Probing User Opinions in an Indirect Way: An Aspect Based Sentiment Analysis of Game Reviews

Full Paper

Björn Strååt
Stockholm University
Department of Computer and Systems
Sciences, DSV
Postbox 7003, SE-164 07
Sweden
bjor-str@dsv.su.se

Harko Verhagen
Stockholm University
Department of Computer and Systems
Sciences, DSV
Postbox 7003, SE-164 07
Sweden
verhagen@dsv.su.se

Henrik Warpefelt
Uppsala University
Department of Game Design
SE-62167 Visby
Sweden
henrik.warpefelt@speldesign.uu.se

ABSTRACT

This paper presents a method for gathering and evaluating user attitudes towards previously released video games. A three-part video game franchise was selected, and all user reviews of these games were collected. The most frequently mentioned words of the game were derived from this dataset through word frequency analysis. The words, called “aspects” were then further analyzed through a manual aspect based sentiment analysis. The final analysis showed that the rating of user reviews to a high degree correlate with the sentiment of the aspect in question. This knowledge is valuable for a developer who wishes to learn more about previous games success or failure factors.

CCS CONCEPTS

• **Applied computing** → Computers in other domains → Personal computers and PC applications → Computer games

KEYWORDS

User experience; user expectations; sentiment analysis.

ACM Reference format:

B.Strååt, H.Verhagen, H.Warpefelt. 2017. Probing User Opinions in an Indirect Way: An Aspect Based Sentiment Analysis of Game Reviews. In *Proceedings of AcademicMindtrek'17*,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org. AcademicMindtrek'17, September 20–21, 2017, Tampere, Finland
© 2017 Association for Computing Machinery. ACM ISBN 978-1-4503-5426-4/17/09...\$15.00
DOI: <https://doi.org/10.1145/3131085.3131121>

September 2017 (Tampere, Finland), 9 pages. DOI: 10.1145/12341

INTRODUCTION

Even though game companies invest a substantial amount of time and money in user experience design, the efforts of user experience specialists are not enough to guarantee a successful game release. While User experience (UX)-specialists work together with game design teams to find and overcome different game design issues, they do not always capture the same issues as the end users find. In a recent article by Patrick Stafford on polygon.com [1], (a news site for games and game design) UX specialists describe their work to find and improve video game issues during the design process. They repeatedly mention that they work with the end user in mind to maximize the experience of the game play and they mention several examples where they have improved a game interface or elements thereof – the separate parts are fine-tuned, but there is a risk that whole may be less than the sum of its parts. An example of this was shown by Strååt & Verhagen [2] who found that experienced UX specialists, who used evaluation tools specifically designed for video games, did not find the same issues that the end users were concerned about in online discussion forum: Where the evaluation group had found usability issues, the end users found issues with the game narrative e.g. which in turn affected their entire experience of the game in question. Even though the users enjoyed separate parts of the game, a narrative that did not meet the players’ expectations spoiled their experience. This implies that no matter how well the UX work is done, if important aspects of the game are left without proper attention, the game experience may suffer. Collecting and understanding the users’ opinion is a cornerstone in every user centered design, games being no exception, and discrepancies in expectations between developer and user can be harmful for the end product.

The fluctuating attitude of the end users can often be seen in video game franchise series, where one game in the series can be viewed as an “episode” in the universe presented by the game. The

Dragon Age franchise is an example of this. The franchise has been largely successful sales wise, but separate games in the series have received negative criticism and poor ratings from users. The userscore, from metacritic.com, of the Dragon Age shows that in some cases, the designers made decisions that were not favored by the users. With the correct information at hand, this may have been averted. In this study, we are not examining game usability or playability; we are examining a post factum situation with a focus on end users' expressed opinion in their own words in the user reviews.

We examine user reviews and the relation between the most frequently mentioned game aspects (combat, story, and character in this case), the sentiment of the aspect, and the rated popularity of the game. We claim that the resulting data analysis can be used to determine, post-mortem, which aspects of a game or game series that the users are most concerned about. While this method will not "repair" a broken game, it will serve as preparation and user research for future productions. Knowledge of the aspects, and the related sentiment, will help developers to decide what their design resources should focus on, and, in cases of user studies, what issues to discuss with their test subjects.

Research Question

We claim that users created reviews of a game will contain aspects that the users deem important, and the expressed sentiment of these aspects will also reflect the users' total judgment of the video game. Thus, knowing these aspects will provide an insight of which design elements users appreciate or dislike. If we select the most frequent aspects that the users seem to find important discussion topics, which of these are relevant for post judgment of a video game?

- Null hypothesis: there is no relationship between the values of aspect X (character, combat or story) and the overall review rating.

2 BACKGROUND

2.1 Game Series

The goal of this study is to see if the user sentiment differs between games that are released in a series. To this end, we decided to examine the user comments of the game series "Dragon Age". At the time of the study, Dragon Age has three installments: Dragon Age: Origin (DA1) [3], Dragon Age 2 (DA2) [4], and Dragon Age: Inquisition (DA3) [5]. The genre of Dragon Age is single player role playing game. The games in the series are played in the same universe, but each installment present a different narrative. Actions and decisions made by the player in a previous installment of the game can sometimes influence the game universe.

We chose the Dragon Age franchise since it is widely known, it represents a relatively common and popular game genre (role playing games), and most importantly, it has received varying ratings from players. The PC version of DA1 received 8.7/10.0

user score on Metacritic, DA2 received 4.5/10.0, and DA3 received 5.9/10.0. These discordant ratings tell us that something changes between each installment, and we believe that the cause can be found in the user comments.

2.2 Metacritic

Metacritic.com is a website that aggregates professional reviewer scores from various online media review sources. Television shows, movies, music and video games (various platforms) are examples of media that are presented. Metacritic calculates an average score called Metascore, based on the various professional reviewers by converting the reviewers' local score into a score of 0 to 100 (e.g. a local score of 8 out of 10 renders a Metascore of 80). These scores are weighted (based on the quality and overall stature of the source) and finalized into a professional Metascore. Regular non-professional users are also allowed to score the media on a scale of 0 to 10. The unweighted average of this score is presented by Metacritic as the Userscore. Non-professional users can also post their own reviews along with their score. The User score does not consider the length or quality of these reviews; a simple four-word comment, such as "this game is good", is valued the same as an analytical 500-word essay. User reviews and scores are posted anonymously under a self-selected user name. The user score is divided into three tiers: Positive, Neutral and Negative, where Positive is ratings 8 to 10, Neutral is ratings 5 to 7, and Negative is ratings 0 to 4. The rating tiers are color coded in green for Positive, yellow for Neutral and red for Negative.

Metacritic has been the subject of many discussions. The validity and value of the professional reviews have been questioned in various video game blogs and online magazines [6] [7], and the site has been used in game and social studies, e.g. as an examination and comparison of player experience vis-à-vis professional reviews [8], or as an important factor in assessing game value and quality [9]. Most commonly, the discussion has been around the professional reviews. In this study however, we have only looked at the User score and user comments.

3 PREVIOUS RESEARCH

In the last two decades, several studies have been made on user experience in video games. Nacke, Drachen and Göbel [10] discuss Gameplay experience (GX) in relation to UX evaluation methods, and highlight several methods, such as interviews, heuristic evaluations, ethnography etc. to evaluate and measure GX. Sánchez et al [11] discuss whether methods for analyzing User Experience are sufficient to measure and analyze experience of video game players, and present a framework to this end. O'Brien and Toms [12] studied attributes of user engagement, such as reasons for users to engage and disengage in an activity with an artifact. They present a conceptual framework model where negative affect, such as uncertainty, information overload, frustration or boredom are causes for disengagement. Zaman and Abele [13] suggest a tool for game designers to make informed

decision during the game development process to enhance player experience.

The use of game reviews as a resource for game user research is not a new phenomenon: Pinelle, Wong & Stach [14] used professional reviews as a source to find common video game issues, which they compiled into a set of video game heuristics; Livingston, Nacke & Mandryk [15] [16] used game reviews in several studies, examining to what degree player experience was influenced by reading negative or positive reviews; Zagal, Ladd & Johnson [17] found that game reviews often include design suggestions and serious discussions on game designer's intention and goals.

The player experience is strongly related to the player's previous experiences and expectations: Kultima & Stenros [18], for example, suggests that the game experience start long before the actual gameplay. They describe the entire cycle of the game experience; the potential user searches and evaluates information about different games from a variety of sources such as online reviews, friends, and game advertisements. The user then selects a game based on the analysis. In short, the user builds up an image of the game that she wants to play, based both on previous experiences and the information that has been gathered. If the actual game experience deviates too much from this image, the user is likely to be disappointed and dissatisfied. The concept of creating an image of an artifact is known as "Character", and is described in detail in the next section.

3.1 The Character of Artifacts

Janlert and Stolterman [19] describe user perception of artifacts as a "Repertoire of Character" (RoC), which we carry within us. This repertoire is based on our previous experiences with similar artifacts. All items have "character" (ibid), i.e. features that the user will recognize at their very first encounter with the artifact. The designer of an artifact relies on the user's ability to recognize the character of the item that they are designing, thus conveying to the user the feeling of an interesting, practical, etc. tool that they want. The challenge of the designer is to realize and clarify these characteristics into the product. Every user has a RoC that has developed over time. We, the users, use our RoC to evaluate new artifacts. Introducing new features in a design requires adherence to existing character-features, which the designer then can develop further. The users' RoC is constantly expanding, shifting and evolving, and designers should stay attentive to these changes to deliver interesting, appealing, and satisfying design [19].

Whenever users first encounter a new artifact, they form preconceptions of its character (ibid). Users construct a personal vision of the product's character, which in turn allow them to determine the appeal of the product (e.g., "It is good/bad"), emotional consequences (e.g. pleasure, satisfaction) and behavioral consequences (e.g. increased time spend with the product) [20]. After a period of use the user develops a mental model [21] of the system. This mental model is constantly evolving as the user's skill improves from consistent use, which in

turn forces the Avant-garde designer to evolve their design in a constant cycle of innovative development.

4 METHOD

In this section, we describe our scientific approach and methods for data gathering and analysis. We use a qualitatively driven mixed methods approach, where quantitative methods supplement and improve the study's results. The qualitative analysis is done through a through manual aspect based sentiment analysis. The quantitative analysis was done through hypothesis testing using a Chi-square test.

Aspect Based Sentiment Analysis

An aspect based sentiment analysis (ABSA) [22] is performed when user sentiment of certain aspects of a multi-aspect entity is to be measured, in a dataset gathered from user comments, such as online forum or user created reviews. Video games have plenty of aspects that the user considers when playing, e.g. playability, graphics, storyline.

Aspects are words or phrases that exist either explicit or implicit in the dataset. Explicit aspects are the actual word in context, and implicit aspects are inferred from the context. For example, if the aspect is *gameplay*, an explicit occurrence could be "*I really enjoyed the gameplay*", and an intrinsic could be "*I really enjoyed the challenges and the features of X*".

The aspects are determined through a word frequency analysis. After the dataset is collected, product or domain relevant words that occur on a frequency above a pre-set threshold are retained for the following sentiment analysis step. The sentiment analysis is then performed either through a scripted natural language processing algorithm, or through a manual read through. The result will show the sentiment for each aspect, for example in terms of positive, neutral, or negative sentiment.

Word frequency and selection. The data collection for our ABSA was performed in the following steps. First, we collected all user reviews on the PC-version of the three games from the Dragon Age franchise: DA1, DA2, and DA3, from Metacritic. As mentioned in the Metacritic description in the background section, Metacritic authors rate their own reviews to reflect their experience of the game in question. This is a rating from 0 to 10, but in effect it will categorize the comment as one of three tiers: low, medium, or high rated. We decided to only work with the reviews of the PC-version (the games exist for multiple platforms) as it was the versions that we were familiar with.

For each game, we did a word frequency analysis, using AntConc¹, to find which aspect that was most frequently used in the reviews. As we had no previous practice of this method in this context, the threshold was set after we saw the results – we decided to pursue the three most frequent explicit aspects that were shared by all three games. These explicit aspects were: *Story*, *Combat*, and *Character*. All reviews that did not contain

¹ AntConc, by Anthony (2012), is a freeware concordance and text analysis tool by Dr. Laurence Anthony at the Faculty of Science and Engineering at Waseda University, Japan (<http://www.antlab.sci.waseda.ac.jp/index.html>).

any of the aspects were omitted from the dataset. As the reviews were rated by the authors, we already had the rating categories.

Since the review rating and the sentiment of the aspect may differ – for example, a high rating review may use an aspect in a negative way – it was important to collect all reviews of all ratings. Figures 1, 2 and 3 show the frequency of the aspects in relation to review rating. As can be seen, the aspects tend to be more frequent in low rated reviews than high and mid rated reviews.

As a result of the data collection, we had a dataset of reviews for each game, regarding the three aspects (story, combat, character). Each review was categorized into its original rating level.

So, in conclusion of this section:

- Aspects were determined through word frequency analysis of all the user reviews
- The three most frequent aspects were combat, story, character
- Each game had a number of reviews
- A review contains at least one of the aspects
- A review is rated as either low, medium, or high
- The dataset contains all reviews, sorted by game, rating, and aspect.

Sentiment Analysis

The sentiment analysis was performed online, through an online crowdsourcing service.² The rating and name of the game was omitted for the evaluators to limit the risk of bias. The evaluators were asked to read a review, or excerpt of a review, which contained one of the aspects, and to determine if the author of the review had used the aspect in a positive, neutral, or negative way. The following quote is an example of an excerpt that the evaluators judged:

“The menus, crafting and combat are so totally and completely cumbersome. Everything is very statically organized and takes so much time. I spent an ungodly amount of hours collecting resources, crafting things, comparing items to what I already owned and it is just so, so, so cumbersome and tiresome, it really damages the game”

The aspect of combat occurs in the quote, and the overall use of the aspect is considered negative.

9358 review excerpts from all three games were analyzed this way, and each aspect was judged by at least three evaluators.

If an excerpt would contain more than one aspect, it would be run again, through a second (or third) sentiment analysis, where that aspect would be in focus for the evaluator.

When the sentiment analysis was done, the dataset was reconstructed with rating and game name.

5 RESULT & ANALYSIS

² www.crowdfunder.com; a data mining and crowdsourcing service where researchers can upload their data e.g. for manual sentiment analysis by anonymous evaluators.

A first glance at the results from the data collection and word frequency analysis, but before statistical analysis, will give us important clues. Consider figures 1, 2 and 3. The amount of user reviews increase for each instalment of the game franchise, but a large majority of the increase is within the negatively rated reviews. This is our first clue that the related aspect is important to the users. The amount of highly rated reviews is approximately the same in all three games and for all three aspects, but the low rated reviews are more than 50 times as many in DA3 than in DA1 regarding “Character” aspect, and 24 times as many regarding the other two aspects “Combat” and “Story”. This is not a statistically validated result, but it gives us an indication if we are looking at something that needs to be further investigated. The amount of low rated reviews that contain at least one of each aspect may indicate that these aspects are part of the reasons that users didn’t appreciate DA2 and DA3 as much as they did DA1. From a video game developer standpoint, we could stop here. It wouldn’t take too long to manually read through a few pages of these comments to get an estimated overview whether the aspects are used in a negative sentiment or not. A developer can, at this stage, get this overview and regard their design choices accordingly.

This however is without the sentiment analysis. Figures 1, 2, and 3 does not show if a low rated review contains a positive sentiment aspect.

Figure 1: Number/%-age of reviews (Y axis) regarding the STORY aspect in all three games.

Figure 2: Number/%-age of reviews (Y axis) regarding the COMBAT aspect in all three games. **Figure 3: Number/%-age of reviews (Y axis) regarding the COMBAT aspect in all three games.**

After the sentiment analysis, we processed the data from an analytical standpoint. Table 1 shows the complete data set for all three games, distributed on review ratings, aspects and sentiment.

We tested the relevance of each of the three aspects for the overall review. We constructed the following null hypothesis:

- there is no relationship between the values of aspect X (character, combat or story) and the overall review rating.

Table 1: Aspects distributed on review ratings, for all three games in the franchise. The numbers are from the evaluators' sentiment analysis.

		Review rating			
Aspect		Low	Mid	High	
Char.	bad	633	257	87	977
	neutral	1038	72	92	1202
	good	68	148	543	759
Comb.	bad	520	211	72	803
	neutral	358	50	48	456
	good	43	83	353	479
Story	bad	993	278	69	1340
	neutral	1056	119	129	1304
	good	72	142	734	948
		4781	1360	2127	8268

Using the Chi square test, we obtained the following values (table 2)

Table 2: Chi square values for each aspect, of all three games in the franchise

	Aspect	Chi square
DA1	Character	120,2
	Combat	100,4
	Story	196,6
DA2	Character	304,9
	Combat	299,6
	Story	426,6
DA3	Character	1072,5
	Combat	374,4
	Story	1250,2
DA series	Character	1541,3
	Combat	813,8
	Story	1963,4

All values exceed the threshold at $p = 0.001$ and 4 degrees of freedom (18,465) thus in all cases we can reject the null hypothesis and conclude that there is a correlation between the aspect value and the overall review value.

To clarify how the Chi square values in table 2 were obtained we refer to table 3 which shows all observations for DA1, Aspect = Character. In this case, Chi square = 120,2.

Table 3: Observations for DA 1, Aspect = Character.

		Review rating			Row total
		Low	Mid	High	
Character	bad	16	23	13	52
	neutral	2	9	20	31
	good	0	13	174	187
	Column total	18	45	207	270

6 DISCUSSION

Our results show that if an aspect occurs in a review, the sentiment of that aspect will reflect the rating of the review. The null hypothesis was falsified for all games, and all aspects.

This implies that the aspects reflect areas, in the games, that are disliked by the users. The relatively high frequency of the aspects is an indication that these areas are the most important ones for the users. It also indicates that the root cause of the low rated reviews is to be found within the game features that the aspects represent.

Our result point out the importance of understanding the user's situation and repertoire: three major game aspects were identified through the user review analysis. According to Janlert and Stolterman's [19] concept of Repertoire of Character (RoC), too many alterations in a known or expected artefact will go against the user repertoire, and put the artefact's popularity at risk. Users can, to some degree, predict or rather anticipate the behavior of the artifact, based on the perceived character and the context where the artifact is meant to be used. When the designer introduces a new feature, s/he must be careful not to take too big a leap from the familiar. Doing this may be perilous as the users feel that the artifact is to dispartate from their repertoire of expected characteristics of the artifact.

The frequency of the aspects implies that they are important to the users, and thus a part of their RoC – this implies that the low rated review authors are disappointed of the aspects as presented in the games. A future research task would be to perform a more qualitative analysis, on user review level, to pinpoint the root cause of the problems that the users experience. A content analysis, for example, of the material would give a more detailed insight. Furthermore, we have only worked with the PC-reviews of the game franchise. A full out analysis of all the platforms for all the games would possibly render a different result, or enhance the one presented in this paper.

It would be unfair to only listen to the users. So far, our focus has only been on discontented users. A natural next step must be to talk directly to the producers; game developers of high-end productions and independent developers alike: What is their take on our results? A voice that is rarely heard is that of the writers. It would give an extra dimension to our research if we gather their opinion. Integrating evaluation of core game characteristics such a narrative that is interesting into the evaluation of the interaction

mechanisms between the player and the game world will create a holistic game evaluation process where technical issues, human machine interaction aspects based on social and behavioral science and narrative and other humanities aspects are taken into consideration.

Could the low ratings of DA2 have been prevented, if the designers had access to this type of analysis? Possibly, but the task of artists and creators may not always be to cater to users' absolute needs. Experimental design in video games should always be encouraged, although it may be a hazardous endeavor. If the developers have more information about their users RoC, they can make more informed decisions, and thus take more risks. It is not the Dragon Age franchise that is the focal point of this study; it was selected because it fitted the purpose. As previously mentioned, the three games have a large amount of reviews, and all three parts have different reviewer ratings. We suggest that this type of study can be performed on any game, if it has a critical mass of user reviews to select aspects from. We decided on three aspects based on frequency, however, it is entirely up to the researcher to decide which aspects to choose, and on what premises to choose them.

REFERENCES

- [1] P. Stafford, "brain wave:the phds changing games," 18 1 2016. [Online]. Available: <http://www.polygon.com/features/2016/1/18/9882430/brain-wave-the-phds-changing-games>. [Accessed 29 1 2016].
- [2] B. Strååt and H. Verhagen, "VOX POPULI - A Case Study of User Comments on Contemporary Video Games in Relation to Video Game Heuristics?," United Kingdoms, 2014.
- [3] *Dragon Age:Origins*, BioWare, 2009.
- [4] *Dragon Age II*, BioWare, 2011.
- [5] *Dragon Age: Inquisition*, BioWare, 2014.
- [6] "Metacritic Matters: How Review Scores Hurt Video Games," 08 08 2015. [Online]. Available: <http://kotaku.com/metacritic-matters-how-review-scores-hurt-video-games-472462218>. [Accessed 18 04 2016].
- [7] "Time to kill Metacritic," 15 10 2014. [Online]. Available: <http://www.mcvuk.com/news/read/time-to-kill-metacritic/0139824>. [Accessed 18 04 2016].
- [8] D. Johnson, C. Watling, J. Gardner and L. Nacke, "The edge of glory: The relationship between metacritic scores and player experience.," in *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, 2014.
- [9] A. Greenwood-Ericksen, S. R. Poorman and R. Papp, "On the Validity of Metacritic in Assessing Game Value," *Eldamos. Journal for computer Game Culture*, vol. 7, no. 1, pp. 101-127, 2013.

- [10] L. Nacke, A. Drachen and S. Göbel, "Methods for Evaluating Gameplay Experience in a Serious Gaming Context," *International Journal of Computer Science in Sport*, vol. 9, no. 2, pp. 1-12, 2010.
- [11] J. Sánchez, L. Gonzáles, F. L. G. Vela, F. M. Simarro and N. Padilla-Zea, "Playability: analysing user experience in video games," *Behaviour & Information Technology*, vol. 31, no. 10, pp. 1033--1054, 2012.
- [12] H. L. O'Brien and E. G. Toms, "What is user engagement? A conceptual framework for defining user engagement with technology," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 6, pp. 938-955, 2008.
- [13] B. Zaman and V. V. Abeele, "Player-Centric Game Design: Adding UX Laddering to the Method Toolbox for Player Experience Measurement," in *Proceedings of Measuring Behavior*, Citeseer, 2012, pp. 128--140.
- [14] D. Pinelle, N. Wong and T. Stach, "Heuristic Evaluation for Games: Usability Principles for Video Game Design," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2008)*, 2008.
- [15] I. J. Livingston, L. E. Nacke and L. R. Mandryk, "The impact of negative game reviews and user comments on player experience," in *ACM SIGGRAPH 2011 Game Papers*, ACM, 2011, p. 4.
- [16] I. Livingston, L. Nacke and R. Mandryk, "Influencing experience: the effects of reading game reviews on player experience," in *Entertainment Computing-ICEC 2011*, Springer, 2011, pp. 89-100.
- [17] J. P. Zagal, A. Ladd and T. Johnson, "Characterizing and understanding game reviews," in *Proceedings of the 4th international Conference on Foundations of Digital Games*, 2009.
- [18] A. Kultima and J. Stenros, "Designing games for everyone: the expanded game experience model," in *Proceedings of the International Academic Conference on the Future of Game Design and Technology*, 2010.
- [19] L. E. Janlert and E. Stolterman, "The Character of Things," *Design Studies*, pp. 297-314, 1997.
- [20] M. Hassenzahl and N. Tractinsky, "User experience-a research agenda," *Behaviour & information technology*, pp. 91-97, 2006.
- [21] D. A. Norman, "Some observations on mental models," in *Mental Models*, Lawrence Erlbaum Associates, Inc., 1983, pp. 7-14.
- [22] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos and S. Manandhar, "Semeval-2014 task 4: Aspect based sentiment analysis," *Proceedings of SemEval*, pp. 27-35, 2014.
- [23] J. P. Zagal and N. Tomuro, "Cultural differences in game appreciation: A study of player game reviews," in *FDG*, 2013.
- [24] I. J. Livingston, R. L. Mandryk and K. G. Stanley, "Critic-proofing: how using critic reviews and game genres can refine heuristic evaluations," in *Proceedings of the International Academic Conference on the Future of Game Design and Technology*, 2010.