Using Spatial Game Analytics to Analyze Player Paths Through Games

Dennis Ramirez, University of Wisconsin-Madison
Matthew Berland, University of Wisconsin-Madison

Abstract: By combining pre-test and post-test measures with spatial gaming analytics, we will investigate how different player types move through a level, and how those differences can inform the design of educational games like *Fair Play*. Of particular interest were the differences between participants who had a high final bias compared to a low final bias, and participants who play games more than an hour a week, and those that do not. Initial analysis revealed differences in the way these groups played *Fair Play*.

Introduction

By connecting outcome variables with specific gameplay actions, we can make better educational games while improving the learning supports embedded in those games. By analyzing specific decisions that players make and connecting them with learning events, learning outcomes, and interpretable visualizations, we can better understand both learners and games. In this paper, we present a new approach for understanding how different sets of learners in *Fair Play* (GLS Studios, 2013) moved through the space and how that movement relates to various measures and outcomes.

Most contemporary educational games rely on pre-test/post-test differences to determine if players successfully achieved content goals. Unfortunately, this approach doesn't tell a game designer much about the value of specific design decisions (such as whether a boss is too difficult or if an Non Player Character, NPC, is positioned correctly). As we design educational games, we would also like to build upon and improve existing design principles (Barab & Squire, 2004), which we cannot do without considering data collected from gameplay. By correlating in-game positional data with outcome variable data, we believe we can identify gameplay patterns within the game that may influence, or predict outcome measures. In this paper, we will illustrate such a technique using the educational game *Fair Play*. We hope this analysis will also be helpful in other games that record positional data.

Spatial Game Analytics: heatmaps

A common problem when developing tools to analyze game data is that a successful approach in one game may not be useful in another. For example, the way players interact changes drastically between a first-person shooter like *Halo* (Bungie, 2001) and turn-based strategy games like *Fire Emblem: Awakening* (Intelligent Systems, 2012). When developing game analytics, there needs to be a balance between general structures used across games and those better suited to a certain game genre. Heatmaps, as an analytic visualization, tend to map to a wide range of games – most games display something that can be transformed into a heatmap (such as the UI), and most games have a representation of a player in the game itself. As such, the contribution of this paper is to present a modality of heatmap helpful for understanding learning through gameplay and to give an example of analysis performed with the heatmap. This type of analysis, referred to as Spatial Game Analytics, "serves as a strong explanatory power for deciphering and understanding player behavior" (El-Nasr, M. S., Drachen, A., & Canossa, A., 2013).

Often used by commercial game companies (Ambinder, 2009; Pruett, 2010; Niwinski & Randall, 2010), heatmaps are a visual representation of a player's movement inside of a game. A player's positional data is recorded and superimposed onto the game map, or UI. The resulting image is then color-coded for frequency (for example, highly traveled spots might be represented by warm colors, while low-activity areas might be rendered in cool colors). Applied to learning games like *Fair Play* (Figure 1), heatmaps can reveal patterns such as points of interest, points of attrition, unused resources, and popularity of NPCs.

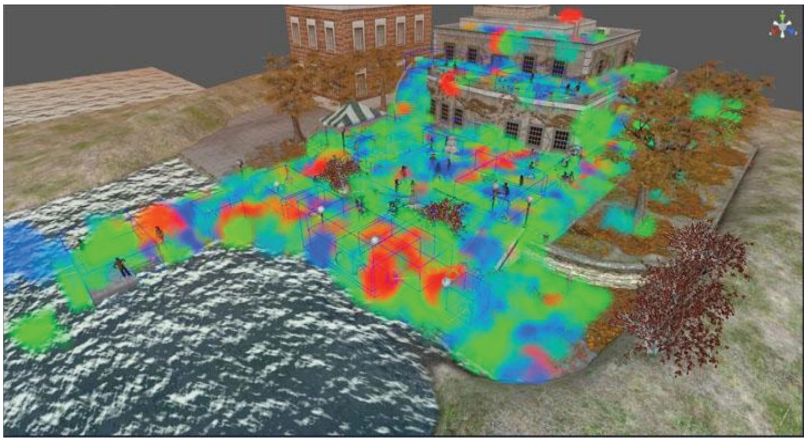

Figure 1: An example heatmap of player positions in *Fair Play*.

Learning Game Telemetry: ADAGE

With current technology, researchers can collect gameplay data with an unprecedented level of resolution. Every mouse click and player action can be recorded, stored, and analyzed. Working with these big data sets poses some problems, such as "how to determine which data are useful, and how to make use of this data in ways that will ultimately inform and improve student learning?" (Behrens, Mislevy, DiCerbo, & Levy, 2012) Effective game-based assessment must accommodate variations in the length, frequency, or content encountered, in order to measure changes in learning attributable to the game experience.

ADAGE is a system that allows users to collect data from games that implement the ADAGE API (Owen, Ramirez, Salmon & Halverson, 2014). An important feature of the ADAGE API is its flexibility to collect data from any player interaction we wish to study. This includes simple things like mouse clicks, and more complex things like changes to the game's internal state. While the ability to log data is helpful, it is by no means revolutionary. Researchers, and industry, have been collecting player data for years. What is helpful, however, is a standardized data framework that is open and available to all developers and researchers. By being open-source, the ADAGE API is more likely to be adopted across games and across institutions. This common language will allow for analysis that does not rely on a specific implementation, or game type, as long as it adheres to the common format. This in turn will allow the sharing of analysis methods that can be critiqued or refined resulting in a deeper understanding of a player's interaction with a system. In this study, ADAGE was used to collect information from players in real-time.

Method

Participants

The study included 58 people (26 female, 32 male, with 50% of the respondents age 18-25 and 37% 26-35) The participants self-selected from invitations sent to a large list of faculty, staff, and students from a large Midwestern university. All 58 of the participants included in this study completed the game, the demographic questionnaire, and the post-test. The participants were uncompensated.

Materials: *Fair Play*

Fair Play is a game created by the Games+Learning+Society Center at the University of Wisconsin-Madison that attempts to address implicit bias, or unconscious assumptions based on group stereotypes, in academia (Gutierrez, et al., 2013). The goals of *Fair Play* were twofold: The first was to explore the possibilities of using videogames as a vehicle for intervention of implicit bias; the second was to educate the general population on issues of implicit bias in academia. While the original project goal was to address gender bias in academia, the focus shifted to implicit racial bias in academia due to the wealth of recorded incidents/analysis indicating that the majority of individuals in the U.S. unconsciously prefer White individuals to Black individuals (Nosek et al., 2007; Nosek, Banaji & Greenwald, 2002). In this study, data was collected from players while they played the first level of *Fair Play*.

Assessment Instrument: Implicit Association Test

The Implicit Association Test (IAT) measures an individual's implicit bias by noting how long it takes the user to correctly associate a given word with a category given a prompt. The most common IAT is the Black/White, and Good/Bad IAT. During a standard IAT, the user is given an association (like black = good) and is then shown a series of words with good or bad intonations that the user must associate with the black or white photo shown. The test measures the delay between first showing the word and a successful classification. Generally, the longer it takes the user to correctly classify the prompt the more entrenched the opposite bias (Greenwald et al 2009). A measure of the user's implicit bias is calculated based on the deltas. A bias score of zero is generally considered to be unbiased, other scores indicate a preference towards one association over another, which is determined by a positive or negative magnitude. For this study, players took the IAT upon completion of *Fair Play*'s first level and the resulting score was used to measure the effectiveness of *Fair Play*.

Procedure

Participants played the first level of *Fair Play*, took the IAT upon completion, and finished with a general demographic questionnaire. During gameplay, information about the player's actions were recorded using ADAGE. After selecting variables of interest (how long a participant played games per week, and their final bias as reported by the IAT), the positional data collected was segmented to create various heatmaps. This visualization can help us to make sense of how players interact with the game allowing us build help systems, or refine the game itself.

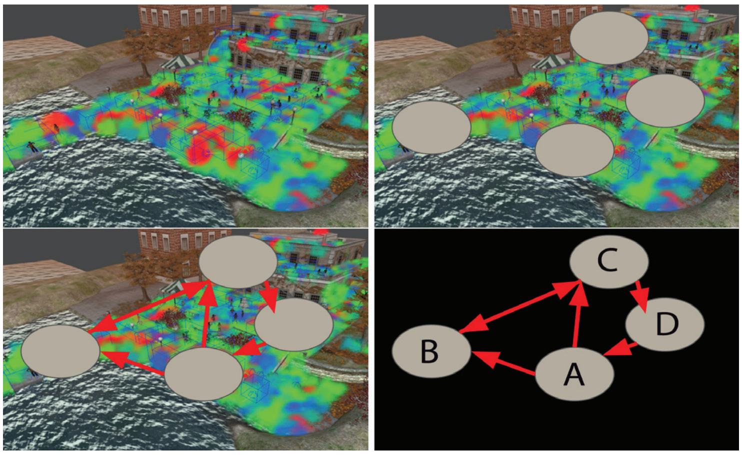

Figure 2: The process of converting heatmap data to a direct graph {A, B, C, D}.

Using heatmaps as a basis, we can also represent player movement as a directed graph. By clustering the heatmap data gathered we can identify areas within a game that players transition in and out of (Figure 2b). By reviewing player movement logs we can then determine if a player moves from one area to another and at what frequency (Figure 2c). By combining this information, we are left with a model of player transitions over time that we can use to make inferences about the player. By converting heatmap data into a directed graph we are also free to run graph analysis over the model to find structures of interest like cycles and most/least visited nodes. For example, in figure 3d we notice that the lower left area B can be completely avoided by visiting nodes A, C, and D. If we have an important event that occurs in B we might consider moving it to A, C, or D. We might also notice that most players quit the game if they are in area D which may prompt us to conduct follow-up interviews with participants to see why they quit in area D (perhaps there was a bug there). Of course, creating a graph to represent transitions does not need to be limited to player movement. We can also use this same method to analyze other state transitions like paths taken during an NPC conversation, or common actions performed by players that could indicate a common strategy. Like heatmaps, these types of analysis are applicable to most games, though the resulting models will be game dependent. By identifying patterns of play that correlate to preferred outcome variables, researchers can then modify their games to elicit that type of behaviour.

Experimental Design: Measures & Data Collection

Data collected includes information about the player's current position, conversation choices, the successful completion of game objectives, and general demographic information. Once gathered, the data is cleaned so that it's usable for our purposes. Specifically, we paid attention to any pre/post measures, and any positional data. Because we use the ADAGE API the time spent cleaning the data was minimized. The most important data to preserve for this purpose is x, y data simply because the resulting visualizations will be 2d, and because elevation isn't really important for Fair Play meaning the z dimension could be ignored safely. 3d data can be used with minor modifications, and It's also possible to collapse multiple dimensions to two dimensions through various dimension reduction techniques such as PCA.

A visualization of all player movement in the first level of Fair Play was produced by running a clustering algorithm over all positional data. In this study, we used a modified version of scikit-learn's K-means clustering (Pedregosa et al 2011). For the representations used in this study, the next step is to segment the 2d representation using the clusters identified. For this study, we used a modified version of SciKit's K-Means visual representation of handwritten digit data which segments a 2d projection by cluster (http://scikit-learn.org/stable/auto_examples/cluster/

plot_kmeans_digits.html). We created additional visualizations by segmenting the data along outcome measures and demographic information in order to compare and contrast the resulting groups. The two factors we focused on for this paper were the player's final IAT bias, and the number of hours the participant played games.

Results

We found differences in movement for players who would receive a low final bias (near zero, signifying little bias) when compared with players who had higher final bias. Players who self reported as having played more games also differed from players who reported playing little or no games.

When all participant data was entered into the heatmap algorithm, we received the visualization in Figure 3.

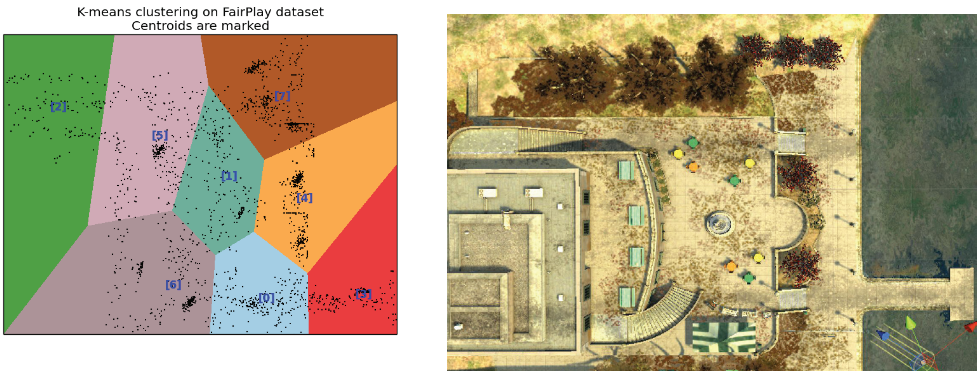

Figure 3: Heatmaps and clusters generated from all positional data (left) and a bird's eye view of the first level of Fair Play (right).

This representation outlined the first level of Fair Play and generally displays predefined areas of interest such as the location of NPCs.

The heatmap data was segmented based on the player's final IAT score. Figure 4a represents 29 players with the lowest final IAT score while Figure 4b represents the 29 players with a high final bias. The resulting visualizations show that players with low bias produced clusters similar to the aggregate while high bias players produced very different clusters. This shows that the distribution of positional data differed upper left and center of the map indicating a difference in the way players with a high final bias moved through the map.

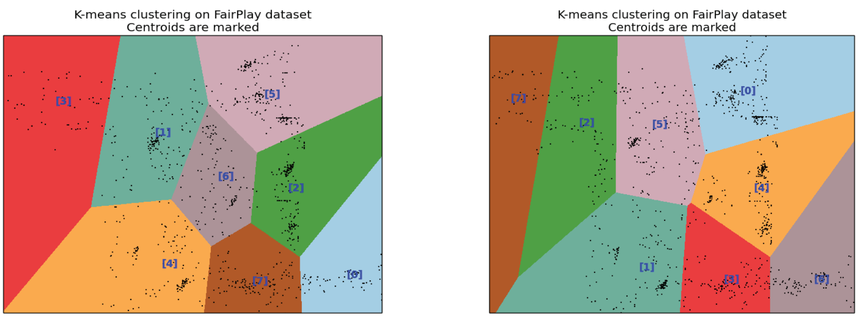

Figure 4: Heatmaps created by segmenting participant data based on final bias. Left represents participants with a low final bias, while the right represents users with a high final bias.

Directed graphs (Figure 5)produced from individual's positional data also support the claim that participants who displayed high bias during the IAT played the game differently than those who exhibited low bias.

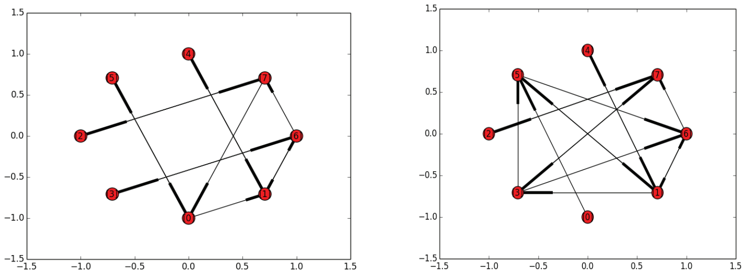

Figure 5: Directed graphs produced from player activity.

Another visualization (Figure 6) was generated based on how often the participant played games. 30 participants reported playing 1 hour of games or less, while 28 participants reported that they played more than an hour a week. While the clustering was more consistent than the visualizations produced by Low Bias / High Bias there were differences in the way the two groups move across the map. Specifically, participants who played more than an hour per week explored areas of the map that were not required while participants who played less than an hour per week did not.

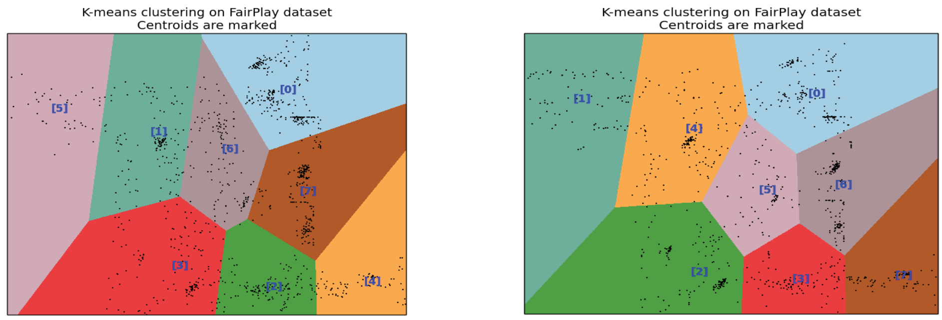

Figure 6: Heatmaps created by segmenting participant data based on hours of gameplay per week. Left, participants who reported playing <1 hour. Right, participants who played >1 hour.

Discussion

In this preliminary study, we used spatial analytics in conjunction with post-test measures and demographic data to analyze gameplay patterns in Fair Play. With this approach, we were able to identify areas of player interest and differences in the ways players moved through the first level.

Identifying Areas of Interest

The aggregate heatmaps successfully displayed the parts of the map that players gravitated towards. As expected, several of these points lined up with NPCs and other critical parts of the game. When the data was segmented by hours of games played per week, we found that participants who reported playing >1 hour per week were much more likely to explore parts of the map that weren't necessary to complete the first level. In our study, the area behind a building (Figure 7) was explored mostly by people who play more than an hour of games per week. This makes sense because there were no obvious indicators suggesting that part of the level was accessible. Further more, this part of the map wasn't necessary to complete the game and could safely be ignored. For the purposes of game design, placing an easter egg, or another NPC, in this area would be a good way to take advantage of the player's interest in a part of the map that isn't being used.

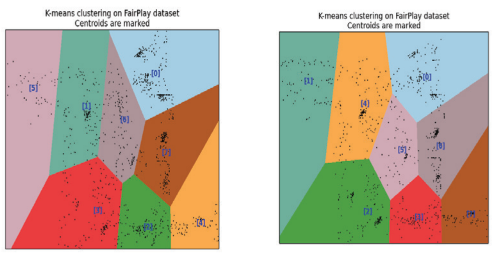

Figure 7: A close up of the upper left corners of the game hours per week condition indicating that participants who played >1 hour per week explored more of the map (left).

Differences in Player Types
The heatmaps generated from low bias and high bias participants differed in the distribution of points and the resulting clusters created. While data from the low bias players yielded heatmap with clusters similar to the aggregate, the high bias participants generated different clusters. Though the analysis was preliminary, by using heatmaps in conjunction with post-test data we were able to find slight differences between players who would have a high final bias and those that would have a low final bias.

The differences between players who played more than one hour a week of games is particularly interesting. Although most educational games target a wide audience, familiarity with the medium influences the way participants interact with the game (where they went, etc). Understanding these differences is important because it suggests that the way the game is presented (and the curriculum and structure necessary) may have to change depending on the participants.

Although the directed graphs generated support the hypothesis that gameplay can highlight differences between participants with high bias compared to participants with low bias, more work is needed to identify cycles and structures within these graphs. Future work will include researching how the resulting cycles and structures relate to, or influence, performance on post-test assessments.

Conclusion

Although there have been studies about the effectiveness of *Fair Play* as an intervention, this study represents the first steps towards evaluating the game with in-game data. This process not only gives us more insight into the relationship between actions made in the game and outcome measures, but also lets us evaluate the design decisions made during development. Because educational games have, built into them, biases that the designer holds about what is important to learn, and the best way to learn them (Squire 2011), being able to evaluate those design decisions is particularly important.

References

Ambinder, M. (2009). Valve's approach to playtesting: The application of empiricism. In *Game Developer's Conference*.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*(1), 1–14. doi:10.1207/s15327809jls1301_1

Behrens, J. T., Mislevy, R. J., DiCerbo, K. E., & Levy, R. (2012). An evidence centered design for learning and assessment in the digital world. In *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 13–53). Information Age Charlotte, NC.

Bungie. (2001). *Halo: Combat Evolved*. [Microsoft Xbox videogame]. Redmond, WA: Microsoft.

El-Nasr, M. S., Drachen, A., & Canossa, A. (2013). *Game analytics, maximizing the value of player data* (pp. 543–583). doi:10.1007/978-1-4471-4769-5

GLS Studios. (2013). *Fair Play*. [Broswer videogame].

Greenwald AG, Poehlman TA, Uhlmann EL, et al. Understanding and using the Implicit Association Test III: Meta-analysis of predictive validity. *J Pers Social Psychology* 2009; 97:17–41.

Gutierrez, B., Kaatz, A., Chu, S., Ramirez, D., Samson-Samuel, C., & Carnes, M. (2014). *Fair Play*: A video game designed to address implicit race bias through active perspective taking. *Games for Health Journal*.

Intelligent Systems. (2012). *Fire Emblem: Awakening*. [Nintendo 3DS videogame]. Kyoto: Nintendo.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit attitudes and beliefs from a demonstration website. *Group Dynamics: Theory, Research, and Practice*, 6(1), 101–115.

Nosek, B. A., Smyth, F. L., Hansen, J.J., Devos, T., Lindner, N.M., Ranganath, K, A., Smith, C.T., Olsond, K.R., Chugh, D., Greenwald, A., & Banaji, M.R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18(1).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830.

Pruett, C. (2010). Hot failure: Tuning gameplay with simple player metrics. *Game Developer Magazine*, *19*(9).

Squire, K. (2011). *Video Games & learning: Teaching and participatory culture in the digital age* (pp. 1–420). New York, New York, USA: Teachers College Press.