International Journal of Marketing, Communication and New Media
ISSN: 2182-9306. Vol. 7, Nº 12, JUNE 2019

Research Paper

Is Your Application Gender Biased?

Milena Lopes*
Carl Vogel**

ABSTRACT

The process of design is underpinned by a solid methodology that comprises stages of research, creation and validation of the ideas in order to ensure that a product or a service will meet the users' needs. However, in the realm of online dating applications, women reported to have a frustrating experience, marked by sexist remarks and offensive behaviour. When design fails to deliver a good experience and to meet women's needs, one question rises: are applications biased in the beginning of the design process? This research proposes a method to investigate biases in the design of mobile experiences.

Keywords: Mobile applications; Gender biases; Online dating; Design methods.

* Trinity College Dublin. E-Mail: lopesm@tcd.ie

** Trinity College Dublin. E-Mail: vogel@tcd.ie

Received on: 2019.04.11

Approved on: 2019.06.30

Evaluated by a double blind review system

1. INTRODUCTION

This work proposes a method to investigate gender bias in the development of mobile applications and to assess the perception of gender by users. The elaboration of the method is part of a broader programme about the impact of gender in design.

The development of mobile interfaces encompasses the stages of the creative design process, some tools that are common to different areas of design, and several creative tools that are specific to mobile devices. However, gender is not highlighted in the process as a factor that can affect the communication, neither in the process of design nor in the tools used to design interactions and the graphical user interfaces (GUI). There are approaches such as the "user centred design" (UCD) that focus on including users in the beginning of the process as a way to meet their needs, but in an environment where development teams are dominated by men and where people sometimes take decision based on their own opinions, it is likely that women's perspective over the matter is not taken into account (Williams, 2014). In order to investigate whether that assertion is true, that is, whether there are gender biases in the creative process of design, we have developed the method reported here that involves the early stage of setting requirements of a design project. Our method was initially developed for online dating applications and that is the reason our examples here will be limited to this topic. Nonetheless, the method can be adapted and used to evaluate other applications.

A previous study was carried out in order to investigate women's perception of online dating applications. For that study, 40 female users of Tinder were interviewed and their responses were analysed through a mixed methods approach, incorporating both qualitative and quantitative data (Lopes & Vogel, 2017a). A subjective analysis made it possible to understand gaps, problems and frustrations through female users' eyes, and

with posterior quantitative analysis it was possible to judge whether there was a problem regarding the female experience. For the quantitative analysis, the responses were transformed into summarised words, grouped into categories and computed in statistical software in order to reveal the number of women that pointed out the same categories of answers. The analysis of the results revealed that 70% of the female users experienced some type of offensive behaviour with regards to gender dynamics (as unwanted direct sexual approach and sexist statements) during conversations with male users and that more than 50% uninstalled the application after a frustrated experience. Moreover, when asked about the developer gender, 85% of female users said they thought the developer was a man while only 3% thought the developer was a woman. Those numbers arguably show that the most popular dating application disregards women's needs and expectations.

That first study revealed three important facts that gave rise to a research plan to investigate gender bias in the context of interface design. First, the high percentage of sexist behaviour drew attention to a possible harmful dynamic within the application. Second, the feeling that the developer is a man by a large majority uncovers the perception of the male dominance in technology by female users. Third, the high rate of overall dissatisfaction indicates a design failure in the early stages of the design process. These facts may be caused by gender biases or worsened by it. For this reason, the method presented here concentrates on the presence of gender biases during the design of the experience and graphical interface, and on the user's perception of gender in design.

This research makes the assumptions informed by the results of Lopes and Vogel (Lopes & Vogel, 2017b) that dating applications do not meet female users' needs, possibly due to gender biases that are implicit in the design of the interaction and the GUI, which privileges males' interests rather than gender-balanced interests. Thus, this is the problem set that initiates the creation of a method to investigate whether there are gender effects in the design of mobile interfaces.

This work is organised in 4 sections. We introduce the topic and the problem we want to solve in this section, describe the design studies that support the creation of the method (see item 2. Background), expand the rationale of the research with the description of the method itself (see item 3. Method), and, to conclude, we discuss the expected results and limitations of the method (see item 4. Discussion and Conclusion).

The Method described here is divided in two stages; that is, there are two different studies within the method, and they complement each other. Using the first, it is possible to anticipate problems with regard to gender biases in the early stages of a project. This stage can elicit information about female and male users' preferences in a manner that allows designers to make better decisions concerning both groups. Using the second study within the method, it is possible to either analyse the perception of gender by users of an application that already exists or to evaluate the perception of gender by users in early stages of a project. In our own validation exercises, we use both research designs in succession.

2. BACKGROUND

In everyday interactions, there are socially constructed ideas that pervade communication between people and the world of artefacts around them. Those are social rules and roles, codes of behaviour, among others, that shape the development of products and technology. As a cultural product, design has encoded meanings that represent the political, economic and cultural values of a society and design dominant codes are social and aesthetic (Buckley, 1986). For example, an experiment conducted by Stonewall and Dorneich (Stonewall & Dorneich, 2016) showed that the perception of professionalism in websites is strongly connected to elements perceived as inherently masculine by users, as dark colours, angular shapes, rugged or business-like images, thick serif fonts, and wood or stone based textures. Those dominant codes imprinted in the area of hardware and software design make room for intrinsic gender biases, which in turn can shape users' behaviour and favour men's interests in the human-human interaction mediated by computers. Friedman and Nissenbaum (Friedman & Nissenbaum, 1996) identified three different types of biases that are implicit in software development: pre-existing, technical and emergent biases. Pre-existing bias is related to social beliefs that are very likely to be embedded in technology (e.g. gender roles, sexism, racism). Technical bias refers to the limitations of the technology itself that fails to address individual needs and preferences. For example, the adoption of the GUI as a main channel of communication with computers left many visually impaired users technologically disempowered. Emergent bias, in turn, refers to improvements or differences that emerge after the system is implemented; a change of behaviour, new discoveries or the appearance of a new population of users will not be addressed by the

system since they were not predicted. Mobile applications are very likely to embody these three types of biases.

Some studies aiming to investigate gender differences in the use of technology or in the production of design revealed findings that reinforce the importance of highlighting gender in the design process. Moss et al. (Moss, Gunn, & Heller, 2006), for example, found out that female users are more likely to use rounded shapes in comparison with men (who tend to prefer straight shapes), to avoid horizontal layouts, to use colours for typography, irregular typography and informal images. An analysis of aesthetics of printed graphic design produced by male and female (Moss & Colman, 2001) identified that both male and female users are more likely to prefer the design produced by people of the same gender as theirs. A different study on gender preferences (Xue & Yen, 2007) also revealed many differences regarding material, texture, form and perceived values, demonstrating that taste is also stereotyped. In the web design realm, a female preference for sites created by other females, for the use of informal language, self-derogation, and for use shortened forms of words in the production of sites were also detected (Moss & Gunn, 2007). Simon and Peppas (Simon & Peppas, 2005) observed that male users feel more comfortable with the use of Internet and website navigation than female, but the authors also address that the difference may occur due the male domination in the field that implicates in developments that attend male needs. Nysveen et al. (Nysveen, Pedersen, & Thorbjørnsen, 2005) investigated the gender differences in the intention to use mobile chats and detected that enjoyment is a strong determinant of intention to use mobile chats for women while usefulness is for men. They found out that expressiveness is more determinant for male users and that they have a high drive for communication, possibly because they feel more comfortable to interact, flirt and express their identity and emotions through chat services than in face-to-face interactions.

A common practice in design is to consider the user as the fundamental key to the development of new technologies. UCD brings the users to the centre of the projects, for which this approach takes in consideration their opinions and needs from the beginning to the end of the process. In this concept, users are seen as co-designers of services and products, and should be involved in many stages of design, as, for example, determining the problem to tackle, setting requirements, testing and evaluating solutions and prototypes. This approach brought a paradigm shift to design, from

drawing attention to technology and activity to consider users' needs as the main purpose for development. Other concepts also put users in the middle of design process, as the universal design approach, which proposes the inclusion of minority groups' needs in the design project (such as people with disabilities, for example). This seems to be a growing trend that could increase awareness of gender biases in design. However, there are some obstacles that limit the adoption of the UCD approach, as described bellow. Despite the attempt to include users in design, companies usually have limited resources to investigate users and many times preference is given to speed up the process of production and to launch products and services as quickly as possible (Oudshoorn, Rommes, & Stienstra, 2004). Moreover, some designers and stockholders are afraid of the leak of ideas to competitors and choose not to involve users in the process.

Gayna Williams (Williams, 2014) draws attention to gender blindness in UCD, and argues that even when the concept is approached there are still many design decisions that are made by individuals in the design teams and that personal opinions influence in the decision making. Considering that teams are largely formed by men, it is more likely that design will be embedded with their point of view. Further, she found out that women do not feel comfortable to bring gender issues to team's discussions because they have learnt to act and think like men (disregarding female needs), they think it is easier to opt for a gender neutral perspective or they do not want to draw attention to the fact of being woman in male-dominated group. Another issue is the idea that the "regular" user is a man (Friedman & Nissenbaum, 1996). Considering that development are usually composed by 20% women and the management positions filled by fewer than 10% (Williams, 2014), and that designers are very likely to bring personal opinions to the project (regardless the users' needs, expectations, political view, lifestyle, and so on), the probability that there are implicit and unconscious gender bias in design is high. For the reasons stated above, we have designed a method to determine the presence of gender biases in the design of applications and to assess user perception of gender biases.

These studies indicate sensitive social factors that pervade the aesthetics and functionality of the design of the GUI and differently affect the perception and use by female and male users. A more focused study would reveal which factors are stereotyping the experience, how those factors affect the communication and solutions

for a more gender balanced service. The method suggested in this work focuses on identifying the presence of gender biases and perceived biases in the design of mobile applications. This is a precondition for identifying the interacting factors.

In order to analyse bias in design, we rely on the design process, which is a methodology based on steps to achieve the creation of a solution for a product, service or space. Herbet Simon (1969) was the first to address principles and guidelines to analyse wicked problems and choose solutions that later would shape the methodology of design, which is also known as design thinking. Contemporary models of the design process are all based on the principles set by Simon. A largely accepted model is proposed by Hasso-Plattner Institute of Design at Stanford and contains five main stages: empathise, define, ideate, prototype and test. In the first stage, the problem or opportunity is analysed so it can be deeply understood; interviews and observations are useful at this point. In the second stage the problem or opportunity is defined and a list of requirements proposed, and it should not happen before studying the users and having dialogues with them in the first stage. The third stage is the ideation, when designers bring out solutions using creative tools like brainstorming, mind map, analogies, and so on. In the fourth stage the idea is implemented (the idea is developed and prototyped), and in the fifth it is validated, that is, tested with users. Different descriptions of design process can include more or fewer stages but those are the basic. In the case fewer stages are proposed, some stages are merged into a single one; in the opposite direction, more stages indicate that a single stage can be split into two or more stages. According to Rogers, Sharp and Preece (Rogers, Sharp, & Preece, 2011), for example, the process of interaction involves four basic activities: establishing requirements, designing alternatives, prototyping and evaluating. In this case, the stage of analysing the problem/oportunity is merged with establishing requirements. For this study we will focus on the stage of "establishing requirements" or "defining" the project, which we believe to be the moment in which user needs are incorporated to design, and, consequently, biases are inscribed to the project.

3. THE METHOD

One may assume that design solutions for the same problem will show differences that are related to who designed them. However, this assumption may be tested empirically. Those can be subtle differences that will not change the user experience or can be

substantial enough to affect the user-mobile device interaction. The early stages of the design process are particularly important to define the application features and behaviour. The process of gathering information, approaching users, setting requirements, designing the experience and generating ideas will inevitably be influenced by the designer or group of designers. Definition of requirements during design process is especially sensitive regarding biases for all the ideas will derive from that list, which can include technical features, values, feelings, communication and engagement approaches, and so on. To investigate whether there are gender biases during the design of applications we propose a user-study method taking in consideration the design creative process, design techniques used to design the experience, and tools used to evaluate the design. The main idea is that female and male participants use some simple design tools to establish design requirements for a general dating application. The design requirements indicated by participants in the first session will be evaluated by others (second session) in order to test whether there is a perceived gender difference. That is, the participants will be asked to guess the gender of each person who designed a dating experience. Moreover, the solutions produced by male and female participants must be compared in order to find significant differences or similarities.

The user study is accordingly divided into two stages. In the first stage (designing the experience), female and male participants must individually complete three tasks that together aim to set requirements for a fictitious online dating application. In the second stage (deducing designers' gender), a different group of participants must try to "guess" the gender of each participant by analysing the solutions of three tasks completed by individuals participating in the first stage - as presented through the materials they will construct (see section 3.3 Procedures). The goal of the first stage is to investigate whether there are identifiable differences between the design produced by men and women and the goal of the second stage is to analyse whether a gender difference is perceived by users.

3.1 Participants and recruitment

The group of participants of first stage (designing the experience) should be composed by equal number of female and male volunteers. They must be representative of a population that is familiar with the task (Blandford, Cox and Cairns, 2008), which should be, in this case, design students, computer science students or professional

designers. The average age should be set in order to have a homogeneous group. The sample size may be composed by a minimum of 30 participants to guarantee representativeness. For our study and its budget limitation, we have defined a maximum of 60, because it could not be exhaustive to the second group to judge the design of the first group. Although a small sample (as of 30 participants) does not necessarily exclude a proper statistical analysis (Sauro and Lewis, 2012), it does limit the power of the study. With such sample only major differences between to groups can be detected without lowering the power of the study or relaxing the confidence interval. Conventionally, the power of a test should be set at 80% and the confidence interval in 95% (Kraemer & Blasey, 2015). This would be ideal for a test with a large sample, but considering the exploratory quality of the study and limitations of sample size, a power greater than 60% and a confidence interval of 90% would be adequate. Using the minimum of 30 participants, 70% power and 90% confidence level, a difference between the two groups would be detected if the difference between the means is 0.8 considering a scale of 0 to 5. However, a larger group would increase the power of the study and, consequently, the precision in detecting gender differences in a biased design.

Participants of second stage (deducing "designers'" gender): female and male designers must be recruited from the same population as the first stage (but who did not participate in the first stage) in order to analyse the design ideas generated by the female and male participants in the first stage. They will try to "guess" each participant's gender by analysing the solution of the three tasks. The minimum number of participants should be 30 for the same reason stated above. In the case a larger group is recruited, it is suggested to split the second group into two and to give half of the solutions produced by the first group to each new group. In this case, each participant of the second group would not judge more than 60 designs.

Participants of the user study can be remunerated or not, however we recommend the remuneration for the time it takes to participants to attend, follow instructions and complete the tasks. Volunteers looking exclusively for remuneration should be excluded from the study since it is necessary that they engage with the topic and understand very well the purpose in order to play the designer role or to evaluate the design requirements. In order to select suitable participants, we suggested to ask volunteers what is their motivation to participate in the study during recruitment.

3.2 Description

In the first stage, participants must get an overview of the design process and explanation about interaction design techniques and be asked to perform three tasks. The tasks involve setting requirements for design - that is part of the second stage of the design process in which the problem (or opportunity) and the user are intensely analysed. Thus, the tasks consist of determining the application behaviour through a set of tools, which is prior to the ideation stage (when the requirements are transformed in solutions) of the design process. Participants must be instructed not to perform the tasks as a user, but as a designer.

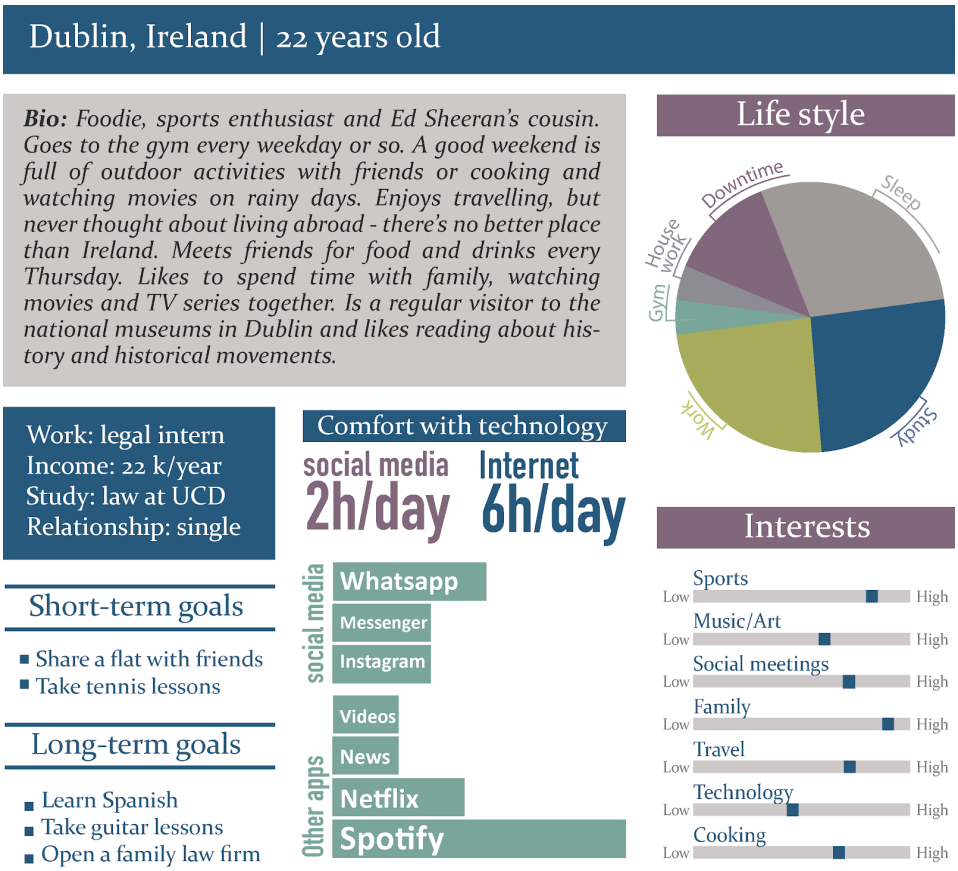
We have selected tools that are used in the stage of "defining the project", such as persona and scenario. Persona is a realistic description of the target user through a character, which is described with details (the user has a name, a lifestyle, a job, an income, expectations, and so on) instead of demographics only. For this method, it was defined that the persona should be gender-neutral; the gender cannot be revealed in the description of the persona or the scenario. For this method, the description of the persona should comprise a gender-neutral name or no name at all, lifestyle, activities and context (described in the scenario). Scenario is a narrative text that describes how the persona will use the application throughout a single day of the routine and how the application will deliver a solution or an experience to the user. The scope of the application must be explained through the description of the persona and the scenario.

In order to create a persona and a scenario it is necessary to study the users first and to understand their personalities and lifestyles. It is impossible to address all the individual differences; however, the persona must have traits and interests that are common to the majority of users of the group of interest. To make the persona seem like a real person, a mix of specific information extracted from different users can be added to the persona (a job, educational experience, pastimes) provided that the information added do not change the main characteristics of the users as a group (familiarity with technology, interests, motivations, behaviour, preferred applications). The persona helps designers to visualize the group they are designing for and the scenario helps designers to understand how the application will pervade the users' routine.

It is important that the participants playing the role of designers understand the users (personas), the context (scenario) and the service (in this case: to develop a dating experience) so they can gather requirements to develop the user experience. Participants

must not be told that gender is the key factor being analysed, but this can be revealed in a debriefing session. The gender dynamic of this research should not be mentioned as it could entice an inclination to consider gender as a design factor and compromise the spontaneity desired in order to determine the existence of gender biases. Decisions of each subject will be evaluated by other participants later. The following description of users and the context in which the interaction occurs is an example of what should be given to the participants:

Fig 1. Persona: the user description



Persona: Foodie, sports enthusiast and Ed Sheeran's cousin. Goes to the gym every weekday or so. A good weekend is full of outdoor activities with friends or cooking and watching movies on rainy days. Enjoys travelling, but never thought about living abroad - there's no better place than Ireland. Meets friends for food and drinks every Thursday. Likes to spend time with family, watching movies and TV series together. Is a regular visitor to the national museums in Dublin and likes reading about history and historical movements.

Scenario: "I met some friends for a drink the other evening. One of them told me about

their experience with this new dating app. I'd never used one before and so had some questions about how it works. My friend opened it there and then to show everybody the main features. I was curious to try it out but not so confident about installing it. My friend let me use it for a few minutes just to get a feel for it. I enjoyed the experience, so when I got back home, I decided to install the app on my own phone, to give it a go. I used the application for half an hour or so and matched with some interesting people. I've been using the app every night since, both to look for new people to match with, and to keep the conversation going with some of those I connected with previously. Everything really seemed to click with this one person, and we have a date coming up next Friday. I'm really looking forward to it."

3.3 Procedures

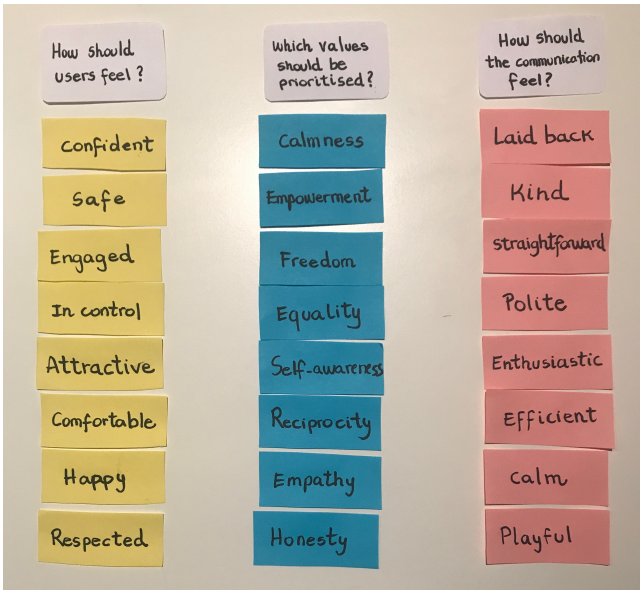
In the first stage, the session must begin with an explanation of some techniques used to set the requirements of a design project and to evaluate the solutions. We estimate that this introductory explanation will take around 20 minutes. It must be presented a generic persona and scenario for the dating experience, as the examples mentioned above.

For the first task, participants will receive three types of cards: value cards (which values should be prioritised?), the feeling cards (how should users feel?), and the engagement cards (how should the communication feel?). They will organize the cards in order of priority respecting each category. After ordering the cards, they will write down the order in their answer sheet, by ranking each card by order of importance. They will assign 1 to the most important card and 8 to the least important card. They will be told that they cannot repeat numbers and must check it after finished to make sure the replies recorded are intended. Rankings with numbers missing or duplicated will be disregarded. To make it more practical to the whole experiment, the cards will be also printed in a random order in a sheet, and, after they finish ordering the cards, they will copy the order in the sheet, assigning values varying from 1 to 8 (number of cards); they will assign 1 to most important card and 8 to the least important card[1]. This is a variation of card sorting technique. The original card sorting technique is used to define the navigation hierarchy of a website (Hudson, 2005) or an application, in which content and information architecture are established by users. Participants in this kind of user study receive the task to organize cards containing topics/categories into groups.

[1] This could also be managed with a computer-based response input system.

For the present method a variation of the technique was developed and the cards are already organized into three groups containing feelings, values and engagement approaches, to which participants assign values and put in order of importance within each group instead of organising by topic. The words in the cards are predefined considering findings of a previous study (Lopes & Vogel, 2017b) and the literature review on hookup culture. Participants will be asked to put cards in order of importance from their designer/developer perspective as in the simulation below (see figure 2).

Fig 2. Card sorting for: emotions, values and engagement



After that, participants should move to the next task and be asked to rate features that could define the experience on a dating application (see figure 3). The features help to predict the application behaviour, and consequently the way people would engage in a virtual environment. The design of the application behaviour highlights the important role of the application in the dialogue, as a second and active part. When designers assign a specific behaviour to the application, they assign a personality to it. That is the concept of the techsona coined by Bødker and Klokmose (Bødker & Klokmose, 2013) to explain the importance of defining an identity for the product or service as well as to the persona, so the interaction can be well planned and the actions and reactions designed. The idea of designing product's personality through feelings and values was explored by Jordan (Jordan, 2002) and it is useful for users to understand how to interact with the product (Mugge, Govers, & Schoormans, 2009). For example, if it is

defined that the application should be straightforward, then it is expected that it reacts and answer to users' requests that way (generating short and prompt responses, for example). The words in this task are also result of previous research (Lopes & Vogel, 2017b). Participants will rate the importance of some personal adjectives that ranges from 1 to 5 (see figure 3). Ambiguous responses (more than one choice or empty answer) will be disregarded.

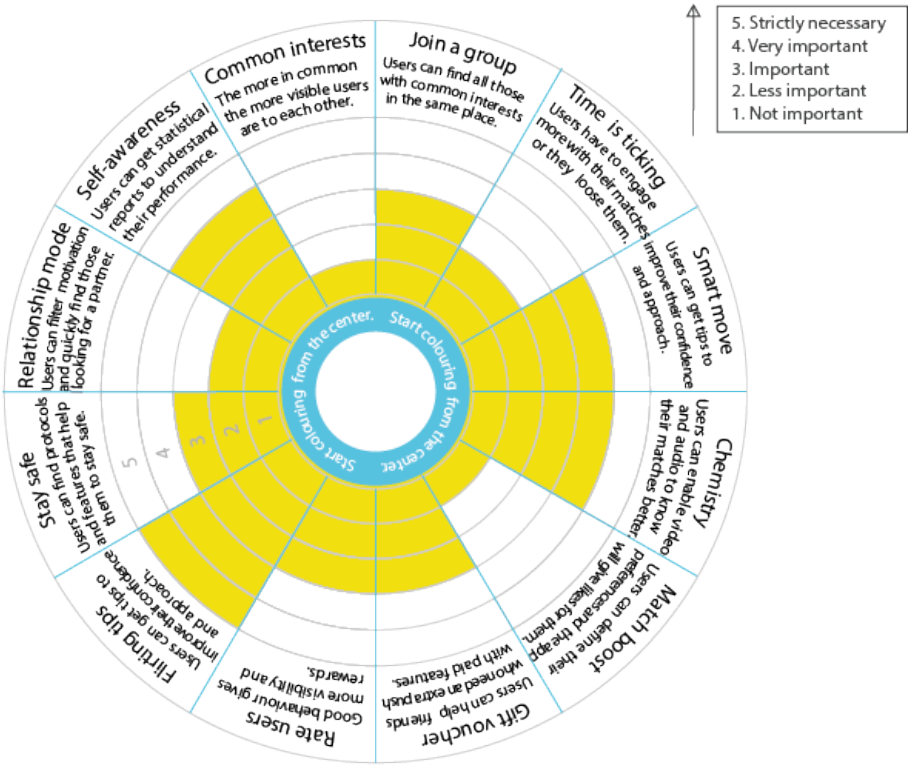
Fig 3. Personality setting table

	1	2	3	4	5
Discreet					
Caring					
Easygoing					
Cheerful					
Open-minded					
Honest					
Creative					
Calm					
Down to earth					
Polite					
Flexible					
Straightforward					
Commited					
Reliable					
Engaging					
Proactive					
Spontaneous					

1- Definitely NO
2- Probably NO
3- Neutral
4- Probably YES
5- Definitely YES

Finally, participants must be asked to complete a feature journey (see figure 4) designed for this study. The feature journey is a graphic wheel displaying possible features that would improve the interaction through the application. Participants will complete the feature journey by rating the importance of each proposed feature. They will begin to colour the centre first and there are 5 levels to cover. If one feature has a value 1 (not important) only the first and more central level related to that feature should be coloured. The participants will start filling the first third of the circle, with features related to the very first stage of the interaction with the interface, and they will assign a value to each of those features. For example, if time saving seems important to fulfil the users' needs, they will give a 4 or 5 and colour 4 or 5 layers, respectively. The same applies to the other two moments of the interaction. Answers that are not coloured as expected (discontinuous shading of a region) must be disregarded.

Fig 4. Feature journey



After they finish the tasks, they will be asked what they think the gender of the persona is. The answers will be added to their task sheets and analysed later as a source to determine whether there is a tendency to consider the regular user as man or woman. All the possible doubts that come up about the procedures will be answered during the design sessions; however, to preserve the gender neutrality of persona and scenario, no addition will be made to the descriptions of those.

In the second stage, a group of participants will judge the design decisions made by individuals within the first group and classify each designer as woman or man. Individuals within the second group will receive the same explanations about the tools for designing interactions and about the design process that the first group received. They will be given the persona and scenario that the first group used as a starting point to define the priorities of the design project. After that, they will try to guess the gender of each designer (participants of the first stage) by analysing their answers for the three tasks proposed. They should be told that the results of the experiment are very important to the research in order to encourage them to reflect before answering and not to give random answers. Furthermore, they can be told that the study will test their accuracy on

perceiving gender and that they will receive a feedback about that, if requested. The information produced by each participant of the first study will be transformed into graphs to help the second group to visualize the information. Each designer's decisions will be organised in a separate sheet containing only his/her choices for the first stage of the experiment. Finally, as another strategy to avoid random answers, they may be asked to justify the answers at the end. However, this must be explained at the beginning of the session, in order to stimulate them to reflect about their answers. The design solution of each participant of the first stage can be judged by every participant of the second stage if the first group is small (N≤60). With a larger number of participants in the first group, the second group should be split into two groups and the solutions produced by the first group divided by the groups for analysis. In this way, each participant of the second group will not analyse more than 60 design solutions.

3.5 Analysis

3.5.1 First stage analysis: In the first task, to each card position will be assigned a value from 1 to 8. Eight is the number of cards. The first card, in the top, will have a value of 8. The next from top to bottom will have a value of 7 and so on. The last card in the very bottom will have the value of 1. The same proceed to the second and third tasks that are both using the same rating system: variations ranging from very important to not important. The data collected can be organized in a spreadsheet and analysed using statistical package, such as R. Considering the nature of the data being analysed and the study goal which is compare two groups (male and female), the statistical test most suitable for this stage is the Mann-Whitney test. When there are significant differences between male and female users' answers then the P value is small, and when there are not the P value is large[2]. All tasks must be analysed separately. The null hypothesis is that there is no difference between the preference order (or the order value assigned to each word) produced by men and women.

3.5.2 Second stage analysis: In the second stage, we want to know if the participants' gender can be correctly guessed and if there is agreement between raters regarding what is feminine or masculine. Standard measures of accuracy rates can be used to assess correctness. In order to attach statistical significance, it is possible to aggregate guesses and assess agreement with the "gold standard" supplied by actual gender. It is also

[2] The confidence interval, and consequently the P value, should be determined after the number of participants is confirmed, since this number will affect the power of the study. It is desirable to have a power of 80% and a confidence of 0.10. However, as noted earlier, power greater than 60% can be adequate.

reasonable to assess agreement independent on convergence on accurate judgements. The inter-rater reliability test is useful providing information on how much raters agree. There are few tests that can be used, considering that there are multiple raters and that the data is nominal and dichotomous, but it is recommended using Krippendorff's alpha and Fleiss' kappa. Alternatively, the Chi-square test can be used in order to determine if there is statistical difference between female and male participants' guesses. In this study we want to know if the majority agree on the gender of the designers. In the case of neutrality, when raters disagree whether the designer is woman or man, we understand that the designer's solutions has both masculine and feminine qualities, that is, it can be considered as neutral. The results of agreement are crucial to verify if gender can be guessed since raters' agreement on a designer being woman or man can correspond or not to his/her gender. With these results it is also possible to verify if raters tend to assign one gender more than other gender and if there is a relation between the participant's gender and the assignment of gender through a percentage analysis.

4. DISCUSSION AND CONCLUSION

This is a method to detect gender biases in the development of mobile interfaces. This method can be adapted to different applications; it is not restricted to online dating realm. For that, the words in the first task, adjectives in the second and features in the third should be adapted to the application being tested. Possible flaws in this method can occur and we discuss below.

The presence of bias in the study is a risk that we should approach realistically. The first creations of the personas, for example, were considered biased by us. Only through deciding mixing two personas, the female and male, to create a third neutral persona that we could approach the neutrality. Yet, as a creation, even the neutral persona can be gendered and biased. The attempt to create a method to study gender biases can be biased by nature, but we believe that the reflections on this topic can mitigate possible biases.

The understanding of the method by participants will depend on the professionalism during the explanation and conduction of the user study by researchers. It is very important that participants fully understand what they need to do and what is expected from them. Otherwise the results can be inconsistent. Apart, it is crucial that all the

participants get the same information and instructions during the study.

The design tools selected for the experiment can be insufficient to detect gender biases during the design of the experience for a particular application, and this can change from case to case. It can be efficient in order to study dating applications but maybe it is not to evaluate one for money transactions. In this case, the method must be adapted as far as possible to attend to other application specificities.

It is considered in this work that one doesn't need to be a professional designer to play the role of a designer in the user study, since the focus is the difference produced by two groups in the same population. However, it is necessary to consider that there will be a difference from population to population and that the ideal situation is to test with professional designers, who are more used to involve users in their developments.

The validity of the method would tell how well it measures gender biases it is supposed to measure. A way to demonstrate construct validity of the method is by comparing the method with other methods that measure gender biases in design in order to determine the correlation between the two measures. A high correlation would indicate that the method is efficient in detecting gender biases. However, a similar method of assessing gender bias in design on the basis of perceived gender is not of our knowledge available for comparison. We relied on the research on design process and design methodology, pilot studies and statistical tests to create the method but we have no means to compare it with a similar method in order to determine whether this is the best approach to measure gender biases.

The proposed method is designed in order to determine whether there are gender biases in the design of applications or not. The user study method described above is the first step towards understanding the connection of gender and production of design. In the case of a positive result (there are gender biases or perceived gender difference), it would be necessary to confirm the results or to investigate other implications of gender biases in a following experiment. In the case of a negative outcome, in turn, it would be necessary to investigate if there is a problem in the design of the experiment (what would implicate a redesign of the experiment) or consider that the frustration of women and the high rate of sexist behaviour in the first experiment do not fall into the first stages of design of the application or that gender biases is also ingrained in the design process of female designers. In the case of a negative outcome, other directions should be taken: the failure to prove gender biases should turn into an attempt to prove the

absence of gender biases.

The research is not at odds with gender diversity, nonetheless, we consider the male-female polarity as a starting point to understand gendered expressions in applications and biases in design, due to the social impact of the historical division of sex roles on gender equality.

Acknowledgements: The first author would like to acknowledge CAPES foundation for funding this research. This research is also supported by Science Foundation Ireland (SFI) through the CNGL Programme (Grant 12/CE/I2267 and 13/RC/2106) in the ADAPT Centre (www.adaptcentre.ie).

REFERENCES

Bødker, S., & Klokmose, C. (2013). From persona to techsona. In P. Kotz´e, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Eds.), *Ifip conference on human-computer interaction,* pp 342-349. Springer.

Buckley, C. (1986). Toward a feminist analysis of women and design. *Design Issues*, 3(2), 3-14.

Blandford, A., Cox, A. L., & Cairns, P. (2008). Controlled experiments. In Cairns, P., & Cox, A. L. (Eds.). *Research methods for human-computer interaction*, Vol. 12, 1-16. Cambridge University Press.

Creswell, J. (2014). *Research design: Qualitative, quantitative, and mixed methods approach*. Sage.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM *Transactions on Infor- mation Systems* (TOIS), 14(3), 330-347.

Hudson, W. (2005). Playing your cards right: getting the most from card sorting for navigation design. *Interactions*, 12(5), 56-58.

Jordan, P. W. (2002). Pleasure with products: Beyond usability. In W. S. Green & P. W. Jordan (Eds.), *Pleasure With Products: Beyond Usability,* pp. 19-47. Taylor and Francis.

Kraemer, H. C., & Blasey, C. (2015). *How many subjects? Statistical power analysis in research.* Sage Publications.

Lopes, M. R., & Vogel, C. (2017a). Gender bias on tinder: Transforming an exploratory qualitative survey into statistical data for contextualized interpretation. In A. Costa, L. Reis, F. Souza, & A. Moreira (Eds.), *Computer supported qualitative research*, Vol. 621, 225-236. Springer.

Lopes, M. R., & Vogel, C. (2017b). *Women's perspective on using tinder: a user study of gender dynamics in a mobile device application*. In Proceedings of the 35th ACM

International Conference on the Design of Communication (SIGDOC'17).

Moss, G., & Colman, A. M. (2001). Choices and preferences: Experiments on gender differences. *Journal of Brand Management*, 9(2), 89-98.

Moss, G., & Gunn, R. (2007). Gender differences in website design: Implications for education. *Journal of Systemics, Cybernetics and Informatics*, 5(6), 38-43.

Moss, G., Gunn, R., & Heller, J. (2006). Some men like it black, some women like it pink: consumer implications of differences in male and female website design. *Journal of Consumer Behaviour*, 5(4), 328-341.

Mugge, R., Govers, P. C. M., & Schoormans, J. (2009). The development and testing of a product personality scale. *Design Studies*, 30(3), 287-302.

Nysveen, H., Pedersen, P. E., & Thorbjørnsen, H. (2005). Explaining intention to use mobile chat services: moderating effects of gender. *Journal of Consumer Marketing*, 22(5), 247-256.

Oudshoorn, N., Rommes, E., & Stienstra, M. (2004). Configuring the user as everybody: Gender and design cultures in information and communication technologies. *Science, Technology Human Values*, 29(1), 30-63.

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction design: Beyond human-computer interaction. 3rd edition.* John Wiley Sons.

Sauro, J., & Lewis, J. R. (2012). *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.

Simon, H. A. (1969). *The sciences of the artificial.* The MIT Press.

Simon, S. J., & Peppas, S. C. (2005). Attitudes towards product website design: a study of the effects of gender. *Journal of Marketing Communications*, 11(2), 129–144.

Stonewall, J., & Dorneich, M. C. (2016). *A process for evaluating the gender and professionalism of web design elements.* In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 60, 750-754.

Williams, G. (2014). Are you sure your software is gender-neutral? *Interactions,* 21(1), 36-39.

Xue, L., & Yen, C. C. (2007). Towards female preferences in design - a pilot study. *International Journal of Design*, 1(3), 11-27.

How to cite this article:

Lopes, M. & Vogel, C. (2019). Is Your Application Gender Biased? *International Journal of Marketing, Communication and New Media*. Vol. 7, Nº12, 103-122.