. 45, no. 4, pp. 1097–1105.
- Volkova, S, Bachrach, Y & Durme, BV 2016, 'Mining User Interests to Predict Perceived Psycho-Demographic Traits on Twitter', in *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*, pp. 36–43, retrieved 29 October 2018 from <https://ieeexplore.ieee.org/document/7474353>.
- Wang, J & Yuan, S 2015, 'Real-Time Bidding: A New Frontier of Computational Advertising Research', in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM '15*, pp. 415–416, retrieved 28 February 2016, from <http://doi.acm.org/10.1145/2684822.2697041>.
- Zerbino, P, Aloini, D, Dulmin, R & Mininno, V 2018, 'Big Data-enabled Customer Relationship Management: A holistic approach', *Information Processing & Management*, vol. 54, no. 5, pp. 818–846.
- Zhang, X, Brown, H-F & Shankar, A 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*, , pp. 5350–5359, doi: [10.1145/2858036.2858523](https://doi.org/10.1145/2858036.2858523).