Analyzing User Behavior in Digital Games [PRE-PRINT]

Anders Drachen
IT University of Copenhagen, Denmark

Alessandro Canossa
IT University of Copenhagen, Denmark

ABSTRACT

User research in digital game development has in recent years begun to expand from a previous existence on the sideline of development, to a central factor in game production, in recognition that the interaction between user and game is crucial to the perceived user experience. Paralleling this development, the methods and tools available for conducting user research in the industry and academia is changing, with modern methods being adopted from Human-Computer Interaction (HCI). Ubiquitous tracking of player behavior and player-game interaction forms one of the most recent additions to the arsenal of user-research testers in game development and game research. Player behavior instrumentation data can be recorded during all phases of game development, including post-launch, and forms a means for obtaining highly detailed, non-intrusive records of how people play games, and by extension their user experience. Behavioral analysis is a relatively recent adoption to game development and –research; however, it is central to understanding how games are being played. In this chapter, the current state-of-the-art of behavior analysis in digital games is reviewed, and a series of case studies are presented that showcase novel approaches of behavior analysis and how this can inform game development during production. The case studies focus on the major commercial game titles *Kane & Lynch: Dog Days* and *Fragile Alliance*, both developed by *IO Interactive/Square Enix Europe*.

INTRODUCTION

Computer games have evolved from simple text-based adventures like *Colossal Cave* and *Akalabeth* to virtually photo-realistic renditions of virtual worlds with advanced mechanics, spreading across a dozen or more genres, offering an increasing number of entertainment opportunities (Bateman & Boon, 2005). This development is to no small degree driven by the evolution of gaming devices, the hardware platforms upon which games software is loaded, are also becoming more and more diverse, and thanks to the increasing connectedness of e.g. mobile networks, users are finding digital games accessible everywhere. The increased complexity of digital games – in terms of the amount of possible user actions and –behaviors that they afford, as well as the breath of interaction options between the user and the software/hardware –, the diversity and the distribution across different hardware devices (Lazzaro & Mellon, 2005; Mellon, 2009; Pagulayan, Keeker, Wixon, Romero & Fuller, 2003), are among the important factors driving an increased focus on the users, the players, of digital games, in the game development industry. Contemporaneously with the development in game design, user-research and user-oriented testing has become progressively more important to industrial development and quality assurance (Kim et al., 2008; Pagulayan, Keeker, Wixon, Romero & Fuller, 2003). The purpose of user-oriented game testing is to evaluate how specific components of, or the entirety of, a game is played by people; allowing designers to evaluate whether their ideas and work provides the experience they are designed for. User-oriented testing is useful in game production, because the perceived quality of a digital game product is generally related to the perceived user experience. Therefore, content testing is receiving increasing attention from industry and academia alike (e.g. Isbister & Schaffer, 2008; Jørgensen, 2004; Kim et al., 2008; Nørgaard & Rau, 2007).

Methods adopted from Human-Computer Interaction (HCI) (Hilbert & Redish, 1999; Kuniavsky, 2003) have begun to replace the traditional informal testing approaches used in game development and game research, with e.g.

usability, playability and user behavior forming keywords in contemporary user-oriented testing and –research (Davis, Steury & Pagulayan, 2005; Isbister & Schaffer, 2008; Medlock, Wixon, Terrano, Romero & Fulton, 2002; Pagulayan, Keeker, Wixon, Romero, Fuller, 2003). Different methodological approaches have different weaknesses and strengths, with e.g. qualitative approaches being excellent for acquiring in-depth feedback from players (users) but requiring substantial resources. In comparison, qualitative approaches are generally better suited for larger participant groups, but less suited for in-depth analysis or study of user behavior and –experience. Given the limited resources of industrial testing, a considerable focus has therefore been aimed towards the quantitative methods.

The automated collection and analysis of **game metrics data** forms one of the new quantitatively-based approaches that have in recent years been adopted from e.g. software development (Renaud & Gray, 2004) to serve in digital game development (Drachen & Canossa, 2009a; Kim et al., 2008; Swain, 2008). Game metrics covers not only **player behavior** (in-game behavior, player interaction with the different components of the game systems, community behavior, customer service evaluation), but also **performance** issues (e.g. server stability, monitoring changing features) and **processes** (turnaround times of new content, blocks to the development pipeline, etc.) (Mellon, 2009; Mellon & DuBose, 2008). **Player metrics**, a form of instrumentation data, are formed by logs or counts of users interacting with the game software, and is notable for being unobtrusive to collect (Blythe, Overbeeke, Monk & Wright, 2004; Dumas, 2003). The term metric should not be confused with the term heuristic (Desurvire, Caplan & Toth, 2004). Heuristics are design principles which assist in guiding game design, whereas metrics are instrumentation data logged from game software.

The player metrics data recorded from digital games, and how they are data mined and analyzed (Kennerly, 2003) varies depending on the stakeholders involved. For examples, at the *management* level, it is of interest to know what install languages that customers (users) have used, for how long they have played a specific game, how many that completed the game or gave up partway there or for example activity levels on game servers. *Community managers* can be interested in information about how users interact with the game website, and in being able to provide e.g. heatmaps (Drachen & Canossa, 2009a; Thompson, 2007) or play statistics to the player community. *Game researchers* can be interested in any type of metrics data, depending on the specific purposes of the research project (Williams, Consalvo, Caplan & Yee, 2009; Williams, Yee & Caplan, 2008). *User-research/Quality Assurance* experts are conversely interested in the actual behaviors expressed by the players, either during game testing or post-launch. Within the context of user-oriented testing, instrumentation data related to player behavior (user-game interaction) are generally referred to as **gameplay metrics** (Swain, 2008; A. Tychsen & Canossa, 2008). Gameplay metrics form objective data on the player-game interaction. Any action the player takes while playing can potentially be measured, from low-level data such as button presses to in-game interaction data on movement, behavior etc. Gameplay metrics data can for example be used to locate design problems or evaluate player behavior patterns (Drachen, Canossa, & Yannakakis, 2009; Kim et al., 2008).

Gameplay metrics can be considered similar to **User-Initiated Events** (UIEs) (Kim et al., 2008), i.e. actions taken by the user, for example moving their game avatar forward, picking up a virtual object, interacting with an AI-controlled entity, or similar. Importantly, UIEs can also be formed by low-level actions such as keystrokes, which are indirectly relatable with user behavior inside digital game worlds. Since playing a computer game is formed as a series of action-reaction cycles, with the player/-s taking an action, and the game software responding, it is sometimes also necessary to consider **Game-Initiated Events** (GIEs), i.e. the actions taken by the game software, either independently of the user or as a response to an UIE. For example, if a player shoots at an AI-controlled agent in shooter-type games such as *Deus Ex, Quake* and *Unreal Tournament*, the AI-agent will initiate aggressive behavior towards the player.

In this chapter, the current state-of-the-art of behavior analysis in digital games is reviewed, and three case studies from the major commercial titles *Kane & Lynch: Dog Days* [KL2] (IO Interactive) and *Fragile Alliance* [FA] (IO Interactive) are presented that showcase different novel forms of behavioral analysis performed via the application of gameplay metrics, for example, it is shown how to take advantage of the spatial dimension of behavior data. The case studies presented are specific to two games (covering single-player and multi-player environments), but are focused on generic behaviors such as player death, navigation and skill, which are common to many games or virtual environments featuring player-controlled characters/avatars (e.g. shooter-type games such as: *Doom 3*, *Unreal Tournament*, *Bioshock* and *Crysis*, and Role-Playing Games/Adventure games such as *Oblivion*, *Neverwinter Nights* and *Dreamfall.* Massively-multiplayer online games such as *Age of Conan* and *World of Warcraft* and online persistent worlds such as *Second Life*, have players taking control of a single avatar/character, and therefore also form potential targets for the behavioral analyses presented. The chapter is in part based on previous work, nominally: "*Towards Gameplay Analysis via Gameplay Metrics*", in Proceedings of the 13th International MindTrek Conference © ACM, 2009; DOI: http://doi.acm.org/10.1145/1621841.1621878. This chapter presents new case study material, updated state-of-the-art, and new sections on e.g. metrics categories and combined methodological approaches.

The case studies are all derived from research carried out in collaboration between the industry and academia, and this is reflected in the case studies being examples of user behavior analysis carried out in practice, with the purpose of evaluating gameplay. The case studies showcase behavior analysis in mid-production phases, expanding on previous work by considering new forms of behavioral data and including multivariate analysis, moving beyond the state-of-the-art. The case studies also focus at the detailed analysis of the behavior of few players, a subject that is virtually nonexistent in the current literature. The case studies are used to build an argument for the usefulness of detailed behavior analysis at the detailed level, in game design-, game production- and game research-oriented contexts. The methods can be directly extended to and applied in the context of, other forms of digital interactive entertainment.

STATE OF THE ART

Compared to the extensive literature available on instrumentation-based user behavior analysis in general software development contexts (De Kort & Ijsselsteijn, 2007; Hilbert & Redish, 1999; Hurst, Hudson & Mankoff, 2007), it may be surprising that there is only limited knowledge available on the use of game metrics for development and research.

Within the academia, publications are separated into those targeting the analysis of behavior in virtual environments in general (Börner & Penumarthy, 2003; Chittaro & Ieronutti, 2004; Chittaro, Ranon & Ieronutti, 2006), and those targeting digital games applications and user behavior inside the virtual worlds of games (Drachen & Canossa, 2008, 2009a, 2009b; Drachen et al., 2009; Ducheneaut & Moore, 2004; Ducheneaut, Yee, Nickell & Moore, 2006; Hoobler, Humphreys, & Agrawala, 2004; Kim et al., 2008; Southey, Xiao, Holte, Trommelen & Buchanan, 2005; Thawonmas & Iizuka, 2008; Thawonmas, Kashifuji & Chen, 2008; Thurau, Kersting & Bauckhage, 2009). Some studies fall outside of these two categories, e.g. (Burney & Lock, 2007) who used simple player metrics to compare user performance in a planetarium dome with conventional flat-screen environments. Within location-aware games, for example those using mobile phones as the hardware platform, some work has been carried out tracking user location (e.g. Coulton, Bamford, Cheverst & Rashid, 2008), which in principle can be applied to virtual environments. Additionally, Thawonmas, Kashifuji & Chen (2008) evaluated behavior analysis in order to recognize AI-driven bots.

Working from a sociological perspective, (Williams, Consalvo, Caplan & Yee, 2009; Williams, Yee & Caplan, 2008) have used game metrics to compare user-report data with actual usage data from the Massively Multi-Player Online Role-Playing Game (MMORPG) *EverQuest 2*, collaborating with *Sony Online Entertainment*. The focus of this research has however been focused across the virtual-world real-world divide, looking at e.g. gender roles or comparing reported number of hours per week played with actual number of hours per week played, rather than player behavior inside the game worlds.

In work related to user behavior analysis, for example (Nørgaard & Rau, 2007) obtained a game metrics via eye tracking, applying the data to user-oriented testing. Gaze tracking also holds potential as an interaction tool (see e.g. (San Augustin, Mateo, Hansen & Villanueva, 2009). This leads into the field of psycho-physiological measures in game evaluation (Nacke, 2009), which also form high-resolution quantitative data about the users, but about the user experience, not user behavior. The potential merging of these to emergent research directions is discussed below.

Within the field of adaptive gaming and AI, related work has been carried out using simple games and a mixture of psycho-physiological measures and user-data (Yannakakis & Hallam, 2007; Yannakakis & Hallam, 2008) or even using outright gameplay metrics in combination with neural network techniques (Drachen et al., 2009). (Southey et al., 2005) used data from the digital game FIFA 99´ to analyze various sweet and hard spots in terms of goal scoring, aiming to semi-automate gameplay analysis, collaborating with the major publisher EA Games. The use of behavioral data for adaptive gaming was highlighted by the major commercial title *Left4Dead* (2007, Valve), which features an AI-director that utilizes player behavior to control various challenge-features, e.g. the number of enemy zombies. It should be noted that the use of basic tracking of player actions to guide e.g. the responses of AI-agents is a stable of contemporary games (Redding, 2009).

In general, a substantial amount of research relevant to behavior analysis in digital games exists in fields such as HCI, visualization and ethnography (Fua, Ward & Rundensteiner, 1999; Hilbert & Redish, 1999; Hurst et al., 2007; Kort, Steen, de Poot, ter Hofte & Mulder, 2005; Kuniavsky, 2003; Peterson, 2004), and it is also from here that most user-oriented methods applied in game development and game research are adopted and adapted (Davis, Steury & Pagaluayan, 2005; Jørgensen, 2004; Kim et al., 2008; Laitinen, 2005, 2006; Medlock, Wixon, Terrano, Romero & Fulton, 2002; Pagulayan & Keeker, 2007; Pagulayan, Keeker, Wixon, Romero & Fuller, 2003).

From the industry, publicly available knowledge about the use of game metrics and gameplay metrics specifically, is rare because metrics data and the associated analyses are treated as confidential information, and therefore not publicly available. The information available about industry practices is therefore limited to a handful of conference presentations (King & Chen, 2009; Lazzaro & Mellon, 2005; Ludwig, 2007; Mellon, 2004; Mellon & DuBose, 2008; Mellon & Kazemi, 2008; Romero, 2008), industry whitepapers (Mellon, 2009), blogposts (Grosso, 2009); online popular articles for specialist magazines and websites (DeRosa, 2007; Goetz, 2006; Kennerly, 2003; Sullivan, 2006; Thompson, 2007) and reports in research articles (Medler, 2009). Game metrics are also mentioned in several game design/development books. For example, Byrne (2005) discusses player metrics in the sense of the abilities of the player characters, using these as design parameters in level deisgn, for example the walk and run speed, height and width, jump distance and interaction distance of player characters. He discusses how different games feature different metrics, e.g. powerups, temporary modifiers, user-definable metrics, etc.

According to Medler (2009), both Microsoft, Maxis and Sony track players using recording systems, and analyze the data using in-house developed analytic tools. However, where some details are known about the approaches of Microsoft (Kim et al., 2008), there is limited knowledge about Sony´s approach. Some development companies chose to partly share the collected data with the player community, for example in the form of diagrams and activity charts. The Steam service provides a macro-level overview of information recorded across different games offered via this service, providing an opportunity for the player community to gain a glimpse of some of the game metrics being collected (see: http://store.steampowered.com/stats/). Similarly, Maxis has allowed the player community access to data collected from Spore, providing the means for user-initiated data analysis (Moskowitz, Shodhan & Twardos, 2009). For games developed using the Flash platform, websites such as mochibot.com and nonoba.com provide analytics tool for developers.

There exist a few specialist companies that perform game/gameplay metrics-based analysis for game developers, to greater or lesser degrees, e.g. *Orbus Gameworks*, *Emergent Technologies* and *Nielsen Games*, indicative of the increasing need for metrics-based analyses in the game industry (Mellon, 2009). A foundational aspect of game metrics work is data mining, which forms the core underlying methodology for recording, preprocessing, integrating, selecting and transforming data for analysis. Data mining techniques, visual analytics and statistical methods are used to analyze the data, which are ideally presented in a format suitable to the targeted stakeholder (which could be the player community). Importantly, while data mining and analytics can provide knowledge from the collected game metrics data, it cannot provide prioritization of results, e.g. deciding which problems to correct first. Essentially, data mining semi-automates the process a process of knowledge discovery (Han, Kamber & Pei, 2005; Hand, Heikki & Padhraic, 2001; Kennerly, 2003).

Focusing on the work carried out on user behavior within virtual worlds (games or otherwise), a substantial part of the current published material stems from the work of *Microsoft Game Labs*, which perform game testing and user-oriented research for the various Microsoft-based game studios. *Microsoft Game Labs* developed e.g. the TRUE setup to capture gameplay metrics together with survey and video data, utilizing this during the user testing of e.g. *Halo 3* and *Shadowrun* (Kim et al., 2008; Thompson, 2007).

Metrics data have notably been applied in the context of Massively Multiplayer Online Games (MMOGs) and similar persistent, social virtual environments (Isbister & Schaffer, 2008; Lazzaro & Mellon, 2005; Mellon, 2009; Mellon & DuBose, 2008; Mellon & Kazemi, 2008), where they form a source of **business intelligence** to the development companies. These game forms are highlighted because they provide a continual service over a period of years (for example, the MMOG *World of Warcraft* has been running since 2003), continually during this period needing information about the users.

When reviewing the available knowledge on behavior analysis in games it is evident that the majority is focused on **character-based games**, i.e. games where the player controls a single character/avatar, which forms the main vehicle for interaction between the player and the game world. For example, the case studies presented by (Drachen & Canossa, 2008, 2009a, 2009b; Drachen, Canossa & Yannakakis, 2009; Hoobler, Humphreys & Agrawala, 2004; Kim et al., 2008) all focus on character-based games. Working with virtual worlds in general, (Chittaro & Ieronutti, 2004; Chittaro, Ranon & Ieronutti, 2006) similarly worked with avatar-based data in the VU-flow application for visualizing movement patterns in virtual environments. From the industry side, (DeRosa, 2007) reported from the use of time-spent reports developed at *Bioware* for the major commercial game *Mass Effect*, similarly a character-based game. Given that character-based games make up if not the majority, then a substantial chunk of the major commercial game titles, this bias is perhaps not surprising. It means however that there is virtually no knowledge about how to perform player behavior analysis in game forms such as Real-Time Strategy (RTS) games, where players control multiple units (generally military); and Turn-Based Games (TBG) such as the *Civilization*-series. Within the work on character-based games, there is a bias towards multi-player or massively multi-player games forming the main focus of behavioral analysis (Ducheneaut & Moore, 2004; Ducheneaut, Yee, Nickell & Moore, 2006; Kim et al., 2008; Mellon, 2009; Thawonmas & Iizuka, 2008; Thawonmas, Kashifuji & Chen, 2008; Williams, Consalvo, Caplan & Yee, 2009; Williams, Yee & Caplan, 2008), in comparison to work on single-player titles (e.g. DeRosa, 2007; Drachen, Canossa & Yannakakis, 2009; Nørgaard & Rau, 2007; Tychsen & Canossa, 2008). Considering MMOGs specifically, it is evident from reports such as (L. Mellon, 2009) and conference presentations such as (King & Chen, 2009; Lazzaro & Mellon, 2005; Swain, 2008) that these "social games" form a basis for an emergent trend in **Metrics-Driven Development**, replacing traditional informal approaches; however, the specifics of behavior analysis and data-mining player activity in these games is unavailable beyond the cases mentioned in these reports and presentations. Data-mining player behavior and – activities in MMOGs is common in the industry, because these games require constant monitoring of the player base due to the persistent nature.

The few the few examples of systems developed for capturing game metrics-like data developed outside of the academia are generally targeted at Virtual Environments (VEs) rather than games specifically (Börner & Penumarthy, 2003; Chittaro & Ieronutti, 2004; Chittaro, Ranon & Ieronutti, 2006). Additionally, these applications are often targeted at analyzing specific features, such as movement, or developed for use with specific applications/games [e.g. 14], and therefore not particularly flexible to accommodate the varied needs of behavior analyses across neither game genres, nor portable across application environments.

Furthermore, the literature on the use of game metrics (in general, not just for behavior analysis) is largely based on data from in-house testing, with the exception of online/social games, where data are commonly derived from installed clients or game server (Williams, Consalvo, Caplan & Yee, 2009; Williams, Yee & Caplan, 2008) and from character-based games (Drachen, Canossa & Yannakakis, 2009). Published work is also somewhat biased towards high-level aggregate counts of specific metrics, although for example the publication of heatmaps (game level maps showing the locations of player character death aggregated for a number of players) and player statistics (e.g. for purposes of ranking) are becoming more and more common as community feedback tools, e.g. for the game *World in Conflict, Half Life 2* and *Team Fortress 2.* Available behavioral evaluations appear to be generally oriented towards single-variable analyses, e.g. level completion times, and spatial analyses showing behavior inside the virtual environment of games are rare metrics (e.g. the position of the player within the virtual environment) (Drachen & Canossa, 2009a, 2009b; Hoobler, Humphreys & Agrawala, 2004).

In summary, there is therefore a substantial room for development of novel approaches towards utilizing gameplay metrics for behavior analysis. There is a need to open up the discussion about how to utilize gameplay metrics analysis in game production and –research, and to broaden the available knowledge beyond the predominant MMOGs and MMORPGs. In this chapter, a beginning towards addressing some of these issues is attempted, presenting detailed multivariate case studies of behavioral analysis from two major commercial game titles, both single-player and multi-player.

LOGGING THE RIGHT BEHAVIORS

As outlined above, the majority of the published work on behavior analysis in digital games, and the work on VWs, is focused on situations where the user controls a single avatar (or character), in a 3D environment. Typically, the digital games represented are First-Person Shooters (FPS) or Role-Playing Games (RPGs). The games can be either single- or multi-player.

Despite this focus on character-based games, behavior analysis would appear to be useful to all forms of games: First of all, the practice of behavior analysis has a strong tradition within software development and website design, where, as noted above, the approach is applied in a variety of situations and contexts. It is therefore likely to expect the method to be equally valuable in a variety of games-related contexts. Secondly, behavioral analysis is already carried out in the game industry on a variety of games, even if the majority of the available published work is restricted to games where the player controls a single avatar or character. Nevertheless, behavior analysis is important to the industry, not just for the purpose of user-research/user-testing, but also in relation to e.g. community feedback. Within MMOGs, data-mining player behavior is vital to the industry because of the persistent nature of these games, requiring constant monitoring of the player base (Mellon, 2009).

A general problem is that research work and practices carried out in the game industry rarely make it to any publication media, whether professional or academic. What little published knowledge there is emerges in industry conferences, events and –publications. The current published work is generally case-based and often fairly shallow in the analytic depth.

The general immatureness of behavior analysis in digital game contexts, and the problems with freely publishing the results of these analyses, which would generally be considered proprietary information by the companies involved, means that it is difficult – at this stage - to provide a consensus about how behavior analysis should be carried out, and to provide guidelines or advice as to which gameplay metrics that it makes sense to track in different kinds of situations.

Another barrier for this is that games vary in their design. It is therefore challenging to provide frameworks or guidelines for how to choose which metrics to track, applicable across all games. Digital games vary even within specific genres such as Role-Playing Games, First Person Shooters, Real-Time Strategy Games etc. Due to the variation, and the number of stakeholders that potentially are involved in an analysis (researchers in academia/scientific institutions; and marketing-, management-, community management-, design departments and user-research/game testing experts in the context of development/publishing companies), the questions asked will similarly vary on a case-by-case basis.

If the focus is narrowed down to *gameplay* metrics (UIEs and GIEs) specifically, and thus behavior analysis, ignoring the broader range of game metrics that find use in business intelligence, it is possible to define three broad categories of metrics, as a function of their generality. These categories do not provide specific guidelines about which metrics to track when, but provide an initial step in this direction.

1) Generic gameplay metrics: There are some features that are shared among all games, generally high-level user information such as total playing time, number of times a game has been played, the real-world time that has elapsed from a game was started until it finished, the ratio of players completing the game vs. those who gave up, playing time per level/segment of the game, total number of player actions, etc. These types of gameplay metrics are typically not useful for detailed gameplay analysis, but excellent for aggregated summaries of player activity. This would also be the kind of high-level gameplay metrics that is relevant to cross-departmental reports.

2) Genre specific gameplay metrics: Although the term "genre" is nebulous at best within a digital game context, it is however widely used in both industry and academia, to describe sets of games that share specific features (REFS). Irrespective of how digital games with shared feature sets are grouped, the advantage in terms of behavior analysis is that these potentially can carry over between games within the group. For example, Drachen & Canossa (2008, future play paper), defined four categories of UIEs and one category of GIEs applicable in character-based games:

1) Navigation metrics: Covers navigation of the character in the game environment
2) Interaction metrics: Covers interactions with objects and entities of the game, initiated by the player via the character.
3) Narrative metrics: Covers navigation through a game storyline, for example quest completion data or navigation through story branches.
4) Interface metrics: Covers all interactions with the game interface, either while playing the game or during periods spent interacting with other interfaces of the game (e.g. game setup).
5) Event metrics: Covers all Game Initiated Behaviors (GIEs), e.g. the actions of NPCs, activation of cut-scenes, or similar.

3) Game specific gameplay metrics: These are associated with the unique features of the specific game. In essence, the unique features lead to game-specific questions associated with user-testing. For example, in a game such as *Batman: Arkham Asylum*, the usage pattern of the main character´s abilities could be of interest, e.g. for evaluating if there are abilities that are over- or under-used.

There exists – to the best knowledge of the authors – no published, generic process for identifying which metrics to track in order to address a specific research question or user-test. Which metrics to track generally flows from the question asked, however, how to log the information (as a frequency? Event-based?) can be difficult to determine, notably during the early phases of a game production or research project. Two examples highlight this difficulty:

1) During the development of *Fragile Alliance* (IO Interactive), the designers wanted to know how long players survived before being killed the first time. This situation shows the ideal situation where the question being asked provides the information necessary to build the analysis to answer it: Survival time can be done via a simple tracking of the timing (in seconds) between each kill event, as a function of a specific player or player tag.

2) The designers were also interested in knowing the ranges at which the different weapons implemented in the game were fired. This question poses a challenge as compared with the first example. Again, the required information is straight forward (distance between shooter and target); however, how to log and analyze the information requires some consideration. First of all, *Fragile Alliance* is a team-based shooter game, meaning that literally thousands of shots can be fired in during a game session! If each shot being fired is to be tracked – in detail – a substantial amount of transmission bandwidth (for transmitting the logged data) and storage space would be required. Furthermore, not all shots hit the intended target. It would be relatively safe to assume that a shot hitting a legal target (e.g. one player shooting another on the opposing team), has a high chance of being intended – it could happen that a lucky player trying to hit a legal target by accident hits another legal target, however, it could be assumed (given the mechanics of *Fragile Alliance*) that the likelihood of this event is small enough to be of minimal influence on a statistical analysis. However, this leaves all shots that do not hit a legal target – for example, shots hitting a wall, car or other game-world object. Should these be included in the analysis? The problem with these shots is that the objects being hit are likely not the indented targets (unless the object in question is destructible and there is a point doing this). Including these in the analysis will therefore not inform about the ranges with which players intend to use specific weapons. Using probability analysis, it is possible to estimate the intended target of a spread of bullets from e.g. a submachine gun, however, this type of evaluation is computationally too cumbersome to employ in the context. In the current situation, the approach chosen was to record the position of the shooter and the legal target being hit, and the weapon type used, for each instance where the shot resulted in a death event only. This provides a compromise between obtaining all data and not overloading the bandwidth requirements. Using simple triangulation, it is based on the recorded data possible to calculate the mean range that each weapon is used, as well as useful parameters such as standard deviation. Using an associated time stamp, the temporal development in the selected variables can also be evaluated.

CASE STUDIES

The case studies presented below form practical examples of behavior analysis in a digital game context. In the first case study, data from a series of playtesting sessions are included, in the second; the focus is on detailed analysis of the gameplay experience of just a few players. The case studies are focused on features such as death, navigation and environment interaction that are generic to character-based games. The approaches described should therefore be broadly applicable to these types of games, which arguably form the majority of major commercial titles along with games from the Real-Time Strategy (RTS) genre; as well as avatar-based social environments featuring 3D environments, e.g. *Second Life*.

Data for the case studies presented are drawn from the Square Enix Europe (SEE) Metrics Suite, developed by the SEE Online Development Team hosted by IO Interactive. The Suite is engineered towards collecting game metrics from SEE produced games, both in-house as well as from various online services, such as the Xbox Live! Service. When performing spatial analyses on player behavior data, preprocessed data are imported into a geodatabase system. From this, data are extracted, plotted, analyzed and visualized using either a Geographical Information System (Longley, Goodchild, Macquire & Rhind, 2005) or a custom tool developed at IOI, *QuickMetrics*. The GIS permits in-depth analysis of spatial metrics, where QuickMetrics is suited for rapid visualization of basic event-driven variables or player navigation data.

It is important to note that the gameplay metrics data used in the case studies are drawn from in-house testing during mid-development of KL2 and FA. This causes two problems: 1) Some absolute numbers are confidential and cannot be reported (percentages are used instead); 2) The data are not obtained from user-research sessions where the experimental conditions are controlled. The data were recorded during playtesting sessions run at the user-research department of *IO Interactive*. The lack of controlled conditions means that there is a risk of bias in the data – i.e. that testers played differently than they normally would. However, given that a controlled laboratory setup would also present players with a different playing environment than that which they normally operate in, it is difficult to avoid this assumption. The exception is studies using remotely connected playdata from users playing in their native habitats (Drachen & Canossa, 2009b; Williams, Consalvo, Caplan & Yee, 2009; Williams, Yee & Caplan, 2008). However, this was not possible in the current context because KL2 and FA are not published games. It should be noted that studies using data from in-house playtesting are fairly common in the literature, as it forms one of two primary data sources for games-based user testing (Kim et al., 2008), with the other being remotely collected data from game servers, e.g. MMOG-servers (Drachen & Canossa, 2009b; Drachen, Canossa & Yannakakis, 2009; N. Ducheneaut & Moore, 2004; Ducheneaut, Yee & Moore, 2006; Williams, Consalvo, Caplan & Yee, 2009; Williams, Yee & Caplan, 2008).

Case study 1: Level analysis by sub-sector in *Fragile Alliance*

In the first case study the focus is on showcasing how in-depth player behavior analysis operates in practice during mid-late development, when a vertical slice is playable but undergoing iterative user-testing in order to ensure the design works as intended (Isbister & Schaffer, 2008; Medlock, Wion, Terrano, Romero and Fullerton, 2002; Pagulayan, Keeker, Wixon, Romero & Fuller, 2003). The setting is a development company but could just as easily be an academic research group. *Fragile Alliance* is a multi-player, online shooter-type game (Figure 1). He players either play as mercenaries trying to accomplish a specific mission, such as a bank robbery; or as police officers trying to prevent this. However, all players start as mercenaries, with the police team being comprised of AI agents. If a mercenary dies, they respawn (are reinstated in the game universe) as police officers, working alongside the AI agents. Apart from the risk of being killed by the police, mercenary players also face the risk of being betrayed by other mercenaries. In *Fragile Alliance*, mercenaries can betray each other, and steal each other's loot. If for example a mercenary player had managed to secure a sum of money from a bank vault, another mercenary could kill the first, and steal his/her money. If a mercenary kills another mercenary he becomes a "traitor" but is allowed to keep all the money collected without sharing (Figure 1). This game mechanics is designed to shift the balance of power from initially being on the side of mercenary team, towards the police (AI and players), as more and more mercenaries are eliminated. After the second death, the player will typically not respawn, but will have to wait for the game round to end (usually after a few hundred seconds depending on the map). A game session will typically consist of multiple rounds being placed on the same map (scenario) and/or different maps, similar to comparable games such as *Unreal Tournament* and *Counterstrike*. The winner of a round is the player who leaves with the most money, irrespective of how these are obtained. Police players earn money by securing money from mercenaries.

Figure 1: Screenshots from Fragile Alliance. (Left) Showing a Traitor clearly marked to the remaining mercenary players. (Right) Teams of mercenaries and police engage (© 2008, IO Interactive. Used with permission).

During development of *Fragile Alliance*, the designers of the game were interested in evaluating if this shift in balance manifested in the actual behavior of the players. A series of playtests (over 100) were run using vertical slices of the game (a specific level map), including 129 players (in-house and external testers). Data from the game sessions were extracted from the SEE Metrics Suite. The dataset included 8943 kill events, with data about the role of the player being killed, who the killer was, the position of both players, the type of weapon used and a timestamp for the event metric. A time-series approach was adopted, with data binned into 15 segment sections, and percentage distributions between the different killer roles calculated. The result showed that the majority of the kills in the game were caused by mercenaries, up to approximately 75-90 seconds of play. For example, from 30-45 seconds, 48% of all kills were caused by mercenaries, and only 35% by the police (of which 27% were by AI-agents). After 90 seconds of play, the pattern changes. Collectedly after the 90 second mark, mercenaries account for 35% of the kills, while the police team caused 55% of the kills (8% of which were kills by AI-agents). The remaining percentages were taken up by specialist roles such as Undercover Cops and Dirty Cops, who are available in specific scenarios. Traitor mercenaries generally did not figure as a major influence in the kill distribution, generally causing 2-7% of the kills. Collectedly, players caused 70% of the kills, the AI-agents 26% - enough to be felt but not enough to wrest control from the players. The remaining 4% were suicides, e.g. players standing too close to exploding object such as cars (Figure 2).

Figure 2: Basic kill statistics for Fragile Alliance, based on 8943 kill events from a series of game sessions including 129 different players. (Left) Distribution of kills by AI-agents or players. (Right) Causes of death by killer role.

Analyzing kill statistics temporally does not address the spatial patterning of the events. Given the focus of Fragile Alliance on scenario-based levels, i.e. the player teams have specific missions beyond elimination of the opposing force, it is essential to ensure that players progress through the level, so that the right events occur in the right locations. For example, in the vertical slice used in the current case study (Figure 3), the objective for the mercenary players to reach the exit and complete the mission. In at least some games, it would therefore be desirable to have the mercenaries reach the exist area. Spatial analysis of gameplay metrics is a powerful tool for analyzing player behavior, as the can be plotted with pinpoint accuracy. This allows fine tuning of gameplay balance.

In the vertical slice used here, a specific level from *Fragile Alliance*, the mercenaries spawn in the bottom of the map, the police AI agents to the top right (Figure 3). The objectives of the mercenaries is firstly to enter a vault, located to the left in the map, and thereafter to reach the level exist, in the top right corner, behind the spawning area of the police. The game level consists of four major sub-sectors: The **spawning area**, where the mercenary players enter the game (red in Figure 2). The **vault area**, where the money they need to steal are located (green in Figure 2), a **subway station** area approximately in the middle between the spawning areas of the two teams (yellow in Figure 2) and finally an area at street level (orange in Figure 2), through the rightmost side of which the mercenary players must go through if they want to escape (Figure 1, right).

Figure 3: The Fragile Alliance vertical slice (game level) divided into sub-sections: Bottom area = spawning area; Middle area = subway; Left area = vault area; Top area = road/exit area. (Left) Locations where police officers were the cause of death events. A broad distribution is apparent indicating that police officers can reach the entire map. (Middle) Locations of suicides. (Right) An example of feedback to the game designers of Fragile Alliance. The level map shows the distribution of about 250 player death occurrences overlain the level map, and has added explanations to guide the interpretation of the map. The map is developed using a GIS.

Combining visualization of the spatial behavior of players with statistics of their temporal (and spatial) behavior permits a more in-depth analysis of the player behavior (Figure 4). Comparing the spatial and temporal behavior shows for example that mercenary players generally turn traitor either in the beginning of the game in the spawning area sub-sector, or later in the game in the road/exit area. Traitors are typically killed in the spawn area (61.25%), but rarely in the road/exit area (8.81%), which indicates that it is a much more risk-filled endeavor to turn traitor early in the game rather than later (it should be noted that further analysis showed that mercenaries turning traitor outside of the spawning area rarely move into the spawning area again – by this point the action has moved to the other segments of the map).

For the mercenaries, the majority of the kills occur in the spawning area sub-sector, where mercenaries enter the game (Figure 4). The AI-agent kills are spread across the entire map, indicating that their search & destroy behavior is working. Suicides occur in the vast majority of cases (76.04%) in the road/exit area, where a series of cars are placed which can explode if taking too much damage. A smaller part takes place in the metro station area, where players can be hit by metro trains while crossing the tracks coming from the vault to the exit/road area to the north in the map (Figure 3). The analysis resulted in designers adding warning noises to the subway trains and increasing the health of cars, in order to bring down the number of deaths caused by these two game features.

In terms of the roles played by players when they are killed, the pattern is generally as was intended in the game design. Police (players and AI) are generally killed in the road/exit area where they spawn (69.32%), and very few are killed in the spawning- and vault areas, where instead the mercenaries are under pressure from the police (44.17%).

Figure 4: (Left) Frequency distribution of the causes of player death as a function of the sub-sector. (Right) Frequency distribution of the player role at the time of death as a function of the sub-sector. Same data source as for Figure 1.

A somewhat larger amount of death events occur in the spawn area than intended by the game design, which could indicate that mercenaries are perhaps a bit too eager to turn traitor early in the game. This could be a gameplay problem, however, it may not necessarily be a user experience problem – the players may find great enjoyment in following this behavior. In order to properly address this question, user experience measures need to be employed (see below). As the game level is iteratively refined and tested, different solutions can be attempted.

The approach to analyzing player behavior in the above example is based on *Fragile Alliance* alone, however, the game is representative of a large number of multi-player shooter games, e.g. *Call of Duty, Team Fortress 2* and the *Battlefield*-series, which have proliferated notably since the release of *CounterStrike*. The methodology outlined is therefore directly transferrable to these games. In other game forms, e.g. single-player games (even non-shooters), causes of death may not be other players and AI-agents, but perhaps environmental effects. The principle of the analysis remains identical, however, and the interest in locating "trouble spots" in game levels/areas where the behavioral patterns are not as intended is also common in user-oriented testing in game development (Drachen & Canossa, 2009a; Isbister & Schaffer, 2008; Kim et al., 2008).

Case study 2: Frustration in Kane & Lynch: Dog Days

Kane & Lynch: Dog Days is a shooter-type game currently in development at IOI. In terms of gameplay, the game follows previous shooters in that the player controls a single character and mainly has to worry about staying alive, eliminate enemies and solve specific tasks. In this case study, the game experience of a single player is investigated in detail with a focus on investigating the causes of frustration exhibited by the player. Frustration would normally be characterized as an unwanted component of user experience (Brave, 2003; Gilleade, 2004; Klein, 2002; Norman, 1988); however, frustration forms a recognized component of the experience people can obtain from playing digital games (Hazlett, 2006; Ijsselsteijn, 2007; Pagulayan, keeker, Wixon, Romero & Fuller, 2003; A. Tychsen, Newman, Brolund, & Hitchens, 2007; Vorderer, 2006). The case study showcases the potential of behavioural data to enable modelling of the navigation of players through a game environment. This is of key interest to game designers because it allows them to observe how their games are being played. User-oriented methods such as playability testing (Davis, Steury & pagulayan, 2005; Pagulayan, Keker, Wixon, Romero & Fuller, 2003) can also locate gameplay problems, however, when integrating gameplay metrics in a data collection suite, it becomes possible to model the second-by-second behaviour of one to – simultaneously - thousands of players.

There are many ways to visualize navigational data. There exists various applications for handling data visualizations, which are flexible enough to handle a variety of contexts. In the current case, a Geographic Information System (GIS), build using the package ArcGIS, was created to provide spatial analysis (Drachen & Canossa, 2009a; Longley, Goodchild, Macquire & Rhind, 2005). Another possibility is to visualize the data directly in the game editor (Figure 5). This form of visualization allows experimenters to see through the eyes of the player in a manner similar to a video recording of a game session, but with the added benefit of having recorded metrics mapped within the game environment, and draw on quantitative results from these.

Figure 5: Gameplay metrics data plotted directly in the game editor. The position of the player was tracked every second (points). Associated metrics tracked included the camera angle at the time (light grey cones), and whether the player character was taking damage (darker grey color). The green line represents the path of the player, calculated as vectors between the tracked positions.

Similar to the first case study, the type of analysis reported here is placed in mid-late development, where at least a vertical slice of a game in production is available. In the first case study, the research questions driving the analysis were pre-defined. This case study is an example of how questions can arise via user-testing in an industrial development (or empirical research) context: The study was made possible due to a serendipitous series of events: During play-test sessions of the KL2, the game´s programmers delivered a version of the game in which a check-point malfunctioned forcing players to repeat a fairly long and challenging segment of play within a specific game level (Figure 6). During the play-test sessions, a user research expert at IO Interactive observed a test participant, who considered himself fairly proficient, become more and more angered as he failed to complete a level of the game, dying several times in the same area of the game level, and having to restart at an earlier point in the game as compared to where he died. The participant manifested frustration through body movements, facial expressions and verbalizations.

Figure 6: Top-down view of the level showing start position, end position and all checkpoints (purple hexagons). The last checkpoint, that was malfunctioning, has been highlighted.

Following the play-test, the user-research experts at IO Interactive wanted to discover if it was possible to recognize, in the patterns of player behaviour captured as gameplay metrics, feelings of frustration. Furthermore, which patterns of interaction and navigation in the game that point towards a state of frustration in the player, and whether these symptoms can be observed in different players.

There are different theories of frustration (Amsel, 1990; Rosenzweig, 1944), outlining different types of frustration, for example failure to understand goals, failure to communicate means available to achieve goals and repeated failure to overcome challenges. For the case study, frustration was defined using the following definition: *Repeated failure to overcome challenges*. This definition formed a compromise between the nature of frustration and the limitations of user-instrumentation data (which cannot show what users feel, only provide indication based on defined guidelines). The malfunctioning checkpoint exasperated the situation because every failure was further punished with a lengthy navigation of the same environment and facing the same challenges without any sort of achievement or feeling of progression.

In order to be able to confirm whether any frustration-detecting pattern was functional, the user research experts did not communicate to the gameplay metrics analysts where the participant manifested frustration. The gameplay metrics recorded during the play-test were given to the analysts with the mandate to individuate reoccurring patterns that could be symptomatic of frustration.

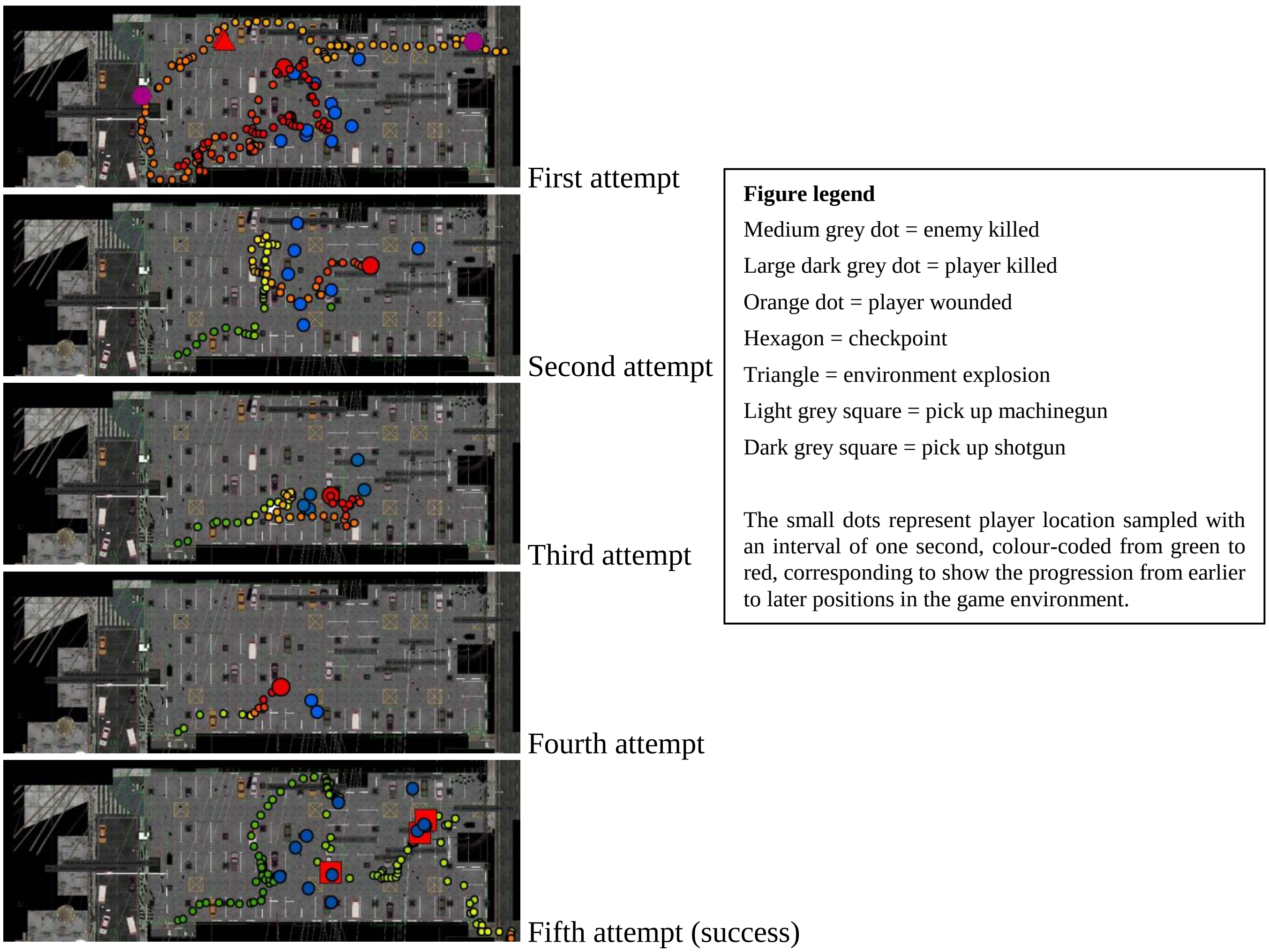
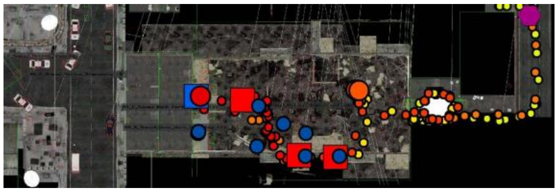


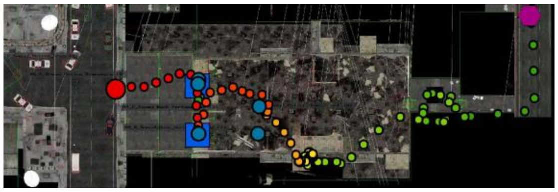
Figure 7: Player paths and in-game events in Kane & Lynch 2: Dog Days, expressed via recorded behaviour data. The images show, from top to bottom, the path of a play-tester and specific events that occurred during the test (e.g. player getting wounded). Each image shows the time from one instance of player death to the next, showing decreasingly less progress in the game from death 1-4; indicative of a behavioral pattern pointing towards player frustration.

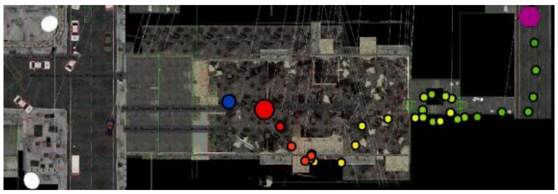
The first instance of behavioural pattern that could indicate a point where the player felt frustration (Figure 7), included several indicators: 1) First of all the player died in the same location four consecutive times, actually regressing in the second, third and fourth attempt (Figure 7). 2) Secondly, the number of enemies killed

decreased considerably in each play trough. 3) Thirdly, the pace of the player becomes considerably faster in each play through, as displayed by the spacing of the small dots, and repeats the same route with no variation (third and fourth death). 4) Also lacking is the presence of special events such as triggering environment explosions or picking up weapons dropped by enemies.
The fourth attempt proved to be the most unsuccessful, lasting only few seconds, showing the play-tester rushing into the enemies and failing almost instantly. After this failure the player appears to regain control, slowing the pace of movement, attempting a new route (leftward turn), killing a considerable amount of enemies and taking the time to pick up dropped weapons (Figure 7).

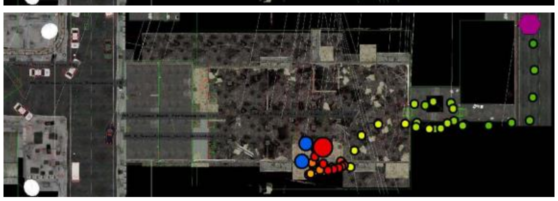
The analysis of the behavioural data were then correlated with the video recording of the play-test, and discussed with the user-research expert who ran the play-test. At the time indicated by the behaviour analyst, the player was evidenced to display signs of frustration, such as irritability, vocalized discontent and a certain blushing. A second set of data were employed to see if the same behavioural pattern occurred later in the play-test (Figure 8).


First attempt

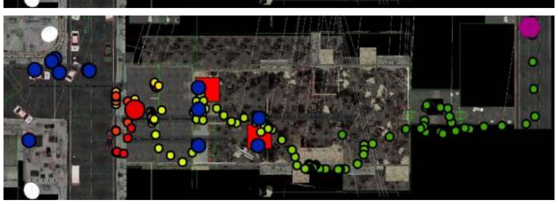

Second attempt

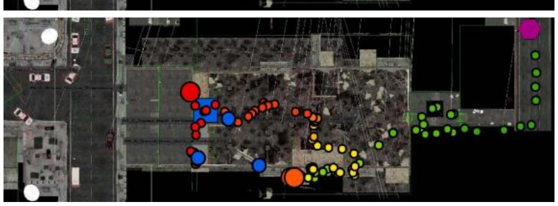

Third attempt


Fourth attempt


Fifth attempt


Sixth attempt


Seventh attempt


Eight attempt

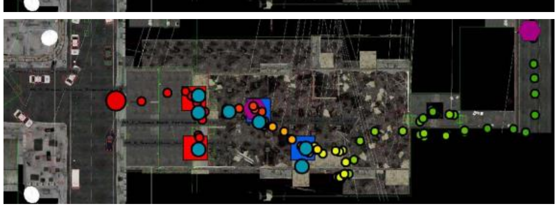

Ninth attempt (triggered final game level checkpoint)

Figure legend

Medium grey dot = enemy killed

Large dark grey dot = player killed

Orange dot = player wounded

Hexagon = checkpoint

Triangle = environment explosion

Light grey square = pick up machinegun

Dark grey square = pick up shotgun

The small dots represent player location sampled with an interval of one second, colour-coded from green to red, corresponding to show the progression from earlier to later positions in the game environment.

Figure 8: Player paths and in-game events in Kane & Lynch 2: Dog Days, expressed via recorded behaviour data. The images show, from top to bottom, the path of a play-tester and specific events that occurred during the test (e.g. player getting wounded). Each image show the time from one instance of player death to the next, showing progressively less progress in the game from death 3-6; indicative of a behavioural pattern pointing towards player frustration.

In this situation (Figure 8), the first attempt should have triggered the malfunctioning checkpoint, which was not working. Following the first instance of death, when the player attempted the challenge a second time, the player is also performing well, displaying proficiency and interest even in sub-tasks that are not vital to survival and game progress, such as searching fallen enemies for ammunition and weapons. In comparison, 3rd to 6th attempts to progress through the game level display a similar lack of progress, and series of death events happening more and more rapidly. The player increases pace without paying attention to secondary tasks and kills fewer and fewer enemies with each attempt. The four elements individuated earlier are present, to an even stronger degree: 1) The player dies in the same location, sometimes actually regressing; 2) The number of enemies killed decreases considerably; 3) The pace of the player becomes considerably faster, repeating the same route with limited or no variation; 4) The player does not give attention to non-vital, secondary tasks such as triggering environment explosions or picking up weapons. Similarly to the first example, the video recordings from the play-test were examined with the user-research expert running the test. Similar vocal and body-language responses showing frustration were found in the play-tester from death event 3-6.

It should be noted that the case study represents a very small sample of participants and a specific game, and the results are therefore not generalizable. Also, the patterns identified are only applicable when frustration is defined as being failure to overcome challenges. Other forms of frustration were not considered. The case study serves to highlight the usefulness of behavioural data to solve problems that arise during user-testing games, as well as during empirical games research. Furthermore, it shows how behavioural analysis can support the practice of placing check points in games such as KL2 using the experience and gut-instinct of designers. The four elements of frustration individuated in this case study regard a single player in a single game level, it will be vital in the future to verify the presence of these patterns in data gathered from the same player but in different levels, from different players and maybe from different games. If the hypothesis can be confirmed, it could be possible to identify universal markers enabling the automatic detection of frustration problems during instrumentation-based evaluation of play experience.

BEHAVIOR AND USER EXPERIENCE

While behavioral analysis addresses specific questions in game development and –research, such as game-space usage analytics, player interaction and player navigation etc., there is an second usage of behavioral data, namely in the combination with **user experience data** to provide a linkage between game design and play experience (Isbister & Schaffer, 2008; Nacke, 2009; Romero, 2008). In essence, gameplay metrics provide opportunity to link fine-grained behavioral information (finer than any other method, baring detailed video analysis) with user experience data. It should also be noted that order to enable metrics-based analysis, an infrastructure is needed to capture the data, which can mean substantial storage needs in the case of large commercial titles.

Gameplay metrics provide information only regarding actions undertaken by players, it is usually not possible to assess reasons and motivations behind the action, unless additional user data are captured (Drachen & Canossa, 2008, 2009b; Isbister & Schaffer, 2008; Lazzaro & Mellon, 2005). Gameplay metrics do not inform whether the player is male or female, or what the player thinks of the game experience. In short, gameplay metrics cannot provide any contextual data, although drawing inferences about the reasons for observed behaviors may be possible. Towards this end, Lazzaro & Mellon (2005) proposed the use of "fun meters" in games, essentially the collection of metrics of user behavior that are indicative of whether or not the players are enjoying the gaming experience. The essence of their argument is that behavior indicates the kind of experience the player is having (with the additional social factor). For example, looking at what people spend their money on in *The Sims Online* as an indicator of what entertains them in the game. Lazzaro & Mellon (2005) noted that many features in games affect enjoyment, and that each of these needs a meter (measure). In turn, these meters require a data source, which relates to the overarching question being asked. Extracting the data for the measures can be difficult, with basically two ways possible: Indirect (asking players) and direct (observing players). The authors highlight the added strength that correlation between data sources, whether all quantitative or mixed quantitative/qualitative, brings to an analysis.

User experience data in game testing is generally obtained using qualitative or semi-quantitative methods, such as user feedback (e.g. attitudinal data) via interviews or surveys, or potentially in combination with usability-oriented testing (Isbister & Schaffer, 2008; Laitinen, 2005, 2006; Romero, 2008). Usability testing generally focuses on measuring the ease of operation of a game, while playability testing explores is users have a good playing experience. In comparison, gameplay metrics analysis offers however insights into how the users are actually playing the games being tested.

Kim et al. (2008) and Romero (2008) presented the TRUE-solution of Microsoft Game User Research Labs, which in brief is a system capable of recording screen capture, video footage, behavioral and survey-data in one coherent framework. The TRUE system uses e.g. small pop-up surveys that activate during timed intervals to quickly assess the user experience of the player, recording simultaneously the position of the player character in the game environment (Kim et al., 2008; Romero, 2008). The problem with this approach is that the interaction flow between player and game is interrupted, and furthermore that the evaluation of the user experience is limited to one or a few dimensions, as surveys need to be kept short to keep the interruption to interaction flow to a minimum.

A promising approach, combining metrics with psycho-physiological methods (Cacioppo, Tassinary, & Berntson, 2007), has not been attempted yet – at least no published studies are known to the authors. It is however not unfeasible that this will occur in the near future, given the development of commercially viable EEG and EMG devices, and the proliferation of psycho-physiological studies of game experience (Mandryk & Inkpen, 2004; Nacke, 2009; Ravaja, Saari, Laarni, & Kallinen, 2006). Projects are already under way, e.g. in Canada and Scandinavia, both places as collaborations between companies and research institutions. The results remain to be seen, but in theory combining these two high-detail methods should be useful to correlate specific behaviors with the perceived user experience.

CONCLUSIONS AND PERSPECTIVES

In this chapter, the state-of-the-art of player (user) behavior analysis in digital games has been reviewed, and the current trends in industry and academia outlined. Case studies from the two major commercial titles *Kane & Lynch: Dog Days* and *Fragile Alliance*, have been presented which showcase how gameplay metrics can be employed to inform behavior analysis in practice (whether in a research or development context). The focus has been on detailed, multivariate work on data from just a few players, operating inside the virtual worlds themselves, taking advantage of the spatial dimension of play. This is a kind of analysis on which there exists virtually no published material. The case studies are based on common features of shooter-type games: Spatial navigation, environment interaction and death events; and are therefore cross-applicable across character-based games. The case studies indicate the usefulness of behavior analysis via gameplay metrics to recreate the play experience and thereby evaluate game design. For example, evaluating challenge and reward systems, locate areas that are over-/under-used, check for areas where players find navigation difficult, and importantly if the players operate as intended by the game design.

Behavior analysis via gameplay metrics addresses one of the major challenges to games-oriented user research, namely that of tracking and analyzing user behavior when interacting with contemporary computer games. As a user-oriented approach, it complements existing methods, providing detailed, quantitative data from – potentially – very large player samples, to supplement qualitative or semi-quantitative data from e.g. playability- and usability testing (Isbister & Schaffer, 2008; Kim et al., 2008; Lazzaro & Mellon, 2005). Additionally, Mellon (2009) highlighted Metrics-Driven Development as a method for utilizing instrumentation data to drive development and design of computer games, focusing on quantifying problems, thus rendering them measureable based on user-instrumentation data.

The field of player behavior analysis remains in its relative infancy, with methodologies lagging behind the general software industry, despite an emergent interest in the industry, and e.g. online games having access to a comparatively broader variety of user measures. One of the primary barriers for the uptake of game metrics in the industry appears to be cultural. Firstly, because metrics-based analysis is a relatively new addition to the industry, developers are reluctant to invest funds in the area. Secondly, metrics do not add to the features of a game. As noted by Mellon (2009): "*The biggest roadblock to our industry reaping the rewards of metrics is in fact our own business practices: if a task does not directly "put pixels on the screen", it is not a game feature and thus it is at the bottom of the funding list*." This leads to situations where the tools that do get build for tracking and logging game metrics are often build by individual systems engineers for their own needs, and thus do not get passed on within companies. A similar affect is present in the academia, where e.g. tools build by PhD-students do not get applied following their graduation. The exceptions to this rule is companies such *Square Enix Europe*, who has invested the time and resources to build team-wide tools for capturing business intelligence data, including behavioral metrics. Similar tools have been reported for companies developing and running MMOGs, e.g. *Lord of the Rings Online* (Mellon & Kazemi, 2008). In the academia, the resources necessary to obtain game metrics data has acted as a barrier for research, although inroads are being made thanks to collaborations with game companies (Drachen & Canossa, 2009a, 2009b; Thawonmas & Iizuka, 2008; Thawonmas, Kashifuji & Chen, 2008).

ACKNOWLEDGEMENTS

This would not have been possible without the stellar work of the Square Enix Europe Online Development Team. Also sincere thanks to the many other colleagues at IO Interactive, Square Enix Europe, the IT University of Copenhagen and the Danish Design School.

REFERENCES

Amsel, A. (1990). Arousal, suppression, and persistence: Frustration theory, attention, and its disorders. *Cognition and Emotion, 43*(3), 239-268.

Bateman, C., & Boon, R. (2005). *21st Century Game Design*: Charles River Media.

Blythe, M. A., Overbeeke, K., Monk, A. F., & Wright, P. C. (2004). *Funology - from usability to enjoyment* (1st ed. Vol. 3). Dordrecht: Kluwer Academic Publishers.

Brave, S., Nass, C. (2003). Emotion in Human-Computer Interaction. In *The Human-Computer Interaction Handbook: fundamentals, evolving technologies and emerging applications* (pp. 82-93). Mahwah: Lawrence Erlbaum Associates, Inc.

Burney, T., & Lock, P. (2007). *Measuring Game-PLay Performance and Perceived Immersion in a Domed Planetarium Projection Environment*. Paper presented at the ICEC

Byrne, E. (2005). *Game Level Design*: Charles River Media/Cengage Learning.

Börner, K., & Penumarthy, S. (2003). Social Diffusion Patterns in Three-Dimensional Virtual Worlds. *Information Visualization, 2*, 182-198.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). *Handbook of Psychophysiology* (3rd ed.): Cambridge University Press.

Chittaro, L., & Ieronutti, L. (2004). *A visual tool for tracing users´ behavior in virtual environments*. Paper presented at the Working Conference on Advanced Visual Interfaces.

Chittaro, L., Ranon, R., & Ieronutti, L. (2006). VU-Flow: A Visualization Tool for Analyzing Navigation in Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics, 12*(6), 1475-1485.

Coulton, P., Bamford, W., Cheverst, K., & Rashid, O. (2008). 3D Space-Time Visualization of Player Behavior in Pervasive Location-Based Games. *International Journal of Computer Games Technology, 2008 Article ID 192153*, 5.

Davis, J., Steury, K., & Pagulayan, R. (2005). A survey method for assessing perceptions of a game: The consumer playtest in game design. *Game Studies; The International Journal of Computer Game Research, 5*.

De Kort, Y., & Ijsselsteijn, W. (2007). People, places and play: Player experience in a socio-spatial context. *ACM Computers in Entertainment, 6*(2).

DeRosa, P. (2007). Tracking Player Feedback to Improve Game Design. *Gamasutra*.

Desurvire, H., Caplan, M., & Toth, J. A. (2004). *Using heuristics to evaluate the playability of games*. Paper presented at the Computer-Human Interaction (CHI) Extended Abstracts.

Drachen, A., & Canossa, A. (2008). *Defining Personas in Games Using Metrics*. Paper presented at the Future Play.

Drachen, A., & Canossa, A. (2009a). *Analyzing Spatial User Behavior in Computer Games using Geographic Information Systems*. Paper presented at the 13th MindTrek.

Drachen, A., & Canossa, A. (2009b). *Towards Gameplay Analysis via Gameplay Metrics*. Paper presented at the 13th MindTrek

Drachen, A., Canossa, A., & Yannakakis, G. (2009). *Player Modeling using Self-Organization in Tomb Raider: Underworld*. Paper presented at the IEEE Computational Intelligence in Games.

Ducheneaut, N., & Moore, R. J. (2004). *The social side of gaming: a study of interaction patterns in a massively multiplayer online game*. Paper presented at the ACM Conference on Computer Supported Cooperative Work.

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006). Building an MMO With Mass Appearl. *Games and Culture, 1*(4), 281-317.

Dumas, J. S. (2003). User-Based Evaluations. In *The HCI Handbook* (pp. 1093-1117): Lawrence Erlbaum Associates.

Fua, Y., Ward, M. O., & Rundensteiner, E. A. (1999). *Hierarchical Parallel Coordinates for Visualizing Large Multi-variate Data Sets*. Paper presented at the IEEE Visualization.

Gilleade, K. M., Dix, A. (2004). *Using frustration in the design of adaptive videogames*. Paper presented at the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology.

Goetz, P. (2006). Too Many Clicks! Unit-Based Interfaces Considered Harmful. *Gamasutra*.

Grosso, B. (2009). Refining the "Metrics Driven" Approach. from http://www.allfacebook.com/2009/06/refining-metrics-driven-approach/

Han, J., Kamber, M., & Pei, J. (2005). *Data Mining: Concepts and Techniques*: Morgan Kaufmann.

Hand, D., Heikki, M., & Padhraic, S. (2001). *Principles of Data Mining*. Cambridge: MIT Press.

Hazlett, R. L. (2006). *Measuring Emotional Valence during INteractive Experiences: Boys at video Game Play*. Paper presented at the Computer-Human Interaction (CHI).

Hilbert, D. M., & Redish, J. C. (1999). *A practical guide to usability testing*: Intellect books.

Hoobler, N., Humphreys, G., & Agrawala, M. (2004). *Visualizing Competitive Behaviors in Multi-User Virtual Environments*. Paper presented at the IEEE Visualization Conference.

Hurst, A., Hudson, S. E., & Mankoff, J. (2007). *Dynamic detection of novice vs. skilled use without a task model*. Paper presented at the Computer-Human Interaction (CHI).

Ijsselsteijn, W., de Kort, Y., Poels, K., Jugelionis, A., Bellotti, F. . (2007). *Characterizing and Measuring User Experiences in Digital Games*. Paper presented at the International Conference on Advances in Computer Entertainment (ACE).

Isbister, K., & Schaffer, N. (2008). *Game Usability: Advancing the Player Experience*: Morgan Kaufman Publishers.

Jørgensen, A. H. (2004). *Marrying HCI/Usability and Computer Games: A Preliminary Look*. Paper presented at the NordiCHI.

Kennerly, D. (2003). Better Game Design Through Data Mining. *Gamasutra*.

Kim, J. H., Gunn, D. V., E., S., Phillips, B. C., Pagulayan, R. J., & Wixon, D. (2008). *Tracking Real-Time User Experience (TRUE): A comprehensive instrumentation solution for complex systems*. Paper presented at the Computer-Human Interaction (CHI).

King, D., & Chen, S. (2009). *Metrics for Social Games*. Paper presented at the Social Games Summit.

Klein, J., Moon, Y., Picard, R. W. (2002). This computer responds to user frustration: Theory, design and results. *Interacting with Computers, 14*(2), 119-140.

Kort, J., Steen, M. G. D., de Poot, H., ter Hofte, H., & Mulder, I. (2005). *Studying usage of complex applications*. Paper presented at the Measuring Behavior.

Kuniavsky, M. (2003). *Observing the User Experience: A Practitioner´s Guide to User Research*: Morgan Kaufman Publishers.

Laitinen, S. (2005, June 23). Better Games Through Usability Evaluation and Testing. *Gamasutra*.

Laitinen, S. (2006). Do usability expert evaluation and test provide novel and useful data for game development? *Journal of Usability Studies, 1*(2).

Lazzaro, N., & Mellon, L. (2005). Fun Meters for Games. Austin Game Developers Conference.

Longley, P., Goodchild, M. F., Macquire, D., & Rhind, D. (2005). *Geographic Information Systems and Science*: John Wiley & Sons Publishers.

Ludwig, J. (2007). Flogging: Data Collection on the High Seas. Austin Game Developers Conference.

Mandryk, R., & Inkpen, K. (2004). *Physiological indicators for the evaluation of co-located collaborative play*. Paper presented at the ACM Conference on Computer Supported Cooperative Work (CSCW).

Medler, B. (2009). Generations of Game Analytics, Achievements and High Scores. *Eludamos Journal for Computer Game Culture, 3*(2), 177-194.

Medlock, M. C., Wixon, D., Terrano, M., Romero, R. L., & Fulton, B. (2002). *Using the RITE method to improve products: A definition and a case study*. Paper presented at the Usability Professionals Association.

Mellon, L. (2004). Metrics in MMP Development and Operations. Game Developers Conference.

Mellon, L. (2009). *Applying Metrics Driven Development to MMO Costs and Risks*: Versant Corporation.

Mellon, L., & DuBose, B. (2008). Acellerating the QA Test Cycle Via Metrics and Automation.

Mellon, L., & Kazemi, D. (2008). Wake up and smell the metrics! Austin Game Developers Conference.

Moskowitz, D., Shodhan, S., & Twardos, M. (2009). *Spore API: Accessing a Unique Database of Player Creativity*. Paper presented at the SIGGRAPH.

Nacke, L. E. (2009). *Affective Ludology - Scientific Measurement of User Experience in Interactive Entertainment.* Blekinge Institute of Technology, Blekinge.

Norman, D. A. (1988). *The Psychology of everyday things*. New York: Basic Books.

Nørgaard, M., & Rau, J. (2007). *User Testing in the Combat Zone*. Paper presented at the International Conference on Advances in Computer Entertainment Technology (ACE).

Pagulayan, R. J., & Keeker, K. (2007). Measuring Pleasure and Fun: Playtesting. In C. Wilson (Ed.), *Handbook of Formal and Informal Interaction Design Methods*: Morgan Kaufmann Publishers.

Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., & Fuller, T. (2003). User-centered design in games. In *The HCI Handbook* (pp. 883-906): Lawrence Erlbaum Associates.

Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., Fuller, T. (2003). User-centered design in games. In *The HCI Handbook* (pp. 883-906): Lawrence Erlbaum Associates.

Peterson, E. T. (2004). *Web Analytics Demystified: A Marketer´s Guide to Understanding how your Web Site Affects your Business*: Celilo Group Media.

Ravaja, N., Saari, T., Laarni, J., & Kallinen, K. (2006). Phasic emotional reactions to video game events: A psychophysiological investigation. *Media Psychology, 8*, 343-367.

Redding, P. (2009). Aarf! Arf arf arf: Talking to the player with barks.

Renaud, K., & Gray, P. (2004). *Making sense of low-level usage data to understand user activities*. Paper presented at the SAICSIT.

Romero, J. (2008). Successful Instrumentation: Tracking Attitudes and Behaviors to Improve Games. San Jose, CA.

Rosenzweig, S. (1944). An outline of frustration theory. In J. M. Hunt (Ed.), *Personality and the behavior disorders* (pp. 379-388). Oxford, England: Ronald Press.

San Augustin, J., Mateo, J. C., Hansen, J. P., & Villanueva, A. (2009). Evaluation of the potential of gaze input for game interaction. *PsychNology Journal, 7*(2), 213-236.

Southey, F., Xiao, G., Holte, R. C., Trommelen, M., & Buchanan, J. (2005). *Semi-Automated Gameplay Analysis by Machine Learning*. Paper presented at the AIIDE

Sullivan, L. (2006, September 8). Video-Game Analytics Track Players´ Behavior. *TechWeb Technology News*, from http://www.techweb.com/wire/192700568

Swain, C. (2008). Master Metrics: The Science Behind the Art of Game Design. NLGD Conference, Utrecht, Netherlands.

Thawonmas, R., & Iizuka, K. (2008). Visualization of online-Game Players Based on Their Action Behaviors. *International Journal of Computer Games Technology*.

Thawonmas, R., Kashifuji, Y., & Chen, K.-T. (2008). *Design of MMORPG Bots Based on Behavior Analysis*. Paper presented at the Advances in Computer Entertainment Technology.

Thompson, C. (2007). Halo 3: How Microsoft Labs Invented a New Science of Play. *Wired Magazine, 15*.

Thurau, C., Kersting, K., & Bauckhage, C. (2009). *Convex Non-Negative Matrix Factorization in the Wild*. Paper presented at the ICDM.

Tychsen, A., & Canossa, A. (2008). *Defining Personas in Games Using Metrics*. Paper presented at the Future Play.

Tychsen, A., Newman, K., Brolund, T., & Hitchens, M. (2007). *Cross-format analysis of the gaming experience in multi-player role playing games*. Paper presented at the DIGRA.

Vorderer, P., Bryant, J. (Eds.). (2006). *Playing Video Games - Motives, Responses and Consequences*. Mahwah: Lawrence Erlbaum Associates Publishers.

Williams, D., Consalvo, M., Caplan, S., & Yee, N. (2009). Looking for Gender (LFG): Gender roles and behaviors among online gamers. *Journal of Communication, 59*, 700-725.

Williams, D., Yee, N., & Caplan, S. E. (2008). Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer-Mediated Communication, 13*, 993-1018.

Yannakakis, N., & Hallam, J. (2007). *Capturing Player Enjoyment in computer Games*. Paper presented at the Advanced Intelligent Paradigms in Computer Games.

Yannakakis, N., & Hallam, J. (2008). *Real-time Game Adaptation of Augmented-Reality Games for Optimizing Player Satisfaction*. Paper presented at the IEEE Symposium on Computational Intelligence and Games.

ADDITIONAL READING

Chittaro, L., & Ieronutti, L. (2004). *A visual tool for tracing users´ behavior in virtual environments*. Paper presented at the Working Conference on Advanced Visual Interfaces.

Chittaro, L., Ranon, R., & Ieronutti, L. (2006). VU-Flow: A Visualization Tool for Analyzing Navigation in Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, *12*(6), 1475-1485.

Drachen, A., & Canossa, A. (2008). *Defining Personas in Games Using Metrics*. Paper presented at the Future Play.

Drachen, A., & Canossa, A. (2009a). *Analyzing Spatial User Behavior in Computer Games using Geographic Information Systems*. Paper presented at the 13th MindTrek.
Drachen, A., & Canossa, A. (2009b). *Towards Gameplay Analysis via Gameplay Metrics*. Paper presented at the 13th MindTrek
Drachen, A., Canossa, A., & Yannakakis, G. (2009). *Player Modeling using Self-Organization in Tomb Raider: Underworld*. Paper presented at the IEEE Computational Intelligence in Games.
Ducheneaut, N., & Moore, R. J. (2004). *The social side of gaming: a study of interaction patterns in a massively multiplayer online game*. Paper presented at the ACM Conference on Computer Supported Cooperative Work.
Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006). Building an MMO With Mass Appearl. *Games and Culture*, *1*(4), 281-317.
Han, J., Kamber, M., & Pei, J. (2005). *Data Mining: Concepts and Techniques*: Morgan Kaufmann.
Hand, D., Heikki, M., & Padhraic, S. (2001). *Principles of Data Mining*. Cambridge: MIT Press.
Kim, J. H., Gunn, D. V., E., S., Phillips, B. C., Pagulayan, R. J., & Wixon, D. (2008). *Tracking Real-Time User Experience (TRUE): A comprehensive instrumentation solution for complex systems*. Paper presented at the Computer-Human Interaction (CHI).
Mellon, L. (2009). *Applying Metrics Driven Development to MMO Costs and Risks*: Versant Corporation.
Pagulayan, R. J., & Keeker, K. (2007). Measuring Pleasure and Fun: Playtesting. In C. Wilson (Ed.), *Handbook of Formal and Informal Interaction Design Methods*: Morgan Kaufmann Publishers.
Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., & Fuller, T. (2003). User-centered design in games. In *The HCI Handbook* (pp. 883-906): Lawrence Erlbaum Associates.
Williams, D., Consalvo, M., Caplan, S., & Yee, N. (2009). Looking for Gender (LFG): Gender roles and behaviors among online gamers. *Journal of Communication*, *59*, 700-725.
Williams, D., Yee, N., & Caplan, S. E. (2008). Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer-Mediated Communication*, *13*, 993-1018.

KEY TERMS & DEFINITIONS

Game metric: Any quantitative measures used during or following game development. Game metrics generally relate to measures of performance, process or players.

Gameplay metric: A specific type of player metric. Any quantitative measure obtained from players of computer games, as pertaining to their actions inside the game environment or during interaction with game menus/interface.

Player metric: Any quantitative measure obtained from players of computer games.

User-Initiated Event (UIE): Any action initiated by a user of digital software, for example, pressing the mouse button or repositioning a digital avatar in a virtual world environment.

User Experience (UX): The subjectively perceived experience of using/interacting with a product, for example the experience of playing a computer game.

User behavior: The behavior expressed by users of a specific product, notably in terms of how the user interact with the product. User behavior takes place in a given spatio-temporal and social context.