Password Blur - A Usable and Secure Password Input Method for Mobile Devices

Zhe Li, Leonhard Mertl, Alice Nguyen, Mario Schneller
Ludwig-Maximilians-Universität München
{li.zhe,l.mertl,alice.nguyen,mario.schneller}@campus.lmu.de

Figure 1. We explored different filter types to improve the usability and security of the common asterisk-based authentication method on mobile devices (from left to right: Color-Halftone, Crystallize, Gaussian Blur and Mosaic)

ABSTRACT

Humans have the ability to understand degraded versions of images, provided they know about the original version of the image. The main focus of this project is to study how this concept can be applied on textual passwords. As a first step we investigated the topic generally by asking 163 people via an online survey whether they would recognize text elements that have been visually degraded by the Gaussian Blur, Color Halftone, Crystallize and Mosaic filters. As a second step we brought this concept to the context of password input. By conducting an in-depth user study with 25 participants we collected significant quantitative and qualitative feedback. The study contained two practical tasks that investigated the usability and security of the new authentication method and compared it to. traditional alternatives, which are Android/iOS default password entry methods and plain text. The hands-on tasks were followed by a qualitative interview. We found, that our filter-based authentication approach is a distinctly more secure alternative to traditional password input methods with its usability performing at a comparably high level. To achieve optimum results we recommend to apply this novel input technique by using the Crystallize filter which showed most promising results in terms of usability and security. As a next step we propose a long-term field-study to get more detailed insights on human perception of different filter gradations and evaluating this method in an extended scope.

Author Keywords

Password; Filter; Distortion; Shoulder Surfing; Security; Authentication

INTRODUCTION

There are more than 2.8 billion internet users world wide [13] which use on average 25 different accounts with 8 log-ins online per day [7]. But even with new authentication technologies like finger print sensors or iris scanners, passwords are still the major approach for data protection and privacy in the web [7]. Although passwords are not getting obsolete in the near future, they are far from being perfectly secure.

Due to the error sensitivity of user inputs on smartphones, the typed characters of a password are shown for a few milliseconds before turning into asterisks. As a consequence basically everybody's private data is at risk to be obtained via shoulder surfing attacks [5]. Shoulder surfing is by definition ”the practice of spying on the user of a cash-dispensing machine or other electronic devices in order to obtain their personal identification number, password, etc.” [1]. Besides the security related issues, there are also usability barriers related to traditional authentication methods. Using asterisks to hide passwords from third-parties has the consequence that users are not able to detect spelling mistakes and therefore have to delete the whole input when they find themselves confronted with incorrect entries.

To challenge this state of the art password input and enhance usability and security for authentication on mobile devices we propose a novel method using image filters to distort the textual password. We investigate if users are able to detect spelling mistakes in their distorted password input and if the distorted passwords can resist shoulder surfing attacks by third-parties. Additionally we evaluate the trade-off between usability and security when applying this method.

RELATED WORK

The idea using the distortion filters on textual password input to exploit the human sense of cognition is based on previous research. It was found that humans are able to pick a known image out of a larger set, even if they only see a degraded version of the original [3, 8, 11]. Harada et al. [9] and Hayashi et al. [10] used this as a basis to develop graphical user authentication (GUA) mechanisms. Hayashi et al.'s approach is called Use Your Illusion. Images for the GUA are distorted by an Oil Paint filter, which eliminates most details in the images, but preserves some details like colours and rough shapes. The user has to select his chosen image portfolio out of other distorted images to authenticate. As a result, the usability of graphical password schemes is maintained and become resilient to guessing attacks. However, the authors point out that the selection of the best filter for this mechanism requires further investigations. Von Zezschwitz et al. [15] used the same concept of human recognition for browsing photos securely. Three filter types (Oil Paint, Crystallize, Pixelate (Mosaic)) with three filter strength (none, medium, high) were examined and successfully proven to conceal the pictures' content.

Using the findings described above, we decided to use the Crystallize and Pixelate filter for our study. The Oil Paint filter showed to be impractical, because when being applied on text, the distorted characters were still too predictable. As a compensation we decided to add Gaussian Blur that also relies on the concept of blur, but in a more extensive way that allows degrading text as needed for our study. To get even broader results we decided to introduce a fourth filter rooted in the family of the pixelate filters [2]. We chose Color Halftone, a filter that divides the image into rectangles and replaces each rectangle with a circle and is therefore a different jet related filter to Crystallize and Pixelate.

METHODOLOGY

Quantitative Online Survey

Our first study in form of an online survey was not designed to provide us with significant data, but rather to give us feedback from many participants in a short amount of time. Our goal was to get a first feeling for the different filters, filter strengths and overall performance of the participants in security and usability related questions. First and foremost, based on the previous work from Hayashi et al. [10] using degraded versions of pictures for the authentication process, we wanted to know if there is a difference in the performance of the participants if they know what a degraded word should mean in contrast an for them unknown word. Furthermore the study should clarify whether there are other factors like password length, random character strings, filter types and filter strengths that could have an impact on the perceptibility of distorted text.

Survey Design

In the survey we tested four different filters (Color Halftone, Crystallize, Gaussian Blur and Mosaic) with two strengths each (light and strong) and six questions per condition. As we wanted the survey to be doable in about 10 minutes, each participant had to answer 20 random questions out of a pool of 48 in total. We differentiated between usability and security related questions. The usability related questions should clarify whether the participants were able to recognize words known to them, what we call Owner Guessability. In the security related questions we tested the Guessability of (to the user) unknown words and character strings, the so called Third Person Guessability.

Results and Discussion

In total our survey had 163 participants with each question answered on average more than 67 times. As mentioned before even though there is no significant data, we could see certain trends, which were addressed in a follow-up study. As seen in figure 2, there seems to be a difference in the Owner Guessability as well as the Third Person Guessability in correlation with both, the filter type and strength. Since we were looking for a filter that has a high usability as well as a high security we chose the filter strengths with the highest Owner Guessability and at the same time lowest Third Person Guessability for our second study. Moreover the survey data shows clearly that there is a difference when the user knows which word should be shown as the Owner Guessability performed overall better than the Third Person Guessability. We could also see in the data, that there is a big difference in the perceptibility when the word given has an actual meaning rather than random characters. Based on this first impression a follow-up study was planned, which tested the filters in a real world scenario and aimed for significant data.

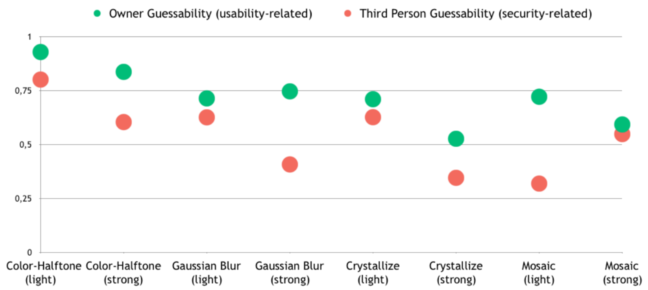


Figure 2. Average Levenshtein distance ratio depending on the word length for each condition: Color-Halftone had the highest usability, while Crystallize and Mosaic were the most resistant ones to observations by third persons

Quantitative and Qualitative User Study

With this follow up user study we examined the performance of filter-based password input in terms of usability and security in comparison to traditional plain text and asterisk-based input methods.

Study Design

25 participants, aged between 18 and 32, were recruited via social networks and personal invitation. To allow a more natural setting the participants performed all input related tasks on their own smartphone. Therefore we provided them with a server that had our prototype running and was accessible through the university's wifi.

We used a repeated-measures ANOVA for analysing the data. The study contained 2 practical tasks that collected quantitative data on usability and security and a concluding survey that aimed for qualitative data. We used again (1) Color Halftone, (2) Crystallize, (3) Gaussian Blur and (4) Mosaic filters. With the added traditional methods Plain Text and Asterisks we counted 6 conditions in total. For each filter we presented the participant with three different password types: short words, random combinations containing special characters and password phrases. Each of the password types had 2 variants. To eliminate learning effects and to counterbalance the combination of passwords and filters we generated a unique, 36 item long password list for each task and shifted the order of filters for every participant by using a 6×6 Latin square [16].

Task 1 that investigated the usability, was split into 2 parts. First, the participants had to enter a given password into an input-field and secondly insert the character x at a specific position. For the first part of task 1, the dependent variables were Levenshtein distance (similarity between entry and the actual password) [4, 12] and entry time. For the second part the dependent variables were editing accuracy (similarity between user entry and the correct position of the x) and editing time. For both parts the independent variables were filter type, password type and entry number.

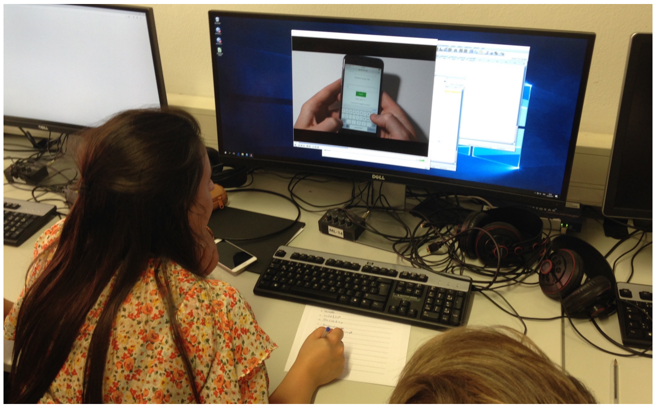


Figure 3. Second part of the study: The participants were asked to look at a recorded video and try to obtain the entered password

Task 2 of the study was considering the security of the different filters and traditional password input methods in the context of shoulder surfing. The participants were asked to act as a shoulder surfer and try to obtain passwords by looking at a recorded video, where a password input procedure was shown in close-up (see figure 3). The participants noted the observed password with pen and paper and transcribed it to a digital form at a later stage to avoid transcription mistakes. In sum the participants were trying to exploit the password inputs of 36 video clips (see 4). The dependent variable was the Levenshtein distance (similarity between noted password and the actual password). The independent variables were filter type, password type and entry number.

Results and Discussion - Usability Task

For the accuracy of the password entry, represented by the Levenshtein distance, no significant effect between the different filters and the password type were found. So we didn't find any evidence for an influence of filters or passwords types on the accuracy of password entries.

We found a significant main effect of password type on entry time (F(2, 48) = 236.723, $p < 0.0005$): The more complicated the password is the longer the users take to enter a password. This expected outcome shows that the study design was set up correctly. Additionally, a significance between the entry time and entry number was found (F(1, 24) = 63.690, $p < 0.0005$), which argues for a learning effect. The test persons performed faster in the second entry for each password type. Based on this results, one could say, that the underlying authentication method is fast to learn, wherefore easy to use.

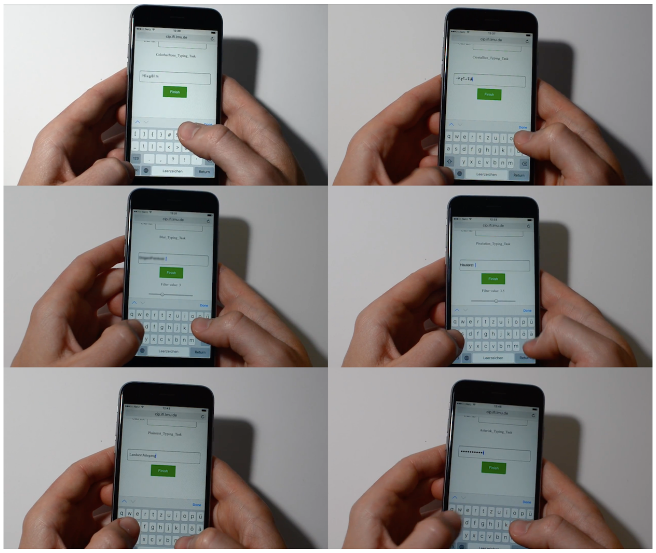
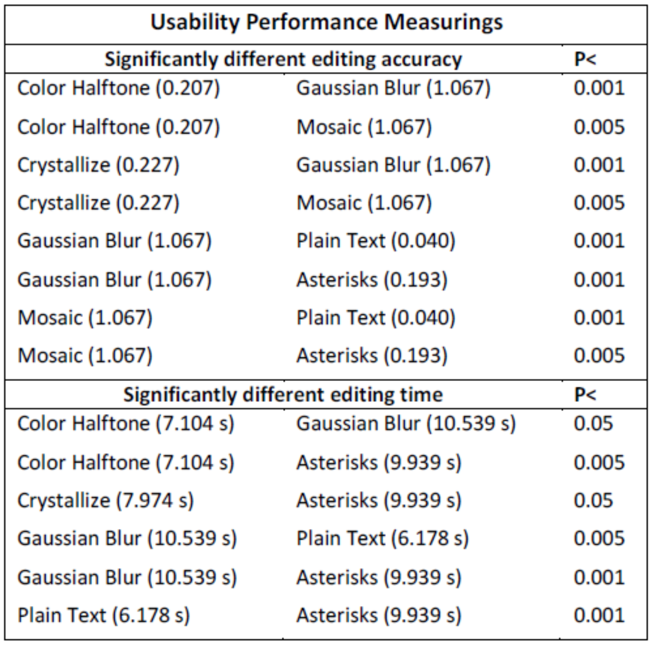


Figure 4. Filter types for Task 2 from top left to bottom right: Color-Halftone, Crystallize, Gaussian Blur, Mosaic, Plain Text, Asterisks

The ANOVA also shows significant results for the editing task, where the test person has to edit the character x at specific position. We found a significant effect of filter types on the editing accuracy (F(5, 120) = 24.600, $p < 0.0005$). Comparing the means based on the filter types, one can see that the Plain Text input performed the highest by accuracy, following by the Asterisks filter. Though, these outcomes were expected. Due to the task design the users could easily count the asterisk positions to insert the x, however, this does not really represent a use case which is why another editing task is conceivable for future investigations. Besides these results, the Crystallize and Color Halftone filters did perform satisfactorily accurate as well (see figure 5). The accuracy performance (editing accuracy) of the filters Gaussian Blur and Mosaic come off as lowest (see figure 5). That denotes that the users insert the x one position falsely next to the expected position on average. Overall, Gaussian Blur and Mosaic performed poorer than the other filters.

Between the filter types significant differences in the editing time was found (F(5,120) = 6.447, $p < 0.005$). Similar performing results as above was found. The users need the most time for the Gaussian Blur filter to edit the password, following by the Asterisks. The test persons score the fastest editing results for the Color Halftone and Crystallize filters (see figure 5). According to these outcomes, one could interpret that the classical authentication method is unsatisfying for editing task, like spelling corrections. Whereas the novel filter-based password technique is applicable for spelling corrections due to the readability for known words. According to the editing results, Color Halftone as well as Crystallize filters seem to be most promising. Future deeper investigations in editing task are required to have more precise results.

Summing up, the filters Color Halftone and Crystallize achieve the best usability results. Gaussian Blur and Mosaic performed the worst in usability and therefore not applicable for the Password Blur concept. In the following section the results from the security task are discussed.



Usability Performance Measurings		
Significantly different editing accuracy		**P<**
Color Halftone (0.207)	Gaussian Blur (1.067)	0.001
Color Halftone (0.207)	Mosaic (1.067)	0.005
Crystallize (0.227)	Gaussian Blur (1.067)	0.001
Crystallize (0.227)	Mosaic (1.067)	0.005
Gaussian Blur (1.067)	Plain Text (0.040)	0.001
Gaussian Blur (1.067)	Asterisks (0.193)	0.001
Mosaic (1.067)	Plain Text (0.040)	0.001
Mosaic (1.067)	Asterisks (0.193)	0.005
Significantly different editing time		**P<**
Color Halftone (7.104 s)	Gaussian Blur (10.539 s)	0.05
Color Halftone (7.104 s)	Asterisks (9.939 s)	0.005
Crystallize (7.974 s)	Asterisks (9.939 s)	0.05
Gaussian Blur (10.539 s)	Plain Text (6.178 s)	0.005
Gaussian Blur (10.539 s)	Asterisks (9.939 s)	0.001
Plain Text (6.178 s)	Asterisks (9.939 s)	0.001

Figure 5. Significant differences were found between the filter types above. The brackets show the mean values for editing accuracy on the right and editing time on the left

Results and Discussion - Security Task

The evaluation of Task 2 showed that there is a significant effect of filter type on the Levenshtein distance between the participant's guess and the actual password (F(5, 120)=186.9, $p < 0.001$). As visualised in the graph below (see figure 6), the traditional Plain Text and Asterisks input methods show distinctly higher values in detectability under shoulder surfing attacks than the novel filter-based input methods.

The probability that a password can be obtained by third-parties is highest for Plain Text input (M=96,5%, SD=4,37%) followed by Asterisks input (M=86,8%, SD=5,54%) while the filter-based methods were much more secure from attacks with the following detectability values in descending order: Mosaic (M=55,8%, SD=16,63%), Gaussian Blur (M=38,9%, SD=8,11%), Color Halftone (M=31%, SD=7,3%) and Crystallize (M=23%, SD=9,65%).

These low values in readability for unknown entries based on filter-based input methods show that this novel authentication method is more secure and less vulnerable against shoulder surfing than classic input methods. Taking the example of choosing password phrases as a password type, which is recommended in recent publications [14], the guessability and chance that the password is obtained in case of a shoulder surfing attack with an asterisk-based (M=84,9%, SD=8,66%) input method is between 32% (highest filter-based value: Mosaic, M=52,5%, SD=17,67%) and 68% (lowest filter-based value: Crystallize, M=16,25%, SD=1,06%) higher than with a filter-based alternative.

Qualitative Interview

In the concluding survey of our qualitative and quantitative study we were not only able to gather demographic information about our participants, but also insights about their subjective feelings while partaking in our study. On average our participants were 24 years old and like we already assumed more than two thirds are using password inputs on their mobile devices multiple times a week or more.

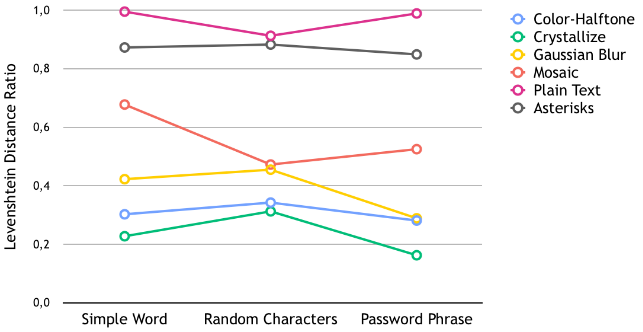


Figure 6. Average Levenshtein distance ratio depending on the word length for each password and filter type. Plain Text and Asterisks input methods show distinctly higher values in detectability under shoulder surfing attacks than the novel filter-based input methods

As security is nowadays an important topic, 84% found it the easiest to read along the state of the art asterisks-based password input after they participated in the security part of our study. This shows, there is still potential for improvement for authentication on mobile devices. Moreover 60% would use the password input based on degraded text on a daily basis. Reasons for this which was named are, inter alia, because it would be easier to correct typing errors and it would be harder for another person to read along.

When asked about the tradeoff between security and usability (which filter was the hardest to read along from a shoulder surfer perspective and at the same time was the easiest to correct) the participants liked the Crystallize filter of all filter types the most with a total of 36% of all votes.

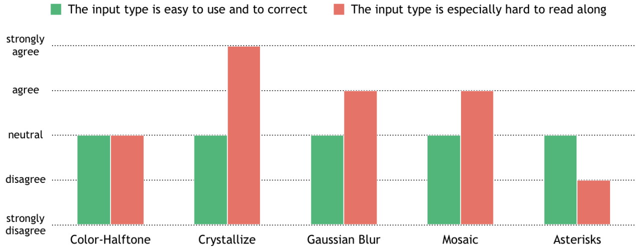


Figure 7. Median of the Likert scale: Perceived usability (green) and security (red) of different password input methods. While adding a character at a given position in the usability part of the study did feel the same for all filters, the participants found Crystallize, Gaussian Blur and Mosaic especially safe against shoulder surfing attacks in the security part

To receive the participants subjective feeling which filter was good usability wise (easy to use and accurate in inserting an additional character at a given position) and which filter was secure against shoulder surfing we used a Likert scale from 1 (strongly agree) to 5 (strongly disagree), see figure 7. The perceived usability among the different password filters and the asterisks-based input are the same and neutral. But in the perceived security there is a big difference. The participants strongly agree, that the Crystallize filter is especially hard to read along. They also agree, that the Gaussian Blur and Mosaic filters are secure against shoulder surfing. The feeling for Color Halftone is neutral and the interviewees think the asterisks-based password input is the weakest regarding the security.

Our data show, that the perceived feeling about the security of the different password input types is quite accurate as our password filters performed overall better in the second part of our qualitative and quantitative study.

CONCLUSION

To summarize, it can be stated that the filter-based authentication approach is less vulnerable against shoulder-surfing attacks but still performing at a constant high usability level and therefore a promising alternative to traditional input methods.

Comparing security and usability results we can see that the filters Color Halftone and Crystallize are most promising and therefore applicable for the Password Blur concept, that ensures a more usable as well as more secure approach for user authentication than the asterisk-based method. Password obscurations by Asterisks are the most observable based on the security study outcomes. This results due to the implementation of Android and iOS, that reveals the characters for a few milliseconds before hiding. Moreover, asterisks-based password inputs are not highly usable when it comes to spelling correction. Spelling mistakes are not able to be specifically detected, wherefore the user has to delete the whole input, which is time-consuming. With our filter-based method, users are faster detecting specific position of known words. Regarding this observation, Password Blur could provide the opportunity for a quick misspelling detection and correction, whereas the classical password input has its limitation. However, further investigation for this part is required to ensure high qualitative results. In summary, the conducted user study shows that filter types significantly influences editing accuracy and time (usability) as well as observation risks (security). On the basis of the study results, we would recommend the Crystallize filter to apply on the Password Blur method.

FUTURE WORK

The results of both quantitative and qualitative data suggest great potential for this novel authentication method to enhance usability and security on mobile devices. A long-term field-study could be conducted among a larger number of people to get even sharper insights on the differences between the filters' gradations and the overall performance. Another interesting topic of research would be to evaluate this concept within the context of messaging [6], where similar issues with regard to shoulder surfing are existing.

REFERENCES

1. Definition of shoulder surfing in english, 2017. Retrieved August 30, 2017 from `https://en.oxforddictionaries.com/definition/shoulder_surfing`.
2. Photoshop elements pixelate filters, 2017. Retrieved August 30, 2017 from `https://helpx.adobe.com/photoshop-elements/using/pixelate-filters.html`.
3. Burton, A. M., Wilson, S., Cowan, M., and Bruce, V. Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science 10*, 3 (1999), 243–248.
4. Damerau, F. J. A technique for computer detection and correction of spelling errors. *Communications of the ACM 7*, 3 (1964), 171–176.
5. Eiband, M., Khamis, M., von Zezschwitz, E., Hußman, H., and Alt, F. Understanding shoulder surfing in the wild: Stories from users and observers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM (2017), 4254–4265.
6. Eiband, M., von Zezschwitz, E., Buschek, D., and Hußmann, H. My scrawl hides it all: Protecting text messages against shoulder surfing with handwritten fonts. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ACM (2016), Pages 2041–2048.
7. Florencio, D., and Herley, C. A large-scale study of web password habits. In *Proceedings of the 16th international conference on World Wide Web*, ACM (2007), 657–666.
8. Gregory, R. L. Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 352*, 1358 (1997), 1121–1127.
9. Harada, A., Isarida, T., Mizuno, T., and Nishigaki, M. A user authentication system using schema of visual memory. In *BioADIT*, Springer (2006), 338–345.
10. Hayashi, E., Dhamija, R., Christin, N., and Perrig, A. Use your illusion: secure authentication usable anywhere. In *Proceedings of the 4th symposium on Usable privacy and security*, ACM (2008), 35–45.
11. Kinjo, H., and Snodgrass, J. G. Does the generation effect occur for pictures? *The American journal of psychology 113*, 1 (2000), 95.
12. Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, vol. 10 (1966), 707–710.
13. Meeker, M. Internet trends 2015-code conference. *Glokalde 1*, 3 (2015).
14. Shay, R., Komanduri, S., Durity, A. L., Huh, P. S., Mazurek, M. L., Segreti, S. M., Ur, B., Bauer, L., Christin, N., and Cranor, L. F. Can long passwords be secure and usable? In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 2927–2936.
15. von Zezschwitz, E., Ebbinghaus, S., Hussmann, H., and De Luca, A. You can't watch this!: Privacy-respectful photo browsing on smartphones. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM (2016), 4320–4324.
16. Williams, E. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry 2*, 2 (1949), 149–168.