Exploring the Impact of Colour-Blindness on Computer Game Performance

Daniela Napoli, Sonia Chiasson
Carleton University
Ottawa, Ontario
daniela.napoli@carleton.ca, chiasson@scs.carleton.ca

ABSTRACT

This study assesses the effects of colour-blindness on casual computer gaming performance and experience. Participants played an online puzzle game in both an unmodified environment and in a colour-blind simulation. Some of our results suggests that the colour-blind filter did not negatively impact player performance. However, post-test questionnaires reveal that users perceived the game to be more difficult and their performance as limited. These contradicting results highlight the importance of further assessing the impact of colour vision deficiency on performance, and the accessibility principles used in game design when colour coding operable cues.

Author Keywords

Accessibility; colour blind; online games; performance

INTRODUCTION

In game design, colours can capture players' attention and guide them while making decisions. Research has suggested that some colours can alter an audience's emotions and bodily reactions. These colours may not have the same degree of effectiveness on colour-blind individuals because they cannot view the interface as intended. But, related literature has yet to verify that.

In this study, I try to assess the colour-blind gamer's performance and experience. Ten participants played six levels of a casual, online puzzle game in its original state and then in a colour-blind simulation. Key aspects of their performance, including mouse clicks and level duration, were not significantly different in the modified condition from the controlled setting. Yet, participants perceived the second iteration of the game as more challenging, and their performance as limited, due to a lack of easily differentiable operable cues.

The results of this research can support the advantages and limitations of accessible interface design guidelines, including the W3C WCAG 2.0. Follow-up research that isolates the colour-blind deficiency variable, preferably with colour-blind individuals, can further support this study's findings. Additionally, the effects of colour-blindness on other game genres would elicit a more comprehensive understanding of this group's performance.

Citation: D. Napoli, S. Chiasson. Evaluating the Impact of Colour-Blindness on Computer Game Performance. Technical Report TR-17-03, School of Computer Science, Carleton University, September 2017.

BACKGROUND

Since the start of this millennium, colour has played an important role in interface design. Wolfson and Case (2000) suggest that players' brains are more alert and their performance is more successful when red colours, rather than blue colours, dominate the screen [11]. The same study suggested that the latter facilitate a long-lasting positive influence on overall interest and performance. One decade later, the effects of blue interfaces were reassessed. Bonnardel's (2010) users not only preferred to see colour while surfing the web, but also navigated blue websites faster than ones with green or grey colour schemes [1]. Plass' (2014) study revealed that developers can elicit positive emotions and increase comprehension rates by incorporating preferred colours — blue, again, and orange — in design [9]. Yet, all this research assumes these effects only apply to users with average colour perception. Do colour-blind individuals have similar experiences?

The core of recent colour vision deficiency (CVD) literature is rooted in Meyer and Greenberg's (1988) approach to adapting computer outputs to remedy or correct the impairment [7] rather than evaluate the usability implications of colour-blindness. In 2007, Jefferson and Harvey developed a corrective algorithm and computer application that allows users to modify colour saturation and contrast so that any information viewable on a desktop monitor can be made more distinguishable [4]. Since deficiencies of any sort are subject to variety, Flatla and Gutwin offer a more adaptive output algorithm [3] and colour augmenting tool called SSMRecolor [6]. In 2016, Reinecke et al. began to consider how other external factors such as operating environment, available light, monitor brightness, and age can impact interface usability [10]. The authors created an image processing tool called ColorCheck and used it to learn that 88 per cent of their study's participants, some with typical vision abilities, were unable to distinguish colour-cues included on popular websites.

To the best of our knowledge, influential literature regarding CVD stops there. Recent studies do not explore how colour-blindness affects information processing abilities and overall user experience. This study will begin to remedy this void by exploring the differences between CVD and non-CVD performances and user experiences through simulation methods. Empirical evidence describing the colour-blind performance can help identify requirements for design accommodation for this user group.

HYPOTHESES

To understand how colour-blindness affects gamers, this study will compare participant performances in an unmodified environment and in a simulated colour-blind environment. The results of the study are expected to prove that groups of operable cues become more challenging to differentiate when they are colour-coded in a limited range of hues. We hypothesize that:

> H1: Users' performance will worsen within the colour-blind state. Poor performance will be measurable through lower scores, more overall clicks, and longer decision-making durations.
>
> H2: Users will perceive their performance as slower, and the game as more difficult, when playing in the colour-blind state.

SYSTEM DESCRIPTION

Setup

Participants used a desktop Windows PC and a standard computer mouse to play an in-browser version of Pet Rescue Saga with Google Chrome in a controlled lab setting. The space was lit with two standard sized panels of fluorescent lighting. The display monitor was uncalibrated, 22 inches wide, and had a resolution of 1080ppi.

Computer Game

The participants played an in-browser version of "Pet Rescue Saga." King Digital Entertainment originally developed the game as an application for Facebook. Since then, it has received over 25 million likes and has become a standalone game for browsers and Android and iOS personal devices. When playing Pet Rescue Saga, the player must click on a group of matching blocks to destroy them and in turn, allow the pets to fall to the bottom of the screen and "rescue" them. In its intended state, each block group is categorized by five distinct colours and corresponding over-lay patterns (Figure 1).

Simulating Colour Vision Deficiency

Due to the limited access to CVD individuals, few participants were expected to be colour-blind. To gauge the CVD performance, participants played within a colour-blind simulation. We opted for a Google Chrome plugin that was compatible called "Spectrum" [5] to mimic a state of colour-blindness. For this study, we focused on simulating a type of red-green CVD called Deuteranopia [8]. This filter reduced the range of visible colours to shades of blue and yellow. As a result, the blocks' patterns became less stark, but were still visible (Figure 2).

Colour-vision deficiencies are not always as severe as the one being simulated in this study. For example, Deuteranomaly (Figure 3) could have been used, but its related changes were less severe and may have limited the expressiveness of the study's results. Furthermore, the effects of study's filter can be compared to other CVD conditions that similarly severely reduce the range of visible hues such as, Protanopia (Figure 4) or Tritanopia.

Figure 1 – The unmodified interface players used in Phase 1 of the user study.

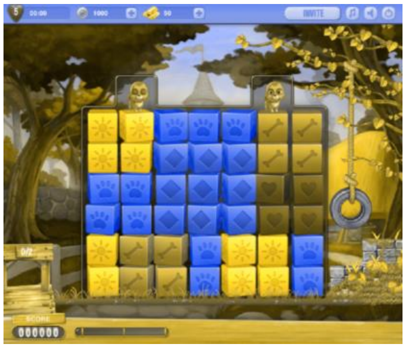

Figure 2 – Phase 2 game interface with red/green colour-blindness, Deuteranopia, filter applied.

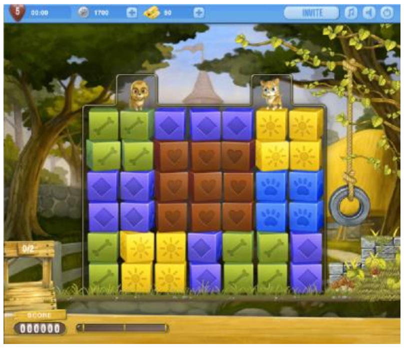

Figure 3 – The less severe effects of more common colour-blindness, Deuteranomaly.

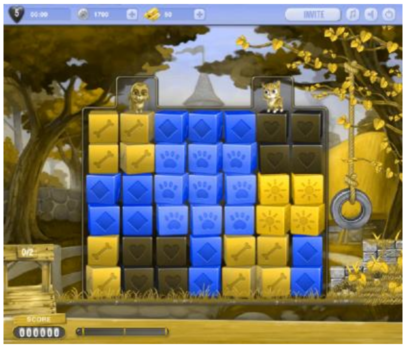

Figure 4 - Tested game with a comparably severe colour-blindness, Protanopia, filter.

METHODOLOGY

Participant Recruitment

Due to our use of the colour-blind simulator, it was not mandatory for our participants to be colour-blind. University students of all colour-vision abilities were recruited with on-campus posters, e-mails, and social media postings. Out of the total 12 participants, nine were aged between 21-39 years old. Eleven were enrolled in graduate courses in the Computer Science field. The same number of participants said they played video games less than five hours per week; the remaining participant played up to 10 hours per week. According to the results of our post-test questionnaires, none had ever played Pet Rescue before the study, but four had previously played games with similar rules and interfaces.

Procedure

The university's Research Ethics Board cleared the following protocol. For this study, participants completed two phases of gameplay that lasted for approximately 20 minutes each. They then answered questions regarding their experience. We anonymized all participants' responses and recorded their on-screen interactions with screen-recording software called Camtasia.

Before gameplay, we measured each participant's ability to see colour with an online version of the Farnsworth-Munsell (1986) colour arrangement test [2]. Based on how the participant arranged on-screen colour samples, the test categorized each user with a CVD type (including "not colour blind") and severity. The test provided some written directions; however, participants typically asked for clarification. Our guidance was not standardized but, we told each participant to "create a spectrum" by ordering the on-screen colour samples according to hue. With this test, 11 participants were identified with average colour vision abilities. Only one participant, P12, was identified with "slight Deuteranomaly."

After the CVD test, participants began Phase 1 of gameplay. During this phase, each participant played six levels of Pet Rescue in sequential order. We withheld game instructions and objectives to limit our influence on player performance. Without assistance, all users seemed to understand that they needed to click on the game's blocks to "rescue" the pets. A few participants requested more information regarding winning techniques; however, the researchers withheld performance improving tips.

During Level 5 of Phase 1, the game introduced a new tool, a rocket, to get rid of blocks that could not be destroyed with typical click-based methods. The game's interface inconsistently instructed users how to use the rocket. Only participants 1, 3, 5, 8, 11 and 12 were exposed to a game induced tutorial, but to ensure all players had equal opportunity to learn this functionality we provided participants 2, 6, and 9 guidance specifically for this new tool. We chose not to provide guidance to participants 4, 7, and 10 since they correctly used the new feature without hesitation. It appeared they relied on their past gaming experiences to infer the rocket's utility, and we assumed further explanation would have been redundant and interfered with the flow of their game play.

During Phase 2, each participant played the same six levels again with the Deuteranopia filter engaged. We did not provide game instructions, and the Level 5 tutorial continued to sporadically appear for only some participants during this phase. We briefly reminded the participants of the rocket tool if game's tutorial was not triggered for them a second time.

After Phase 2, each participant completed post-test questionnaires. The questionnaires requested the participants to reflect on the game and their performance, explain how they used colour to make decisions, and list any other methods in deciphering operable cues during this study.

DATA COLLECTION

Game Recordings

When reviewing the screen-recordings, we focused on collecting three main pieces of information to evaluate user performance: 1) player score, 2) number of clicks, and 3) level duration.

For player scores, we recorded the number that was automatically generated by the game. We are unsure how player scores were calculated; however, it seemed to favour individuals who broke large groups of blocks at once.

During playback, the screen-reading software illustrated each time the mouse is clicked with a pulsing, red halo around the cursor. We counted the number of illustrated clicks respective to gameplay interface when reviewing the recordings. We excluded clicks that navigated between levels or closed dialogue boxes as they were not reflective of game strategy. We categorized each in-game click depending on the consequence of the player's move: a "successful" click broke a block to help rescue a pet, and an "unsuccessful" click failed to destroy the selected blocks, suggesting the player did not click on a group of blocks with matching colours/over-lay patterns.

To measure level durations, we recorded the time difference between the first and last successful or unsuccessful clicks. Unlike observations regarding clicks, we included any time (typically milliseconds) spent reading dialogue boxes as we considered this variance reflective of their individual abilities to process information within the unmodified and colour-blind conditions. In-game tutorials occurred before the first clicks and were not included in the level durations.

After reviewing the data, we decided to omit the following two sets of results to calculate valid averages: P4 did not play all six levels in sequential order, and could have been advantaged/disadvantaged when playing the game in an unintended order of increasing difficulty; P12 was the only person diagnosed with a mild CVD and perceived the interface and the colour-blind filter differently than all other participants, thus rendering their performance as a theoretical outlier.

Post-Test Questionnaires

When reviewing the post-test questionnaires, we recorded all yes/no and short-answer questions verbatim for further observation. Higher participants responses on Likert-scale based questions suggested better performance and ease.

RESULTS

The data of the remaining 10 participants was analyzed. According to the quantifiable portions of the participants' performance, the study's conditions were not significantly correlated with changes in player scores, mouse clicks, and decision-durations. At times participants performed better in with the colour-blind filter in Phase 2, than in the unmodified state during Phase 1. These findings challenge my original hypotheses as they suggest that players were not impacted by visual changes in operable cues.

However, the qualitative information gathered from participants suggest that user perceptions of their own performance were aligned with my original hypotheses. Post-test feedback suggests that the colour-blind filter forced players to change their original strategies. Participants often associated this forced change with a decline in their performance. This dichotomy suggests further research should be made to reassess the effects of colour-blindness on player performance.

Lower Overall Scores

Average player scores were fitted to the trend lines illustrated in Figure 5; this graph suggests a declining tendency in average player scores over the course of the game. Players scored somewhat lower when playing with the colour-blind filter (M = 22564.8, SD = 5677.0) than in the unmodified state (M = 23259.4, SD = 5404.2). A Pearson's product-moment correlation was run to assess the relationship between study's conditions and players' scores. There was no significant correlation with scores, $r(8) = -0.2$, $p = 0.5$, and game state.

The negative, but insignificant, change in mean player scores between phases are consistent with my first hypothesis; players score less points when playing within the confines of colour-blindness. As previously mentioned, our insights related to score differences are limited since it is unclear how Pet Rescue awards points to players. Though, while observing participants, it appeared that players who destroyed larger groups of blocks would be rewarded more points than players who destroyed smaller groups. If this is true, an insignificant difference between phases could suggest that players were able to map large groups of operable cues regardless of available hue range. Yet, since average scores were lower, and slightly negatively correlated, the colour-blind filter could have limited players' abilities in accomplishing strategies that would award them with higher points.

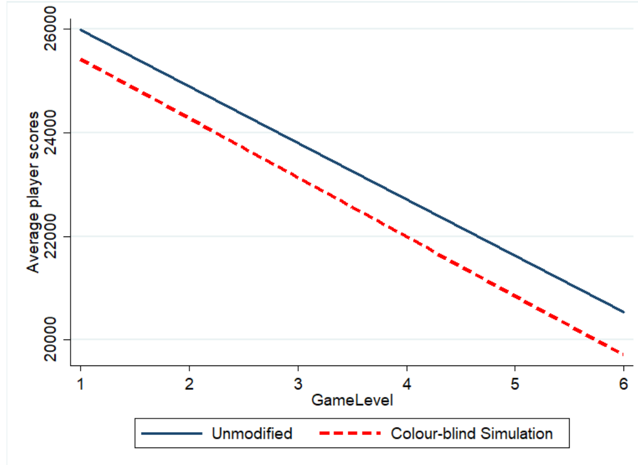

Figure 5 - Trends in average player scores during both study phases

More Overall Clicks and Less "Successful" Choices

As shown in Figure 6, average mouse clicks increased as the game progressed. This graph also shows that although there was a gradual increase between both phases, players clicked more in the unmodified state (M = 9.9, SD = 7.8) than when playing with colour-blind filter (M = 8.4, SD = 5.2).

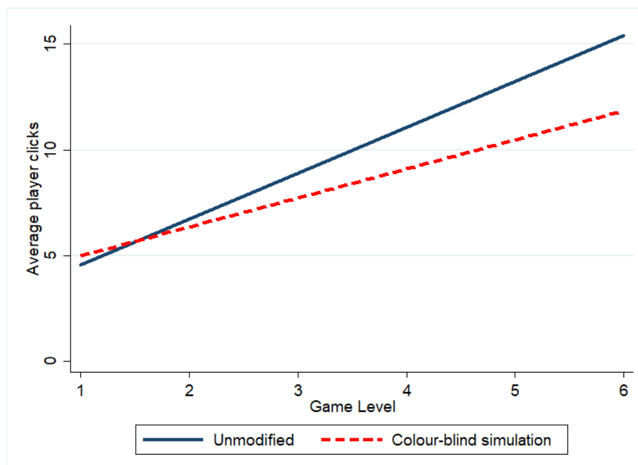

Figure 6 - Trends in average mouse clicks during both study phases

A Pearson product-moment correlation test suggests that there was a small negative correlation between conditions, $r(8) = -0.1$, $p = 0.2$, and mouse clicks. These findings seemingly counter my second hypothesis because they suggest that players took fewer clicks to complete levels during the second phase, and thus the colour-blind simulator may not have negatively impacted the success of their choices.

However, we cannot be certain that this small negative correlation accurately reflects the effects of colour-blindness on the trends of player mouse clicks. It is possible that participants became acclimated with the game by the time they had a second chance to play during Phase 2 (the colour-blind simulation stage), and player performance may have naturally improved due to learning effects. Giving participants an opportunity to learn the game and re-strategize may have positively skewed mean mouse clicks made within the colour-blind state. This could be a limitation of choosing not to counterbalance conditions between participants, and will be discussed in later parts of this paper.

Longer Decision-making Durations

To measure decision-making durations, I divided the level intervals by the total number of clicks (i.e., decisions) to estimate how long it took players to scan the screen, strategize, and choose a block during gameplay. As reflected in Figure 7, average decision-making durations increased as the game levels increased. Decisions also too longer to make during Phase 1 (M = 2.3, SD = 1.4) than in Phase 2 (M = 2.2, SD = 1.2).

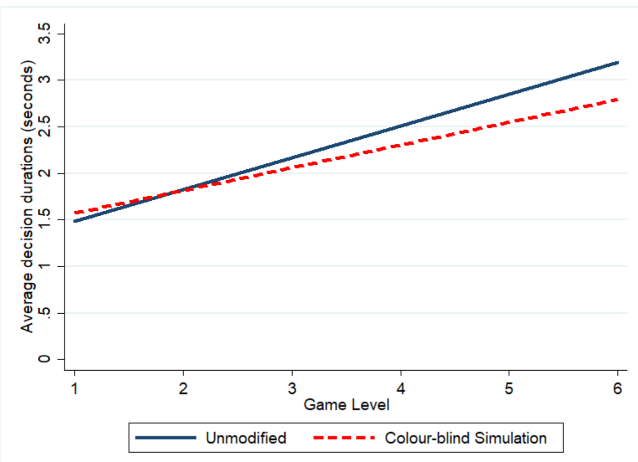

Figure 7 - Trends in decision-making durations during both study phases.

These results suggest that player performances decreased as the game became more difficult, but increased when given the opportunity to repeat the same level. A Pearson correlation test suggests that the study's conditions had a negative, albeit insignificant, correlation with player decision-making durations, $r(8) = -0.1$, $p = 0.5$, with game conditions. As previously discussed, participant's improved performance may not be due to the colour-blind filter but rather their acclimation for the game.

Perceived Performance

The results of the post-test questionnaires suggest that participants' perceived performance was consistent with our original hypotheses. It is important to understand any negative perceptions of game difficulty and performance as they may reveal aspects of poor user experience or usability.

After playing the game in an unmodified state during Phase 1, participants rated the game closer to "very easy" (M=3.6, SD=0.5) and their performance as "fast" (M=3.0, SD=1.0). In the short-answer sections, most players noted that the variety of colours allowed them to match related blocks easily, and execute their next moves quickly. After playing Pet Rescue within the colour-blind simulation, participants rated the game as difficult (M=3.0, SD=0.5) and their performance as slow (M=2.0, SD=0.3).

Several participants reported that they felt they made more errors within the colour-blind simulation than in the unmodified state. Most noted an induced reliance on the blocks' over-lay patterns, rather than obvious colour changes, to differentiate between groups of matching blocks. A couple of participants noted that they focused on slight changes in colour shades rather than hues to group matching blocks. Often, participants associated any induced changes in their techniques with the changes in their ratings of game difficulty and personal performance.

The Colour-Blind Participant's Performance

P12 reported that she played games, such as Backgammon and memory training puzzles, for less than 5 hours a week on average. She was the only participant diagnosed with a colour-vision deficiency, slight deuteranomaly, per the results of the Farnsworth-Munsell evaluation. This type of colour-blindness is generally a milder version of the deutan-based condition simulated in this study.

When discussing the effects of the simulation, P12 noticed changes in the interface's saturation and contrast — *"Everything looks less bright and greyer"* — with the filter engaged. According to the post-test questionnaires, P12 focused on grouping blocks according to symbols rather than colours. They reported no change in their strategy between conditions. Also, there was no change in their rating of the game's difficulty (very easy), nor the perceived speed of performance (fast), between phases.

Unlike the other participants, P12's average score increased during the second phase of the study. With even greater success than the other participants, P12 made fewer clicks and spent less time making decisions within the colour-blind simulation. Overall, P12's performance improved over the course of the study. This positive trend in P12's performance suggests that the colour-blind simulation did not impede on their ability to learn the game, and decipher operable cues. The significance of P12's performance trends will be discussed in later sections.

DISCUSSION

On average, players did not seem to be significantly impacted by the colour-blind filter. Although, since there were no strong correlations between performance indicators and the two conditions, this may suggest that future research could capture more distinct trends. These trends that could be supportive of our original hypotheses, or reflect the advantages of accessible design guidelines.

Lessons Learned

We intended to gather a fair, baseline measure of performance when introducing a new game to participants in its unmodified state. However, we may have indirectly advantaged performance within the colour-blind simulation by allowing them to play the same six levels for a second time. Between-participant design may have allowed participants to learn, and perform within one game state. Differences in the performances between average players, and "colour-blind" players may have allowed us to assess the impact of CVD. If between-participant design was not feasible (a small sample group available), counterbalancing the states may have mitigated the learning effects that could have advantaged colour-blind performance.

Using a game with variable difficulty may have also limited our ability to infer that changes in performance were due to the visible changes in the interface. Future studies may want to explore a game of stagnant difficulty so that researchers can explore differences in learning curves, and reduce the number of confounding variables within the colour-blind player's performance. This may also allow researchers to understand the long-term implications of CVD.

Authentic behaviours and thought-processes are unique to individuals who have colour-vision deficiencies. The results of an induced colour-blind study may be argued as ecologically invalid. Yet, recruiting colour-blind individuals is difficult due to the relatively small population. Inducing colour-blindness through Spectrum, or other CVD simulation software, is regarded as a more convenient solution. Since the software did not seem to negatively impede on the colour-blind participant's performance, these tools could be effective. Future studies should formally evaluate these tools' ability to mimic the colour-blind experience.

The Advantages of the WCAG

Our results contradicted all original hypotheses, but it did provide some insight on current accessible design principles. The Success Criterion 1.4.1 section in WCAG 2.0 addresses many ways to incorporate colour in accessible interface design. According to these guidelines, information that is conveyed through colour must be paired with different signifiers including patterns, text and selection tools to ensure clarity for any type of perception ability.

Pet Rescue Saga's interface is consistent with these suggested techniques and it uses colours, symbols and a selection border to help organize the blocks onscreen. The mostly positive results of this study allude to the effectiveness these colour-blind accommodations. The participants also used these pieces of the design interface to cope with the CVD filter:

> *P6: In [the colour blind] version, I relied mostly on the symbols. I think I may have checked the colours, but as a secondary thing. I'm not even sure now whether the colours were unique for each symbol. I think decisions were slow at the beginning, till [sic] I got used to relying on the symbols rather than the colours...*

Furthermore, P12's techniques disregarded colours entirely. Their ability to perform, and improve their performance, over the course of the study suggests that the game's secondary cues were enough to ensure their success. Further research exploring the techniques of colour-blind users could confirm the value of these secondary cues, and their effectiveness when playing casual puzzle games.

CONCLUSION

Our analysis of the quantitative performance indicators does not suggest that colour-blindness has a significant impact on players' scores, successful decisions, nor decision-making durations. However, we do not assume these results prove a null hypothesis. We expect that the effects of colour-blindness on gamer performance would be better assessed with further research that addresses this study's limitations and further isolates of the colour-blindness variable. We also suggest that further research should evaluate the effectiveness of using secondary cues (such as over-lay patterns) that are often mentioned in accessibility design guidelines and relied on by participants who need to re-think their strategies in interacting with colour-coded cues that have been compromised.

ACKNOWLEDGEMENTS

S. Chiasson acknowledges funding from NSERC for her Canada Research Chair and Discovery Grant.

REFERENCES

1. Nathalie Bonnardel, Annie Piolat, and Ludovic Le Bigot. 2011. The impact of colour on Website appeal and users' cognitive processes. *Displays* 32, 2: 69–80. https://doi.org/10.1016/j.displa.2010.12.002

2. Colblindor. Farnsworth-Munsell 100 HueColor Vision Test. Retrieved July 20, 2011 from http://www.color-blindness.com/farnsworth-munsell-100-hue-color-vision-test/

3. David R Flatla and Carl Gutwin. 2011. Improving Calibration Time and Accuracy for Situation-specific Models of Color Differentiation. *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*: 195–202. https://doi.org/10.1145/2049536.2049572

4. Luke Jefferson and Richard Harvey. 2007. An interface to support color blind computer users. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, 1535.

https://doi.org/10.1145/1240624.1240855

5. Yehor Lviski. 2014. Spectrum Experiment. Retrieved from http://lvivski.com/

6. Rhouri McAlpine and David R. Flatla. 2015. Real-Time Mobile Personalized Simulations of Impaired Colour Vision. *Sigcomm 2015*: 421–434. https://doi.org/10.475/123

7. Gary W. Meyer and Donald P. Greenberg. 1988. Color-Defective Vision and Computer Graphics Displays. *IEEE Computer Graphics and Applications* 8, 5: 28–40. https://doi.org/10.1109/38.7759

8. Jeremy Nathans, Darcy Thomas, and David S. Hogness. 1986. Molecular Genetics of Human Color Vision: The Genes Encoding Blue, Green, and Red Pigments. *Science, New Series* 232, 4747: 193–202.

9. Jan L. Plass, Steffi Heidig, Elizabeth O. Hayward, Bruce D. Homer, and Enjoon Um. 2014. Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction* 29: 128–140. https://doi.org/10.1016/j.learninstruc.2013.02.006

10. Katharina Reinecke, David R. Flatla, and Christopher Brooks. 2016. Enabling Designers to Foresee Which Colors Users Cannot See. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2693–2704. https://doi.org/10.1145/2858036.2858077

11. S. Wolfson and G. Case. 2000. Effects of sound and colour on responses to a computer game. *Interacting with Computers* 13, 2: 183–192. https://doi.org/10.1016/S0953-5438(00)00037-0