Gender Effects in Mobile Application Development

Milena Ribeiro Lopes
Department of Computer Science and Statistics
Trinity College Dublin
Dublin, Ireland
lopesm@scss.tcd.ie

Carl Vogel
Department of Computer Science and Statistics
Trinity College Dublin
Dublin, Ireland
vogel@scss.tcd.ie

Abstract—Growing interest in the impact of gender on technology includes concerns with regard to gender bias and its implication in the user experience. Past explorations of gender and design revealed significant differences between female and male users with regard to the perception and the acceptance of dating apps. A novel study was designed to investigate whether gender differences in the user experience could be caused by gender bias in mobile app development. Significant differences in the definition of design priorities emerged with regard to the participants’ gender and the persona’s perceived gender.

Index Terms—gender bias, design, mobile applications

I. INTRODUCTION

The advance of gender studies and growing public concern to assure gender equality brought to light aspects of gender differences in the use and development of technology. Gender may be understood as clusters of perceptions, attitudes, conventions and behaviours held individually and socially in relation to biological sex distinctions. As cluster concepts, they do not reduce to binary oppositions; nonetheless, two prominent distinctions are labelled “man” and “woman”. Some studies [1]–[3] point out barriers in the inclusion of the female perspective during the design process and reveal gender differences in users’ preferences and users’ creative outcome that are disregarded due to the predominance of a male perspective in the tech industry [4], [5] caused by deep rooted gender dynamics.

Causes of gendering in artefacts have been theorised [6], from which two are connected to problematic understanding of users: I-methodology [7] and the inscriptions of gender stereotypes [2]. Those issues are caused by the lack of user research [1], [2], [8], the misrepresentation of female users during user research [9], and bias in development [3].

Technological developments are assumed to be gender biased, for women are not well represented in design teams and in user research [4], [5], [10], designers are likely to embed developments with their own opinions [2], [6], [7], [11]–[14], women have low influence in teams’ discussions [1], [9], [15], and for a speculative unconscious tendency to consider men to be the regular user of technological artefacts [16], [17].

Strategies have been suggested to counteract bias and gender issues in the design process, such as participatory design [18], value sensitive design [16], reflective design [19], undoing design [20] and de-gendering design [6]. However, those strategies mostly provide guidelines to improve sensitivity and reflection on gender issues during the design process. Gender bias in development has not been empirically demonstrated.

A previous study with female users of Tinder [21] revealed that 57% of the interviewees uninstalled the application due to a negative experience and 70% experienced offensive gender dynamics. This suggests that the unsatisfactory female user experience and the disregard of their needs may had been caused by gender bias in design. Support for this hypothesis may be found in studies that point out gender differences regarding users’ preferences in design [22]–[24] and designers’ preferences in graphical design [25] and web design [24]. Given the same user, and the same service, will men and women reveal any statistically significant difference in the definition of the project’s requirements? This research question cannot be answered by previous studies since the user is not controlled and no statistical tests support the significance of differences. This hypothesis may be tested empirically. Moreover, some argue that there is a proclivity to consider the user as a man [16], [17], and we also test this assumption. To investigate whether gender has a measurable impact in the development of mobile applications, we use a variation of the method we developed to investigate gender bias in design [26]. The main idea is that participants use simple design tools to establish requirements for the design of a fictitious online dating application.

This research is not at odds with gender inclusivity and the wide gender spectrum. However, we narrow our investigation to a female-male dichotomy that brings historical weight to the shaping of social interactions and gender roles. We assume that historical patterns of behaviour may be reproduced in new technologies and incited by the user-device interaction, negatively affecting women’s experience. Thus, this study of gender in mobile application design adopts a male-female binary approach underpinned by values of gender equality, equity and women’s empowerment.

II. RESEARCH DESIGN

Participants individually completed tasks that aimed to set requirements for the design of a fictitious online dating application. They were presented with the description of

1 The study was approved by the Research Ethics Committee of the School of Computer Science and Statistics (SCSS) at Trinity College Dublin before being conducted.
the user as a “persona” and of the context and service as a “scenario” in the first page, and three tasks (with respective task explanations) in pages that followed. 56 undergraduates studying computer science at Trinity College Dublin (TCD) volunteered. University students studying computer science were deemed appropriate because their subject inherently involves designing applications. The sample was composed of 28 women and 28 men, and their age mean was 19 years old. Participants had the option to indicate their gender as “female”, “male” or “other”, as suggested for gender inclusivity in surveys [27] but each indicated either “female” or “male”.

A. Gender neutrality

Participants were not told that gender differences were being analysed. However, a debriefing email with the study results were offered to participants who indicated interest. A neutral persona was developed with the help of an HCI research team at Trinity College Dublin and a gender expert to avoid marking the texts with gender stereotypes. In order to create a neutral persona, two personas (male and female) were mixed into a single one. The first versions of a neutral persona and neutral scenario revealed stereotyped ideas and gendered style of writing. They were redesigned until the team was not able to say whether the text described a male or female user.

The persona was determined to be gender neutral in the description of activities and to have no name, since even a neutral name could appear to be more feminine or masculine for participants. The scenario was adapted to a first person narrative in order to incorporate gender-neutral quality. With this approach, it was possible to verify whether participants would assign gender to the target user. The scenario of Figure 1 and persona of Figure 2 were given to participants.

“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.”

Fig. 1. Scenario for task 1

B. Procedures

Each participant received the same explanations about the techniques and instructions about how to perform the tasks. In the first task (see Figure 3), participants were asked to priority sort “feelings”, “values” (which values should be supported?), and aspects of “communication” (how should the communication feel?). Participants were asked to assign values varying from 1 to 8 (number of words in each group) in order of importance – 1 to the most important and 8 to the least important. In the second task, participants were asked to rate personality traits that would define the experience on a dating application (see figure 4). The personality helps to shape the application’s behaviour, and consequently the way people engage with the application in a virtual environment. 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 Klokmos [28] to explain the importance of defining an identity for the product, so the interaction can be well planned and the actions and reactions designed. The idea of designing a product’s personality through feelings and values was explored by Jordan [29] and it is useful for users to understand how to interact with the product [30]. In our work, to assess personality, participants had to rate the importance of some personal adjectival expressions through a Likert scale ranging from one (definitely not important) to five (definitely...
The items rated are detailed in Figure 4.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Join a group</td>
<td>Users can find all those with one common interest in the same place.</td>
</tr>
<tr>
<td>Relationship mode</td>
<td>Users can filter by motivation and quickly find those looking for a relationship.</td>
</tr>
<tr>
<td>Time is ticking</td>
<td>Users have to engage more with their matches or lose them due to inactivity.</td>
</tr>
<tr>
<td>Gift voucher</td>
<td>Users can help a friend who needs an extra push by giving them a voucher for paid features.</td>
</tr>
<tr>
<td>Stay safe</td>
<td>Users can find protocols and features that help them to stay safe during encounters.</td>
</tr>
<tr>
<td>Self-awareness</td>
<td>Users can get statistical reports to better understand what they’re doing wrong and well.</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Users can enable video and voice features to get to know their matches better before meeting.</td>
</tr>
<tr>
<td>Match boost</td>
<td>Users can define their preferences and let the system give likes on their behalf.</td>
</tr>
<tr>
<td>Smart move</td>
<td>Users can get tips to improve conversation and keep matches interested in them.</td>
</tr>
<tr>
<td>Flirting tips</td>
<td>Users can get tips to improve their confidence and approach.</td>
</tr>
<tr>
<td>Rate users</td>
<td>Good behaviour gets more visibility; bad behaviour gets penalties.</td>
</tr>
<tr>
<td>Common Interests</td>
<td>The more they have in common, the more visible their profiles will be to each other.</td>
</tr>
</tbody>
</table>

Fig. 5. Task 3: Feature rating list.

Finally, participants were asked to rate putative system features through a Likert scale ranging from one (definitely not important) to five (definitely important): Figure 5 shows the features available for rating. Those feature ideas derive from results reported in previous studies with Tinder users [21], [31], which made it possible to identify dating application features that women and men indicated as problematic or as desirable.

After they finished the design tasks, participants were asked to indicate the persona’s perceived gender and the level of difficulty in designing without taking into account their own preferences. With respect to perceived gender, it was posed an open-ended question. Since the persona was contrived in a gender-neutral fashion, indications of gender would reveal the participant’s perception/projection of the persona’s gender. With respect to difficulty in designing for preferences other than the designer’s own, three options were available: “easy”, “medium” and “difficult”. Since the persona was contrived in a gender-neutral fashion, indications of gender would reveal the participant’s perception/projection of the persona’s gender.

C. Analysis

As the goal of this study was to identify measurable gender differences in the design of mobile applications, the analysis of the data set is composed of four different tests, which were defined before recruiting participants and conducting the study. The data were tested for independence in four different ways: differences by participant’s gender, differences by persona’s perceived gender, differences by participant’s gender for female persona and differences by participant’s gender for male persona. In each of these cross classifications of responses, median responses to items were examined. In this, differences that arose were analysed within all the tasks, together, not each task in isolation. Thus, Section III (FINDINGS) is not organised by task, but by the cross-classifications indicated above. A focus was retained on the effects that were statistically significant using a Wilcoxon test on item response, conditioned on the distinction at stake in each case. As a non-parametric alternative to a t-test, this test does not compare medians or means directly, but rather the groups’ distribution. However, a difference in the distribution is often associated with a difference between the medians of two groups, which is why the medians are displayed in the tables. With the medians it is possible to visualise which group assigned priority to each factor analysed. In order to analyse factors that revealed significant differences but the same median for the two groups, a closer look in the distribution helps to visualise the difference, indicated in the columns with headings “priority by/for” in the tables. This study’s focus in gender-relative differences in prioritisation in the responses – see Table I for a summary of

2The importance ordering here is the reverse of that in the first task.

3Bonferroni corrections were not applied, since only the four tests described were used, which were determined before the study was conducted and are not a result of statistical fishing.
available response categories across the design tasks.

<table>
<thead>
<tr>
<th>Task</th>
<th>Greatest</th>
<th>Least priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: feeling qualities</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>T1: value qualities</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>T1: communication qualities</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>T2: techsoma personality qualities</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>T3: feature qualities</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

### III. Results

Considering the nature of the data analysed, the nonparametric Mann-Whitney Test (Wilcoxon) was used to detect significant differences between the design requirements pointed out by men and women, as measured in the ranks and scores assigned to design factors. Each factor was analysed separately. The null hypothesis was that there was no difference in the values assigned by men and women. In discussing the results, we emphasise mainly the effects that are significant ($p < 0.05$).

#### A. Perception of the persona’s gender and personal bias

The first analysis refers to the perception of gender-neutral personas. An open-ended question was posed to enable responses within a broader gender spectrum, but all participants indicated either “female” or “male”. Among women, 13 believed the persona was a woman while 15 believed the persona was a man. Among men, 16 believed the persona was a woman and 12 believed the persona was a man$^4$ (see Table II). One may not reject the null hypothesis of no interaction between participant gender and perceived persona gender.

Participants were asked to evaluate if it was easy or difficult to make design decisions for a persona without taking their own personal beliefs into account. Among female participants, 6 found it easy, 5 found it neither easy nor difficult, and 17 found it difficult. Among male participants, 10 found it easy, 1 found it neither easy nor difficult, and 17 found it difficult$^5$ (see Table III). Therefore, it is not safe to reject the null hypothesis that there is no interaction between participant gender and reported difficulty in designing for a self-dissociated persona.

#### B. Test 1: Differences by participants’ gender

When comparing the values assigned by men and women, men tended to prioritise “calm” communication and the “relationship mode” feature more than women, and women to embed the application with a “honest” value and a “honest” behaviour (see Table IV) more than men.

*C. Test 2: Differences by persona’s perceived gender*

To determine whether participants (both men and women) tended to assign priorities differently depending on target user gender, the same test was carried out, comparing persona’s rather than the participant’s gender. The results show that participants tended to think that feeling “happy”, and that enhancing “empathy” and “reciprocity” would be more important to male users, while enhancing “privacy” and “equality”, and imbuing the application with a “discreet” and “caring” personality would be more important to female users (see table V).

#### D. Test 3: Differences by participant’s gender—female persona

With regard to decisions made for a female persona, men were more likely than women to prioritise “calm” communication and to embed the application with the “relationship” filter (see table VI) when designing for a female persona.

#### E. Test 4: Differences by participant’s gender—male persona

When designing for a perceived male persona, men were more likely to prioritise a “calm” communication through the application and women to embed the application with a “cheerful” and “down to earth” personality (see table VII). Recall again from Section II-C (Analysis) that the tables report the median response to items, by category, and that the Wilcoxon test is a comparison of response distribution, by

---

$^4$ Thus, a $\chi^2$ test of the independence of participant gender and judgements of persona does not reveal a significant difference within the computer science group: $\chi^2 = 0.28668, df = 1, p = 0.5927$.

$^5$ The interaction between participant gender and perceived difficulty is not significant in the computer science group: $\chi^2 = 3.6667, df = 1, p = 0.1599$. 

---
An aim of this study was to ascertain any differences in assigning priorities in design according to participants’ gender, to progress the discussion of gender bias in design. Therefore, we next attempt to contextualise these results based on findings of previous studies [21], [31] and the wider literature.

Considering participants’ gender, women revealed a tendency to imbue the application with the value of “honesty” and an “honest” personality, which suggests a female claim for transparency in interaction through dating apps within that group. Men were more likely than women to prioritise a “calm” communication and to provide the application with a “relationship mode” feature in which users can easily find people who are looking for a relationship. It is natural to speculate about why men prioritised the “relationship mode” feature. A closer look in each persona reveals that men were more likely than women to prioritise a “calm” communication when designing for both male and female personas and to prioritise “relationship mode” when designing for female personas. The “relationship mode” feature being preferred by men for a female persona (Table VI) but not a male persona (Table VII), reveals a masculine perception of women’s needs. The study on gender differences in the experience of Tinder [31], however, revealed that men and women were equally likely to look for a relationship on dating applications, although they both tended to think that women are usually looking for relationships while men are looking for sex. This reveals a gender stereotype which does not entirely correspond to reality. According to that study, men were statistically more likely to report “easy dates” as a benefit, but as open as women to engage in new relationships.

Participants were more likely to prioritise the feeling of “happiness” and to provide the application with “empathy” and “reciprocity” values when designing for a male persona. This might reveal that both men and women thought a male user would need more responsive interactions to be engaged than women and that feeling happy would be more relevant for the application’s adoption by men when compared to women. With regard to female persona, men and women both tended to embed the application with “privacy” and “equality” values, and a “discreet” and “caring” personality. This might also indicate a gender stereotype, since the results of previous studies [21], [31] indicated that happiness and empathy could be critical for women, that offensive behaviour was the most reported negative aspect of their experience of ODS (with regard to gender dynamics), and more than a half uninstalled Tinder for having an overall negative experience, while men were more likely to use Tinder for longer, even though boredom was the most frequent “negative” aspect of Tinder among men. Moreover, imbuing the application with the value of “privacy” and a “discreet” personality may indicate a taboo around female sexuality, a need to “hide” subscription to a dating application.

Women tended to assign a “cheerful” and “down to earth” personality to the application when designing for a male persona but not for a female persona, suggesting they have a perception that men seek a joyful but informal interaction. That impression of women is in accordance with the study on women’s experience of Tinder [21] in which interviewed

**TABLE V**

<table>
<thead>
<tr>
<th>Task</th>
<th>Factor</th>
<th>Female persona</th>
<th>Male persona</th>
<th>Prioritised by</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: feeling</td>
<td>Happy*</td>
<td>6</td>
<td>4</td>
<td>Male persona</td>
</tr>
<tr>
<td>T1: value</td>
<td>Privacy*</td>
<td>2</td>
<td>3</td>
<td>Female persona</td>
</tr>
<tr>
<td>T1: value</td>
<td>Empathy*</td>
<td>6</td>
<td>4</td>
<td>Male persona</td>
</tr>
<tr>
<td>T1: value</td>
<td>Reciprocity***</td>
<td>7</td>
<td>5</td>
<td>Male persona</td>
</tr>
<tr>
<td>T1: value</td>
<td>Equality**</td>
<td>3</td>
<td>5</td>
<td>Female persona</td>
</tr>
<tr>
<td>T2: personality</td>
<td>Discreet*</td>
<td>4</td>
<td>3</td>
<td>Female persona</td>
</tr>
<tr>
<td>T2: personality</td>
<td>Caring**</td>
<td>4</td>
<td>3</td>
<td>Female persona</td>
</tr>
</tbody>
</table>

* – $p < 0.05; ** – $p < 0.01; *** – $p < 0.001.

**TABLE VI**

<table>
<thead>
<tr>
<th>Task</th>
<th>Factor</th>
<th>Women</th>
<th>Men</th>
<th>Prioritised by</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: comm. Calm*</td>
<td>5</td>
<td>3</td>
<td>Men</td>
<td></td>
</tr>
<tr>
<td>T3: feature Relationship*</td>
<td>3</td>
<td>4</td>
<td>Men</td>
<td></td>
</tr>
</tbody>
</table>

* – $p < 0.05; ** – $p < 0.01; *** – $p < 0.001.

**TABLE VII**

<table>
<thead>
<tr>
<th>Task</th>
<th>Factor</th>
<th>Women</th>
<th>Men</th>
<th>Prioritised by</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: comm. Calm***</td>
<td>7</td>
<td>4</td>
<td>Men</td>
<td></td>
</tr>
<tr>
<td>T2: personality Cheerful*</td>
<td>4</td>
<td>4</td>
<td>Women</td>
<td></td>
</tr>
<tr>
<td>T2: personality Down to earth*</td>
<td>4</td>
<td>3</td>
<td>Women</td>
<td></td>
</tr>
</tbody>
</table>

* – $p < 0.05; ** – $p < 0.01; *** – $p < 0.001.

**IV. Discussion**

Female and male participants assigned female and male gender to the neutral persona in balanced way, showing no consistent tendency in assigning gender to the user. Nevertheless, there is a strong indicator of participants’ biases towards their own preferences in decision making, since 60% of participants found it difficult to make decisions for a user without taking their own preferences and experiences into account, even when the user is defined. Only 29% of participants found it easy to avoid personal bias. Thus, the assumption that there can be gender bias in the design process even when a user-centred design is approached [1] has mitigated support. A tendency to design for oneself [2], [7], [13], [14] is evident.

Note that if the same percentage categories of replies (see Table II) arose for ten times the number of participants, the tendency towards opposite-gender perception would be statistically significant ($\chi^2 = 6.0148, p = 0.01419$).
women revealed their perception that men are looking more for sex and less for romantic relationships.

V. CONTRIBUTIONS, CHALLENGES, AND CAVEATS

Many assumptions regarding gender biases in technological developments were posed in the past decades, indicating that gender stereotypes could be ingrained in products and services [32], [33] and that male domination in technology affects the design process [4], [5]. We investigated whether there are gender differences in design decisions that could support such claims.

This study revealed gender biases in design. A positive outcome might be the creation of tools to improve gender awareness in design process and methods for counteracting the inscription of gender stereotypes. We introduced the use of a gender-neutral persona, which can be explored in future research to promote attention to gender needs in a balanced way in the design process. Our findings may not generalise to all stages of the design process. The study is limited in the number of participants and possibly traits unique to that sample and to the application designed. However, we think that comparable differences would also emerge in a larger, more heterogeneous group and also in other design stages.

ACKNOWLEDGMENT

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 CNG Programme (Grant 12/CE/I2267 and 13/RC/2106) in the ADAPT Centre (www.adaptcentre.ie).

REFERENCES