Placing a Value on Aesthetics in Online Casual Games

Erik Andersen, Yun-En Liu, Richard Snider, Roy Szeto, and Zoran Popović
Center for Game Science
Department of Computer Science & Engineering, University of Washington
{eland,yunliu,mrsoviet,szetor,zoran}@cs.washington.edu

ABSTRACT
Game designers frequently invest in aesthetic improvements such as music, sound effects, and animations. However, their exact value for attracting and retaining players remains unclear. Seeking to estimate this value in two popular Flash games, we conducted a series of large-scale A/B tests in which we selectively removed aesthetic improvements and examined the effect of each component on play time, progress, and return rate. We found that music and sound effects had little or no effect on player retention in either game, while animations caused users to play more. We also found, counterintuitively, that optional rewards caused players to play less in both games. In one game, this gameplay modification affected play time three times as much as the largest aesthetic variation. Our methodology provides a way to determine where resources may be best spent during the game design and development process.

Author Keywords
games, aesthetics, analytics, A/B testing

ACM Classification Keywords
H5.0. Information interfaces and presentation: General.

General Terms
Experimentation, Human Factors

INTRODUCTION

Game designers must decide how to focus limited resources on various aspects of their games. Aesthetic elements such as music, sound effects, and animations, for example, are believed to be crucial components of the video game experience. Schell [9] argues that good aesthetics can make the player more likely to tolerate imperfections in game design, and can draw a player into a game they might have otherwise ignored. However, the exact effect that aesthetic quality has on how long players play games is unknown. Discovering this information would be of considerable use to game designers, in order to focus their effort on the areas of their games with the greatest impact.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Performing large-scale experiments to test the effects of these game elements on player behavior can be difficult. However, the rise of online casual and "indie" games [12] gives researchers a cheap and attractive option for gathering massive amounts of experimental data. Kongregate, one of the most popular websites for free online games, attracts 10 million players who collectively spend 23 million hours playing games per month [1]. The popularity of these websites and the ease with which a new game can be uploaded and distributed makes them an ideal testbed for experimentation. Game companies have started to gather data about player behavior remotely through internet analytics and beta testing; unfortunately, most of this data remains unpublished. A well-known advocate for such testing is the social games company Zynga, who used a metrics-driven approach [4] to bring FarmVille to a peak of almost 88 million users [2]. By instrumenting our games to record player behavior, we also are able to measure player engagement quantitatively on a large scale.

In this study, we evaluate the importance of aesthetic quality in two Flash games, Refraction and Hello Worlds, that we developed and uploaded to Kongregate. Through A/B testing, we varied these elements one by one and measured how much progress players made in the game, how long they played, and how likely they were to return. We expected to find that each of these aesthetic improvements significantly increased player retention; however, we found no significant effects for music and sound effects in either game. Animations caused players to play significantly longer. In contrast to what is commonly believed, we also found that the presence of optional rewards decreased play time. Further research is necessary to determine whether these results generalize to other games.

In contrast to in-house playtesting, our experiments are conducted "in the wild" where players do not know they are part of an experiment. This gives us the advantage that players are playing under their own motivation and that our findings are based solely on their observable behavior; however, there is also the disadvantage that we cannot know what players are thinking or feeling. These quantitative measurements can be used in conjunction with other quantitative and qualitative data to provide strong empirical reasoning for design decisions. Our methodology can help determine where resources should be spent during the game development process, and also suggests a new way of gathering huge numbers of participants for answering research questions relevant to games and user interfaces.

	Hello Worlds	Refraction
Music	eleven alternating songs	title screen, intro and ending movie, song for each of the 7 worlds
Sound	picking up a coin using a door and error sound on unusable doors applause at end of each level bouncing on trampoline	picking up and placing pieces on the board navigating the menu fireworks and spaceship flyaway at the end of each level overpowered, underpowered, satisfied, and unsatisfied targets
Animation	walking, jumping, climbing ladder, falling, turning left and right fading between stages counting points and fireworks on level finish combination view gliding rewind bar projector animation on worlds picking up coins and stars unusable door quake	intro and ending movies, intro level penguin animation laser splashes and pulses animated piece description when player hovers over the piece fireworks and spaceship flyaway at the end of each level animated title screen, moving title bar, level select screen zooming arrows, blinking animations, and moving pieces in tutorials magnifying glass

Figure 1. Description of various aesthetic additions in Hello Worlds and Refraction that we removed to evaluate their importance. Removing music and sound had no significant effect in either game. Removing animations had negative effects on player progress, time played, and return rate in Refraction, and time played and return rate in Hello Worlds.

AESTHETICS

The word "aesthetic" has multiple meanings in the context of games [8]. Here, we are concerned with content that adds style and artistic depth to the player experience. Aesthetics can also be used to create a sense of theme; however, our goal is to evaluate our investment in aesthetics, not to determine whether our particular choice of art or theme was optimal. Others have studied the effects of aesthetics on user behavior. In user interfaces, Kurosu and Kashima [6] and Tractinsky [11] showed that the aesthetic quality of an interface can affect how usable the interface is perceived to be. In games, surround sound [10] and music [7, 3] have both been shown to increase enjoyment and sense of presence.

LARGE-SCALE A/B TESTING

To collect data on a large scale, we rely on A/B testing and player metrics. In A/B testing, researchers present two conditions randomly to users in the same population and see how they respond, allowing them to assign quantitative values to different aspects of their systems. For example, Amazon uses various metrics to discover which version of a website layout will maximize sales [5]. The same method also applies to games; Zynga, for example, used this approach to find that pink fonts caused players to click on an advertisement for PetVille far more often [4].

For our tests, we used two Flash games that we developed. Hello Worlds is a puzzle-platformer game in which the player's character inhabits as many as four worlds at the same time. Hello Worlds has been played over 900,000 times on Kongregate since its release in May 2010. The game was developed by a team of undergraduate computer science students and an artist as part of a game design capstone course. The visual style of the game is relatively simple, but the game contains music, sound effects, and some animations. The music and sounds were taken from free audio websites. A list of the aesthetic elements that we examined can be seen in Figure 1.

Refraction is a puzzle game in which players use pieces to split and redirect lasers in order to satisfy targets that all want some specific fraction of the laser. A team of graduate and undergraduate computer science students created this game as part of a larger games for learning project. Refraction was released to Kongregate in September 2010, and has been played over 100,000 times. The game features extensive artwork and animations created by three artists. Similar to Hello Worlds, and probably to most Flash games, the music and sound effects were selected from websites for audio under the Creative Commons license. Figure 1 also shows the aesthetic qualities that we examined for Refraction.

METHODOLOGY

To evaluate the importance of music, sound effects, animations, and other game-related aesthetic variables, we conducted a series of A/B tests over a two week period. For each experiment, we changed the versions of Hello Worlds and Refraction on Kongregate, each time replacing the game with a version that tested a single aesthetic variable. In each trial, the uploaded game would remove the target aesthetic variable with 50% probability. We also removed all of the components that suggested the presence of that aesthetic, such as buttons for toggling on and off sound.

We tracked players with variables stored in the local Flash cache. We focused solely on new players; veteran players who had already played the game were not included. There is always some possibility that a player might clear the cache or use a different computer and be treated as a new player.

We measured player engagement in three ways. First, we counted the number of levels that each player completed during their first run-through of the game. Second, we measured the amount of time that each player spent during this first run-through. Since players occasionally idle for long periods while playing Flash games, we aggregated moves the player made in 30-second intervals. If two or more consecutive intervals (a full minute) had no player actions, this time was discounted from the total play time. Finally, the third measure of engagement that we used was return rate. If the player loaded the game again within a day after the first play session, we counted them as a returning player.

DATA ANALYSIS AND RESULTS

Our data was not normally distributed; for example, user play times exhibited a spike near 0 and a long falloff (Fig-

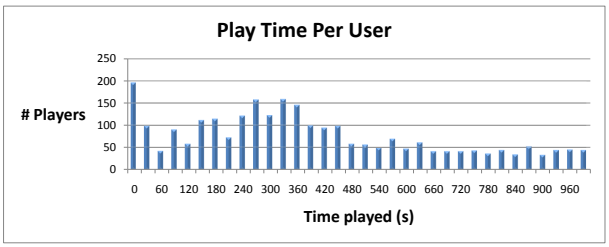

Figure 2. Typical user play times for their first play of Refraction. Many people quit immediately after loading the game. The data is not normally distributed, so we must rely on nonparametric tests.

ure 2). We therefore relied on a nonparametric test, the Wilcoxon / Kruskal-Wallis 2-sample test, to analyze levels completed and time played. For return rate, we used Pearson χ^2 analyses to compare the percentages.

Music and Sound Effects

The first experiment evaluated the importance of music. In Hello Worlds, 4915 people played the game, 2491 with music and 2424 without. There was no significant change in the median number of levels completed (M = 7 levels) (Z = 1.293, p = 0.20), nor was there an effect in the time played comparing those with music (M = 458 seconds) and those without (M = 450 seconds), (Z = 1.061, p = 0.29). No significant effect was found for return rate, $\chi^2(1, 4915) = 0.18$, p = 0.67, with 16.7% of those with music returning, and 16.25% of those without music. In Refraction, 3883 people played the game while testing music, 1939 with music and 1944 without. No significant effect was found for levels completed (Z = -0.502, p = 0.62); players with and without music completed the same number of levels (M = 10). Players with and without music played the same amount of time (M=540), (Z = -0.446, p = 0.66). Finally, no significant effect was found for return rate, $\chi^2(1, 3883) = 2.67$, p = 0.10, with 11.6% of those with music returning, compared to 9.9% of those without music.

Next, we evaluated the importance of sound effects. In Hello Worlds, 4984 people played the game, 2491 with sound effects and 2493 without. Players with sound (M = 7) did not complete significantly more levels than those without (M = 6), (Z = 0.193, p = 0.85). No significant effect was found for time played (Z = 0.265, p = 0.27) between those with sound (M = 450) and those without (M = 390). 16.9% of players with sound returned, compared to 16.7% without, which was not significant $\chi^2(1, 4984) = 0.03$, p = 0.86. In Refraction, 3892 people played the game during the experiment, 1951 with sound and 1941 without. No significant effect was found for levels completed (Z = -0.568, p = 0.57) between those with sound (M = 11) and those without (M = 10). No significant effect was found for time played (M = 600 for both groups), (Z = -0.720, p = 0.47). Finally, no significant effect was found for the return rates of 8.2% with sounds and 7.5% without, $\chi^2(1, 3892) = 0.531$, p = 0.47.

We found little payoff for our investment in music and sound effects, as we did not observe significant effects on player retention. It may be the case that many players are playing

Figure 3. Example of animation in Refraction. Removal of the animations in an A/B test caused significant decreases in level completion, play time, and return rate.

with their audio turned off, perhaps because they are in a location where sound would disturb others, such as at work or school. We unfortunately cannot know remotely if players are listening to the audio. Although music tempo has been shown to affect player enjoyment in racing games [3], our games do not rely heavily on audio feedback, potentially reducing any effect on player behavior.

Animation

In Refraction, 7765 people played the game while testing animations, 3938 with animations and 3827 without. Figure 3 shows one such animation. A significant effect was found for levels completed (Z = -2.473, p = 0.01). Players with animations (M = 11) completed more levels than players without animations (M = 9). A significant effect was also found for time played (Z = -5.973, $p < 0.0001$). Players with animations (M = 570) played longer than players without animations (M = 480), a difference of almost 20%. Finally, the return rate dropped from 13% to 11% for players without animations $\chi^2(1, 7765) = 5.046$, p = 0.02. In Hello Worlds, 5050 people played during this experiment, 2493 without animations and 2557 with animations. The median levels completed was 6 for both groups, and was not significant (Z = -1.302, p = 0.19). However, people played longer with animations (M = 390) than without (M = 330), and this was significant (Z = -2.851, p = 0.004). Players also returned significantly more with animations (19.2%) than without (16.5%), $\chi^2(1, 5050) = 6.35$, p = 0.01.

Animations serve an integral purpose, as they are the primary means of feedback about game state. Removing them makes it more difficult for the player to evaluate what is actually happening. Animation is also often used to give the user hints about how different objects in the game world work. It is therefore not surprising that the lack of animation would cause players to quit because gameplay and usability were negatively impacted.

Modifying Gameplay

We also compared the effect of aesthetic and gameplay variations. Most levels of Hello Worlds have optional coins scattered throughout the level that the player can collect to gain a higher score. Similarly, many levels of Refraction contain coins that present an additional challenge for each level. Refraction also rewards the player with cards for certain achievements. Prior to this experiment, we believed that

these optional rewards helped to keep players hooked and feel a sense of accomplishment, similar to the achievements that have recently become popular in online games. Therefore, they would encourage players to play longer and return more often. Removing these optional rewards, however, had the opposite effect. During this experiment, 4162 people played Hello Worlds without coins and 4207 people played with them. Removing the coins increased the median number of levels that players complete from 6 to 9, ($Z = -10.382$, $p < 0.0001$). Since removing coins might cause players to focus more on completing the levels rather than collecting optional rewards, this is not surprising. However, the median player played about 50% longer without coins ($M = 510$) than with coins ($M = 330$), ($Z = 4.652$, $p < 0.001$). We also found a significant effect on return rate, $\chi^2(1, 8369) = 11.98$, $p < 0.01$. 20.2% of players without coins returned, compared to 17.2% with coins.

The same effect was observed in Refraction. 2294 people played Refraction without coins and 2281 people played with them. Removing the coins increased the median number of levels that players complete from 11 to 17, ($Z = -11.469$, $p < 0.0001$). Users also played about 20% longer without coins ($M = 690$) than with coins ($M = 570$), ($Z = -3.004$, $p = 0.003$). We did not find a significant effect on return rate, $\chi^2(1, 4575) = 1.10$, $p = 0.295$. 10.6% of players with coins returned, compared to 9.6% of players without.

Further experimentation is necessary to explain why optional rewards do not improve player retention in our game in the way that we expected. Perhaps players try to complete these challenges even when they are difficult, and instead of giving up and continuing to play, they become frustrated and quit. Or, the presence of large numbers of achievements or optional challenges cause players looking for a quick, fun experience to go elsewhere, because the game looks more complex and difficult than it really is. Regardless, even modifying this non-essential gameplay mechanic affected player behavior more than any of our aesthetic experiments. In Hello Worlds, removing coins affected play time three times as much as animation, the strongest aesthetic effect.

CONCLUSION

We used large-scale A/B testing in two online casual games and found that music and sound had little effect on player behavior, while animations caused users to play more. We also found that a minor gameplay modification affected player retention more than aesthetic variations, although understanding the reasons for these effects and how they generalize to other games and genres requires further research. Our methodology can be used by other game designers to evaluate game prototypes and to determine the best use of resources. One can envision a development process that first tests the importance of key costly parts of the game as soon as the basic game mechanics are worked out, and then proceeds with development based on the findings of those tests.

We believe that games are valuable for gathering large numbers of participants for HCI research, as large-scale data is often expensive to collect. While creating games can be expensive, once they exist, they can attract large numbers of users at low cost. Since players are not aware that they are being experimented on, results obtained this way match more closely to their "true" behavior. For these reasons, experimentation within games may help researchers discover principles of player behavior and motivation. We have only scratched the surface of this potential by studying aesthetics and optional rewards in two online Flash games, and more nuanced tests in a wider variety of games will help develop more general theories.

ACKNOWLEDGMENTS

We thank Kongregate and the additional creators of Refraction and Hello Worlds: Ethan Apter, Brian Britigan, Eric Butler, Seth Cooper, Michael Eng, Mai Dang, Happy Dong, Jeff Flatten, Justin Irwen, Christian Lee, Marianne Lee, Emma Lynch, Stephen Sievers, and Blake Thompson. This work was supported by an NSF Graduate Fellowship, NSF grant IIS0811902, DARPA grant FA8750-11-2-0102, the University of Washington Center for Game Science, Adobe, Intel, and Microsoft.

REFERENCES

1. L. Alexander. Gamestop buys social gaming hub kongregate. *Gamasutra*, 2010.
2. E. Caoili. Farmville, top facebook games continue to shed users. *Gamasutra*, June 2010.
3. G. Cassidy and R. A. MacDonald. The effects of music on time perception and performance of a driving game. *Scandinavian Journal of Psychology*, 51:455–464, 2010.
4. K. Graft. Dice 2010: Zynga's reynolds on 'social' first and foremost. *Gamasutra*, 2010.
5. R. Kohavi, R. M. Henne, and D. Sommerfield. Practical guide to controlled experiments on the web: listen to your customers not to the hippo. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 959–967, New York, NY, USA, 2007. ACM.
6. M. Kurosu and K. Kashimura. Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In *CHI '95: Conference companion on Human factors in computing systems*, pages 292–293, New York, NY, USA, 1995. ACM.
7. S. D. Lipscomb and S. M. Zehnder. Immersion in the virtual environment: The effect of a musical score on the video gaming experience. *Journal of PHYSIOLOGICAL ANTHROPOLOGY and Applied Human Science*, 23(6):337–343, 2004.
8. S. Niedenthal. What we talk about when we talk about game aesthetics. In A. Barry, K. Helen, and K. Tanya, editors, *Breaking New Ground: Innovation in Games, Play, Practice and Theory: Proceedings of the 2009 Digital Games Research Association Conference*, London, September 2009. Brunel University.
9. J. Schell. *The Art of Game Design: A Book of Lenses*. Morgan Kaufmann Publishers, Burlington, MA, 2008.
10. P. Skalski and R. Whitbred. Image versus sound: A comparison of formal feature effects on presence and video game enjoyment. *PsychNology Journal*, 8(1):67, 2010.
11. N. Tractinsky. Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 115–122, New York, NY, USA, 1997. ACM.
12. D. Wesley and G. Barczak. *Innovation and Marketing in the Video Game Industry: Avoiding the Performance Trap*. Gower Publishing Ltd., Farnham, Surrey, England, 2010.