

How Playstyles Evolve: Progression Analysis and Profiling in *Just Cause 2*

Johanna Pirker¹, Simone Griesmayr¹, Anders Drachen², and Rafet Sifa³

¹ Graz University of Technology, Austria,
simone.griesmayr@student.tugraz.at, jpirker@icm.edu,

² Aalborg University & The Pagonis Network, Denmark
drachen@hum.aau.dk

³ Fraunhofer IAIS, Germany
rafet.sifa@iais.fraunhofer.de

Abstract. Evaluating progression of players in a game can take a variety of forms, but ideally combines playstyle or performance analysis with one or more aspects of progression, e.g. through a level- or mission-based structure. Furthermore, visualization of the results of analysis are essential to ensure that action can be taken on them. In this paper behavioral profiling through Archetype Analysis is combined with progression analysis, expanding on previous work in the area, and extending it into the context of Open-World Games. The proposed methodological framework is applied to the case of the action-adventure title *Just Cause 2*, focusing on the main storyline. The results show how players navigate the content of the title, and how some playstyles remain constant throughout the game, whereas others emerge or disappear with player progress. Additionally, player performance as a function of progression is evaluated across a number of key metrics.

Keywords: Game Analytics, Progression, Playstyle, Player Behavior, Cluster Analysis, Visualization

1 Introduction

Progression is a key element of game design, vital to balancing, and therefore an important focus of investigation in Game Analytics [16]. The concept of progression in game design can refer to a variety of components, however, adopting the definition of Lopez [15], generally refers to a pattern of advance towards a goal as well as the act of navigating towards the goal. This structure is important in ensuring the user experience and player engagement. Progression ties directly in with other elements of the game, including mechanics, rewards and risk, difficulty/complexity, etc.

From a behavioral analytics standpoint, evaluating progression can be performed in a variety of ways, but fundamentally combines performance analysis and measures of progression, e.g. through a level- or mission-based structure [2, 13, 16–18]. The performance component can be developed in a variety of ways

depending on the involved number of behavioral features and the specific questions asked of the analysis. A commonly adopted approach is profiling, which seeks to condense high-dimensional behavioral datasets into user profiles that contain the underlying patterns of variance in the data. Towards this, a number of studies have investigated approaches such as description, segmentation as well as machine learning, for the latter notably cluster analysis [2, 7, 9, 17, 18].

Being able to analyze and visualize progression is notably important in large development teams, where different stakeholders (designers, artists, programmers etc.) need to collaborate on constructing the indented progression ramp, and the need for checking if the design intent is implemented across all of the components of a game. Essentially the behavior of the player needs to be modeled on an ongoing basis as the player navigates through the content of the game, and results compared against design intent or combined with e.g. churn prediction models. This is the case for all digital games, irrespective of their genre, design or platform. A key element in progression analysis is acknowledging the dynamic nature of gameplay and the impact this has on progression, i.e. that as long as players are active, progression is taking place, either directly by moving through the content of the game, or indirectly, by considering and thinking about how to address particular challenges of puzzles. Progression inertia or even cessation is one of the major sources of player frustration, and therefore not a desirable situation [1]. Furthermore, visualization of the results of analysis are essential to ensure that action can be taken on them [13, 16, 19]. In the work presented here, behavioral profiling and progression analysis are combined towards evaluating player progression and visualizing analysis results.

1.1 Contribution

In this paper behavioral profiling through the clustering with Archetypal Analysis (AA) is combined with progression analysis, expanding on previous work in the area, and extending it into the area of Open-World (or Sandbox) Games. The proposed methodological framework is applied to the case of the major commercial action-adventure title *Just Cause 2*, using the main storyline of the game as a proxy measure of progression along multiple vectors. Results are visualized using Sankey flow diagrams, which permit visual inspection of the patterns in the behavioral data. The results show how players navigate the content of the title, and how some playstyles remain constant throughout the game, whereas others emerge or disappear as a function of player progress. Additionally, player performance as a function of progression and player profile are evaluated across a number of key metrics.

1.2 Just Cause 2

Just Cause 2 (JC2) is an action-adventure game set in an open-world environment, also commonly referred to as a *sandbox* game. Similar to other sandbox titles such as the *Elder Scrolls*- and *Grand Theft Auto*-series, it is characterized by free-roaming gameplay in a large 3D environment, in the current case

covering over 1000 square-kilometers. The game was published by *Square Enix* in 2010 and sold over six million units, retaining an active player base today. The gameplay is reminiscent of other action-adventure third-person shooters, with the player controlling the game protagonist as he travels around the Panau environment whether to complete missions, explore, collect items, or just have fun with the mechanics of the game. There are computer-controlled enemies everywhere, and a substantial part of the gameplay is combat-focused, which also means that combat performance metrics are important for the current aim of developing behavioral profiles.

A complete description of the gameplay of JC2 is out of scope of this paper, but in brief, JC2 is set in the fictional tropical island nation of Panau. Panau is ruled by a ruthless dictator. The player, taking on the role of Rico Rodriguez, an agent for an outfit referred to as the Agency, is tasked removing the dictator from power. This is done via progressing along different vectors notably the seven Agency-related missions, as well as missions from a number of Rebel Factions which helps wrest control from the dictators forces and into the hand of the factions. At the end of each mission chain for each Rebel Faction is a Stronghold mission, which form some of the most difficult tasks in the game. Upon completion of a Stronghold mission, the control of an administrative district is gained by the local Rebel Faction.

Progression in JC2 follows a directed gameplay approach [15], where all mechanics in the game are available from the beginning of the game, but that the player is gradually encouraged to use and become familiar with the mechanics of the game progressively. Overall, progression in the game towards reaching the point where the dictator can be toppled, is determined by *chaos*, a measure of the overall efforts of the player towards destabilizing Panau, whether through completing faction missions, attacking military posts, searching for and blowing up status of the dictator, etc. Thus the overall chaos achieved by a player provides a proxy measure of progress across the various progression vectors in the game. Furthermore, Agency missions unlock with chaos progression. Throughout the game, the player is via missions encouraged to navigate around Panau, which can be done using a wide variety of vehicles operating in the water, on land and in the air. Barring a few missions players can travel via any means they can capture. The player also has a grappling hook and reuseable para-sail/parachute which can be used for aerial movement as well as in combat.

2 Profiling and Progression: Related Work

Behavioral profiling in games has been explored by a number of researchers across academia and industry, providing a number of case studies showcasing how profiling can be performed using different techniques and in varied game contexts [4, 7, 9, 10, 14, 16].

There are particular challenges associated with behavioral profiling in major commercial game titles, notably related to volume, variety, and volatility [5]. These problems have been approached using a variety of techniques from de-

scription and segmentation to machine learning. In the latter category, cluster algorithms [8] and other dimensional reduction techniques have been broadly applied across different game formats. Cluster analysis is an exploratory approach to building behavioral profiles, and aims at identifying groups for players with similar behaviors and identify the most important behavioral features in terms of the underlying patterns in the dataset [4, 9, 6, 14].

The work presented here forms an extension of earlier work on behavioral game research, notably: Sifa et al. [17], who investigated player progression in the action-adventure, single-player *Tomb Raider: Underworld*, and Drachen et al. [2], who investigated player progression in the browser-based Massively Multiplayer Online Game (MMOG) *Glitch*. Sifa et al. [17] utilized Simplex Volume Maximization (SIVM) to develop behavioral profiles of the seven levels of *Tomb Raider: Underworld*, finding some behavioral profiles to be present across the levels, others to be present in only some of the levels. They noted that the variations in the profiles present for a level appeared related to the specific challenges and gameplay of these levels. Drachen et al. [2] adopted a similar approach, but focused on time rather than in-game segments. The authors developed profiles on the auction house behavior of players using k-means clustering, and used Sankey diagrams to develop interactive visualization of the pattern of players moving into and out of profiles across monthly time bins. This visualization approach is adopted here. The work in the above mentioned publications in turn rests on other research in game analytics focused on behavioral profiling in games, e.g. [4, 6, 14].

3 Dataset and Features

The dataset was provided by Square Enix, the publisher of JC2. The data comprises complete play histories from over 5000 JC2 players who played the game in 2010. A vast amount of behavioral features were collected including the in-game geographical coordinates for all player actions registered, as well as timestamps for these events. The dataset contains a detailed set of metrics from the gameplay, e.g. the total kills, the total chaos gathered by the player, the kilometers driven or the number of stronghold takeovers the player completed, and more.

3.1 Feature Definition and Pre-processing

The gameplay of JC2 gives rise to a huge variety of potential behavioral measures, across spatio-temporal navigation, combat performance, progression through the main storyline and numerous side quests to general exploration. This means that progression analysis can target a number of dimensions in the game, however, for all the free-roaming gameplay, JC2 features a central storyline, experienced across seven Agency missions. Progressing along these missions in turn require the player to reach specific levels of Chaos, which in turn means completing Rebel Faction missions and causing mayhem and destruction in general, which means that the Agency missions work as a proxy for multiple vectors of

progression in the game, and it was therefore chosen to make these the vector of progression in the work presented here.

In order to keep the profiling task manageable, a subset of features was selected based on the core mechanics of the game, following the principles of Drachen et al. [6], and the focus on the main storyline. This means for example including measures of distance traveled and locations discovered, which informs whether players focus on straight navigation between story points or are exploratory in their behavior, as well as combat performance metrics such as kill/death ratios (K/D ratio) and player death frequencies, and finally features related to more advanced mechanics such as base jumping, stunt driver points and the variety of vehicles used. The specific feature set chosen here does not impact on the analytical framework, which can be employed for any feature-set in any game context, but of course impacts on the kinds of conclusions that can be derived from analyzing them [2, 17]. The features are described in Table 1.

Table 1: Overview of the behavioural features used for AA

Feature	Description
Total playtime	Total time played until the completion of the corresponding agency mission
Chaos/minute	Chaos gathered per minute (Chaos increases by completing missions, destroying buildings).
Kill/Death Ratio	(K/D ratio), the number of kills the player has scored divided by the number of deaths suffered
Kills/minute	Number of kills the player has scored per minute.
Deaths/minute	Number of deaths the player has suffered per minute.
Hijackings	Number of times the player hijacked a vehicle
Base jumps	Number of times the player used the game feature base jumps.
Stunt driver points	Number of points the player gathered by performing stunts with vehicles.
Stronghold takeovers	Number of completed stronghold takeovers.
Kilometers driven	Distance covered by the player in kilometers.
Locations discovered	Number of discovered locations.
Locations completed	Number of completed locations.

In further pre-processing, all players with scores outside the 1-99th percentile were excluded from the analysis, in order to remove extreme outliers. Outliers can in other contexts be the focus of study, but in the current case the most extreme outliers were caused by faulty tracking and similar errors, and therefore excluded. The behavioral features were normalized using regular variance normalization, following the principles of [4, 9].

4 Analysis

We applied AA models to all seven agency mission bins. For each the optimal number of clusters (k) was determined. To identify the number of clusters for the AA, we analyzed the residual sum of squares for all k values less than or equal to 20 and chose the number of clusters by utilizing the elbow criterion. The residual sum of squares indicated that the data consists of three main archetypes.

4.1 Progression Profiling via Cluster Analysis

Profiling game player behavior is an important practice in game analytics to obtain insights about how players interact with the game [9, 3]. Unlike the commonly used predefined-segmentation based profiling methods, clustering based profiling approaches reveal more interpretable and actionable insights about the game play behavior. Cast as low-rank matrix decomposition problems, many clustering methods can be grouped under the same formalism for better comparison. In this work we will concentrate on two factor decompositions as the methods we used here follow that scheme. That is, given a data matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, our aim is to find two lower rank matrices $\mathbf{P} \in \mathbb{R}^{m \times k}$ and $\mathbf{C} \in \mathbb{R}^{k \times n}$ such that $\mathbf{X} \approx \mathbf{P}\mathbf{C}$. For the purpose of profiling, the factor matrix \mathbf{P} contains the prototypical players that define the profiles and are selected by special procedure enforced by the algorithm whereas the matrix \mathbf{C} contains the mixing or belongingness coefficients indicating how much a player belongs to the particular profiles defined by \mathbf{P} [3, 9]. It is important to note that under this framework we can categorize the algorithms with respect to their constraints. For Archetypal Analysis [11] \mathbf{P} is constrained to contain extreme vectors that are defined as convex combinations of particular data-points whereas \mathbf{C} is constrained to be column stochastic. Additionally, keeping the stochasticity constraint for \mathbf{C} , for a more relaxed archetypal decomposition such as Simplex Volume Maximization, \mathbf{P} is selected to be a datapoint lying on the data-simplex [12]. For more information about constrained matrix decomposition and clustering we refer the reader to [3, 9, 17].

In this study we present a profiling analysis with two components: (1) Behavioral profiling of the JC2 players via the application of archetype analysis (by analyzing the resulting \mathbf{P} -matrices) and the (2) generation of a Sankey diagram-based visualization (by using the resulting \mathbf{C} -matrices) for the temporal behavior of the players across the defined profiles.

4.2 Player Profile Descriptions

The analysis with AA shows that the player profiles change over the different missions. Archetypal analysis provides an indication of the number of clusters between 4 and 6. Table 2 gives an overview of the distribution of the various main profiles among the different missions. One particular player group ("Low Performers") stands out in mission 2-3.

With AA we can identify eight main clusters, which appear and disappear in different missions and can have slightly different metrics depending on the mission-design. **Average Player:** This player cluster only appears in mission 6 and describes players with overall average values. **Drivers:** Players that drove the most kilometers and also performed the most base jumps and black market orders (cluster appears in mission 2). **Elite Players:** Players that played on the highest difficulty level with the highest kill/death ratio, for the rest of the features average values. They have an average exploration performance. In later missions (3 and 4) they also have higher values in hijacking cars and stunt driving points and less black market orders, base jumps (mission 5) and completed below average faction missions and stronghold takeovers (missions 6) and have a small number of kilometers driven (mission 7). **Explorers:** This player type is very constant and can be observed in slightly different forms over all missions. This cluster can be described as players with the slowest pace, a kill/death ratio above average, but the least amount of chaos collected per minute. They hijacked the most vehicles and discovered the most locations. In mission 2 they've already played in average three times longer than other player types and collect most stunt driver point. From mission 3 they have played at least twice as long as other player types and completed far most faction missions and stronghold takeovers and have discovered and completed nearly three times as many locations. **Low Performers:** Players with the worst kill/death ratio of all player types and most deaths per minute. Average values for the rest. One can also observe average or slightly above average performance regarding exploration and driving (mission 3-7). **Low Performance Drivers:** This cluster only appears in mission 1. Players with a fast pace, a very low kill/death ratio and most deaths per minute, but collected a lot of chaos per minute. They drove with many different vehicles and on average drove more kilometers than other clusters. **Rushers:** This player group is also very constant and appears in all missions but mission 6. It described players that completed the game the fastest. They collected the most chaos per minute, achieved a high kill/death ratio and the most kills per minute. They care little about exploring the world and discovered and completed the least locations (mission 3-5). **Stunt Drivers:** This group is only observed in the first mission and describes players who collected the most stunt driver points, drove with more different vehicles than the other player types, and also drove the most kilometers. Skill scores pretty similar to player type "Low Performance Drivers".

4.3 Player Progression along the Missions

Fig 1 gives an overview of various metrics, which can give the players feedback on their progress and/or their performance. *Progress* describes the a pre-defined in-game metric to describe the actual progress along the game. *K/D-ratio* describes the ratio between kills and deaths. In particular missions three and four seem to be "rich of action". In particular the last two missions seems to require the players to drive a lot according to the average metrics *Kilometers driven*. *Stunt driver points* can be collected by extreme driving activities. In particular in early missions more new locations (*Locations discovered*) are discovered.

Table 2: Overview of main AA types by missions (in %)

Type	Mission						
	1	2	3	4	5	6	7
Average Players	-	-	-	-	-	44.57	-
Drivers	-	2.27	-	-	-	-	-
Elite Players	19.00	-	3.69	6.15	12.43	21.71	14.74
Explorers	2.84	3.37	2.78	4.49	4.36	10.86	3.91
Low Performers	1.23	70.62	55.81	43.42	45.04	22.86	47.97
Low Performance Drivers	18.10	-	-	-	-	-	-
Rushers	56.23	23.74	37.72	45.94	38.17	-	33.38
Stunt Drivers	2.61	-	-	-	-	-	-

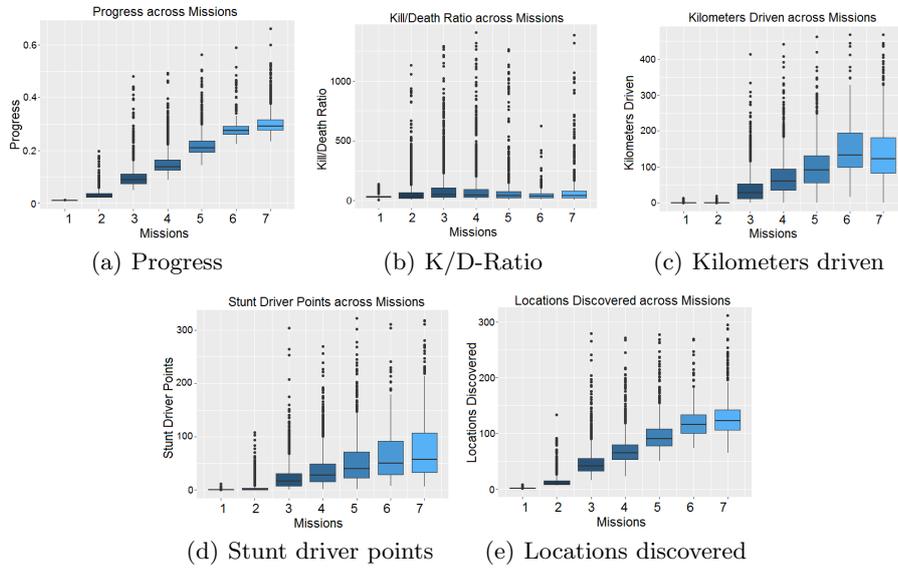


Fig. 1: Player progression over the missions based on different in-game metrics.

4.4 Visualizing player behavior along the main story line

Sankey diagrams consist of nodes, in our case our clusters, and links, the flow of players between these clusters. The diagram was developed using an rCharts implementation of the d3.js sankey plugin for the R language⁴. Hovering the mouse over any link or node shows the respective player count and information about new and departing players. Fig 2 illustrates how players change their playstyles (their player profile) over the seven missions and gives also a very good overview of player-groups that stop, how they split in subgroups, or merge.

⁴ <https://github.com/timelyportfolio>

As an interactive visualization, Sankey diagrams are an powerful and easy-to-use tool supporting different forms of behaviour analysis of players.

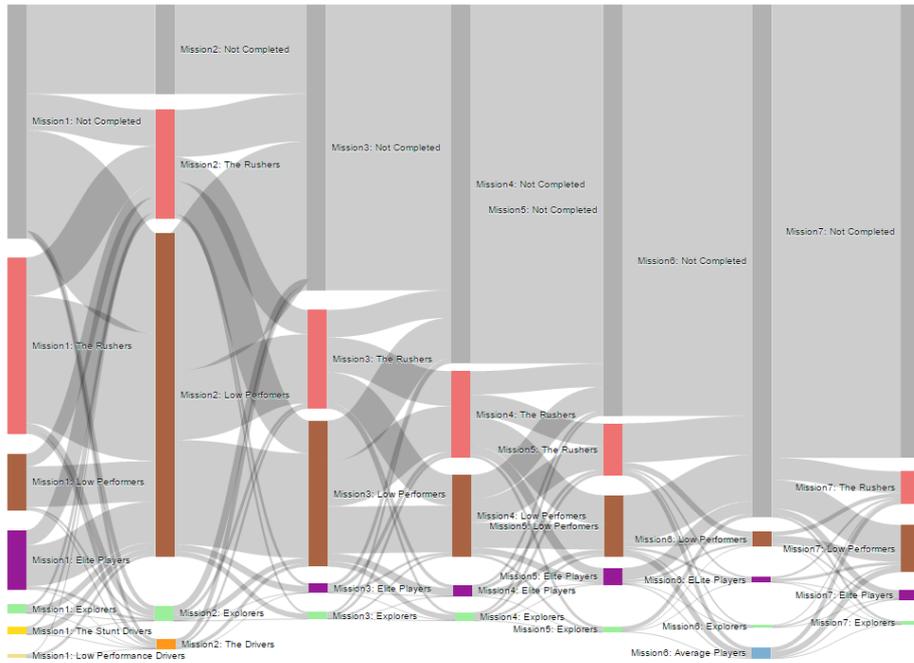


Fig. 2: Sankey Diagram based on archetypal analysis

5 Results

Based on the performed analysis and visualizations different results with focus on player behaviour along the game progress can be presented.

5.1 Player Behaviour over the Missions

The analysis clearly indicates variations in the player behaviours over the duration of the game, despite the consistent presence of the three main clusters throughout all agency mission bins. Players do not remain in a single cluster when they progress through the game, they move relatively freely across clusters from agency mission to agency mission. While the dominance in the exploration based features stays the same for the Cluster "Explorer", the other features, like the skill features change over the course of the main storyline from agency mission to agency mission. Another quite consistent feature for all three clusters

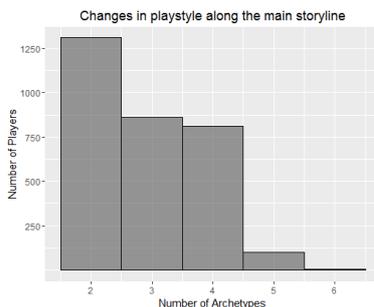


Fig. 3: How players change clusters along the main storyline: all players change the cluster at least once

is the total time played, the "Explorer" cluster yields in the highest values of total play time across the whole main story line.

How many profiles enter players on average over the course of the game? The analysis indicates that all players change their in-game behaviour along the storyline. As illustrated in Fig. 3 all players change the cluster at least once. On average players enter 2.91 clusters over the course of the game.

How does in-game behaviour and performance change over the various missions? The analysis and the visualization clearly indicate that player behaviour changes over the different missions, which is also due to the nature of the mission design. The Sankey Diagram (Fig. 2) illustrates how players change the player types along the course of the seven missions.

How can we describe player behaviour of the different player profiles? Fig 4 refers to different in-game metrics and how these metrics change over the seven missions for the different main player profiles as described above. *Progress* is a metric, which is in particular high for Explorers. *Kill/Death-Ratio* is a distinct feature for Elite players. As illustrated in subfigure (c) the *total time played* (in minutes) is particular high for Explorers. Starting from mission 2 they start to strongly separate from the other clusters. This value is very low for Rushers. *Stunt driver points* are also very high for Explorers, but also for the group of Drivers. *Factions Missions* are popular for Explorers, but interestingly, only in Mission 6, also for the Low Performers and Elite Players.

6 Conclusion and Discussion

In this paper behavioral analysis has been presented for *Just Cause 2*, focusing on the relationship between playstyle and progression in the game. It is to the best knowledge of the authors the first time an OWG has been the focus of progression analytics, and provides insights into the behavior of players in these kinds of titles. The results and visualizations directly expand and refine the approaches of [2] and [17] via combining OWGs, Archetype Analysis,

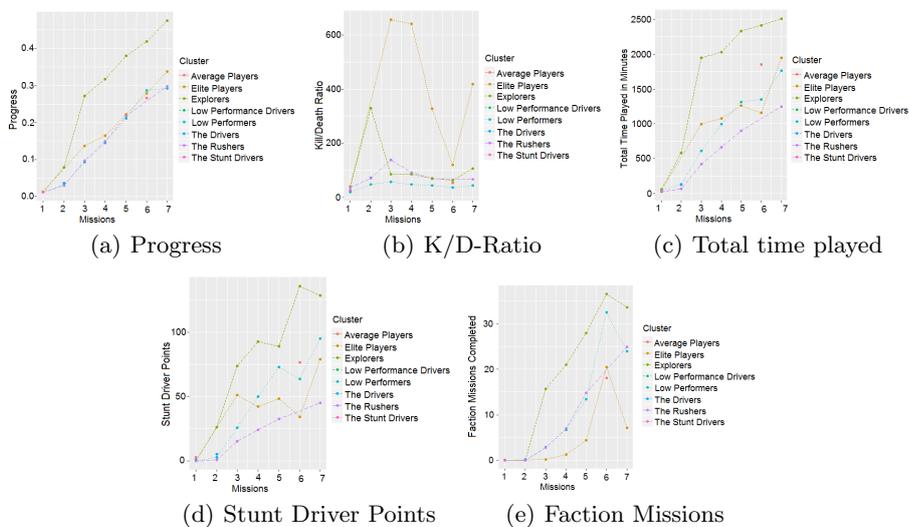


Fig. 4: Player progression over the missions based on AA clusters.

Sankey visualizations and in-depth exploration of key behavioral metrics. The progression of the player clusters exhibiting similar playstyles has been explored across a number of behavioral metrics related to progression in an OWG, such as discovered locations, K/D ratios or distance traveled. The approach provides a template for similar analyses in other games for evaluation of behavior and game design. Future work will explore different perspectives of progression and extend the work presented here to other games. The work presented here focuses initially on using the main story line of JC2 as a proxy measure of progression, expanding to integrate other metrics. However, there are many viewpoints that could be adopted on player progression in games. For example, progression can also mean gradual acquisition and upgrading of weapons, developments of player skill, or exploration of the environment. Similarly, in other games, progression is a viewpoint that can be applied in a range of situations. Irrespective of the game and the specific aspect of progression studied, it is a key aspect of game design and vital to balancing. The work presented here combines profiling and progression towards enabling the detailed evaluation of player progression.

Acknowledgments The authors would like to express their sincere gratitude to Square Enix for making the *Just Cause 2* dataset available for analysis.

References

1. A. Canossa, A. Drachen, J. R. M. Srensen, Arrgghh!!! Blending Quantitative and Qualitative Methods to Detect Player Frustration, Proc. FDG, 2011.

2. A. Drachen, S. Baskin, J. Riley and D. Klabjan, Going Out of Business: Auction House Behavior in the Massively Multi-Player Online Game Glitch, *Journal of Entertainment Computing*, 5, pp. 20-31, 2014.
3. A. Drachen, and R. Sifa, and C. Thureau, and C. Bauckhage, A Comparison of Methods for Player Clustering via Behavioral Telemetry, in *Proc. FDG*, 2013.
4. A. Drachen, R. Sifa, C. Bauckhage, and C. Thureau, Guns, swords and data: Clustering of player behavior in computer games in the wild, in *Proc. IEEE CIG*, 2012.
5. A. Drachen, C. Thureau, J. Togelius, G Yannakakis and C. Bauckhage. Game Data Mining. In (M. S. El-Nasr, A. Drachen and A. Canossa, Eds.) *Game Analytics: Maximizing the Value of Player Data*, Springer Publishers, 2013.
6. A. Drachen, A. Canossa, and G. Yannakakis, Player Modeling using Self-Organization in Tomb Raider Underworld, in *Proc. IEEE CIG*, 2009.
7. A. Normoyle, and S.T. Jensen, Bayesian Clustering of Player Styles for Multiplayer Games, in *Proc. AAAI AIIDE*, 2015.
8. C. Aggarwal and C. Reddy, Eds., *Data Clustering: Algorithms and Applications*, Chapman & Hall/CRC, 2013.
9. C. Bauckhage, A. Drachen and R. Sifa, Clustering game behavior data, in *Transactions on Computational Intelligence and AI in Games*, 7(3), pp. 266-278, 2015.
10. C. Bauckhage and R. Sifa, k-Maxoids Clustering, *Proc. KDML-LWA*, 2015.
11. C. Cutler and L. Breiman, Archetypal Analysis, *Technometrics* 36.4, 1994.
12. C. Thureau, K. Kersting and C. Bauckhage, Yes we can: simplex volume maximization for descriptive web-scale matrix factorization, in *Proc. ACM CIKM*, 2010.
13. G. Wallner, S. Kriglstein, An introduction to gameplay data visualization, in: *Game Research Methods*, pp. 231-250, ETC Press.
14. K. Shim and J. Srivastava, Behavioral Profiles of Character Types in EverQuest II, in *Proc. IEEE CIG*, 2010.
15. M. Lopez, *Gameplay Design Fundamentals: Gameplay Progression*, Gamasutra, Nov. 26, 2006, <http://www.gamasutra.com>
16. M. Seif El-Nasr, A. Drachen, and A. Canossa, *Game Analytics - Maximizing the Value of Player Data*, Springer, 2013.
17. R. Sifa, A. Drachen, and C. Bauckhage, Behavior Evolution in Tomb Raider Underworld, in *Proc. IEEE CIG*, 2013.
18. R. Sifa, F. Hadiji, J. Runge, A. Drachen, K. Kersting, C. Bauckhage, Predicting Purchase Decisions in Mobile Free-to-Play Games, *Proc. of AAAI AIIDE*, 2015.
19. W. S. Cleveland. *Visualizing Data*. Hobard Press, 1993.