

The Impact of Negative Game Reviews and User Comments on Player Experience

Ian J. Livingston*
Department of Computer Science
University of Saskatchewan

Lennart E. Nacke†
Department of Computer Science
University of Saskatchewan

Regan L. Mandryk‡
Department of Computer Science
University of Saskatchewan

Abstract

Game reviews and player ratings have an effect on the commercial success of games. They are used extensively by game developers to gauge the success of their titles and by potential buyers to make more informed purchase decisions. However, their potential influence on player experience remains uncertain. We investigated how game reviews and user comments influence players' affective states and experiences during game play. We found that both professional reviews and user comments (especially the negative comments) affected experience measured through game ratings, and that this effect was not mediated by changes in players' moods. Our results are important to the game industry because of the meaningful negative effect that user and critic comments can have on individual player experience and the resulting commercial success of a game.

CR Categories: K.8.0 [General]: Games;J.4 [Computer Applications]: Sociology—Psychology

Keywords: biasing effects, games, reviews, player experience, negative affect

1 Introduction

In games, experience is king. Game reviews can be viewed as expert experience reports giving players an idea of what to expect from the game they want to buy [Larsen 2008]. In addition, reviews act as more than just purchase guides, they play a role in shaping our understanding of the reviewed games [Zagal et al. 2009] and of ourselves as gamers [Ivory 2006]. Game publishers have acknowledged the importance of reviews for selling their games, and use review aggregation sites to gauge the success of their titles and to determine funding for game franchises. For example, the website *Metacritic* aggregates reviews from both professional game critics and game players and is used by players to decide whether or not they will purchase a game. However, due to the voice these online review sites provide to game players, the role of professional critics has found competition in consumer word-of-mouth and peer comments. Thus, it is currently unclear who—critics or peers—has the most persuasive power.

Due to the prevalence of game reviews and user comments, players form opinions about games before having played them. What remains uncertain is how exposure to these reviews influences player experience, whether the source (e.g., critic or peer) matters, and

whether the player experience results from the player's opinion of the game, or is mediated by a change in their mood. The biasing of attitudes and opinions is not a new topic in social psychology, economics, or marketing. However, *biasing* caused by textual reviews has received little attention in the fields of game user research and human-computer interaction, especially regarding its effect on player experience.

We present a study examining the effects of review text on player experience. The review text was manipulated to be either positive or negative in tone, either relevant to the subsequent gameplay or not, and derived from either an authority (i.e., critic) or peer (i.e., user). Our results show: 1) Players who read the negative text rated the game significantly lower than players who read positive text; 2) The text had to relate directly to the game to produce biasing effects; 3) The authority of the source (expert or peer) did not produce different biasing effects; 4) The effects could not be solely attributed to changes in mood as a result of reading the text; and 5) Negative text has a larger biasing effect than positive text.

Important for our understanding of biasing effects and player experience in general, our results have particular relevance to the game industry where it is common practice to release games to critics and the media prior to the public release date to get early reviews.

2 Related work

Previous work has shown that negative stimuli have a significantly greater effect on an individual than positive stimuli. In their literature review, Baumeister et al. [2001] concluded that *bad is stronger than good* in almost every area of psychology. For example, it has been shown that negative information is more attention grabbing [Baumeister et al. 2001; Fiske 1980], influences impression more, [Fiske 1980], is remembered better [Prato and John 1991], and is processed more carefully cognitively than positive information [Baumeister et al. 2001]—resulting in a greater biasing effect of negative stimuli.

The persuasive biasing power of experts and peers has been explored in the domains of movies [Reinstein and Snyder 2005], e-commerce [Sundar et al. 2009], advertising [Yi 1990] and news [Sundar et al. 2007]. However, there has only been little investigation in the domain of game research. One exception is a recent industrial study investigating effects of game ratings on perceived game quality and game value [Jenkins et al. 2010]. In this study, players were shown a high, low, or no game score and an accompanying comment, such as "Game of the year!" The authors found when players were exposed to a high review score (e.g., 90%) they were more likely to rate a game higher than if they were exposed to a low review score (e.g., 61%). The authors concluded that showing players high game review scores will cause higher player ratings, but attributed their results to anchoring [Tversky and Kahneman 1974]. *Anchoring*¹ is a biasing effect where an individual's ability to estimate is biased toward a previously known value. Thus,

¹For example, people asked to estimate if a price is more or less than 10 dollars are more likely to guess lower values than those who are asked if a price is more or less than a 1000 dollars.

*e-mail: ian.livingston@usask.ca

†e-mail: lennart.nacke@acm.org

‡e-mail: regan@cs.usask.ca

Copyright © 2011 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for 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, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.

Sandbox 2011, Vancouver, British Columbia, Canada, August 10, 2011.
© 2011 ACM 978-1-4503-0775-8/11/0008 \$10.00

we cannot know whether the results seen in Jenkins et al. [2010] are due only to effects of anchoring, or whether they change the player’s perceptions of the game. In our experiment, we eliminated all anchoring effects; we did not present review scores, ratings, or value to players. Instead players only read an affective text stimulus. The use of affective text unrelated to a task to bias participants via a change in their mood is known as *affective priming* [Yi 1990]. We investigate affective priming through the use of movie reviews.

If players reading game reviews are invested in evaluating the information in the review, they will expend cognitive effort in the systematic processing of that information [Chaiken 2003]. They may also use cues, such as the length of the review or the review source, to give weight to the information they have read—a mechanic known as heuristic processing [Chaiken 2003; Sundar 2007]. Heuristically-processed cues are based on biases players have; for example, that a number of corroborating positive ratings are indicative of quality. The latter is called the *bandwagon heuristic*: if others think something is good, then I will think that it is good, too [Sundar 2007; Sundar et al. 2007; Sundar et al. 2009]. Another related heuristic cue is the *authority heuristic*: if an expert thinks something is good, I will agree [Sundar et al. 2009].

In the present study, we explore the effect of game reviews, user comments, and movie reviews on player experience.

3 Experiment

Our study asks five main research questions:

1. Does the positive or negative tone in review text affect player experience?
2. Does the relevance of the review text to the task matter?
3. Does the authority (expert or peer) of the authors matter?
4. Can the experience differences be explained by changes in mood?
5. How do our results compare to reading no review text?

3.1 Games

We used two little-known games: *Tag: The Power of Paint*, a 3D puzzle shooter, and *Fancy Pants Adventures* (FPA), a 2D platform game. The primary factor in choosing games was obscurity—the games had to be of professional quality but publicly unknown, so that users did not have prior first or second hand experience with them. No participant expressed prior knowledge of either game when asked after they completed the study.

3.2 Experimental Design and Affective Text Stimuli

We used a 3 Source (Game Critic, User Comments, Movie Critic) by 2 Valence (positive, negative) between-subjects design. The sources were either relevant to the task (Game Critic and User Comments) or not (Movie Critic) and were either authoritative (Game or Movie Critic) or peers (User Comments). We also included a control group, who did not read any review text, resulting in seven groups.

We created the review text by modifying real reviews and user comments to ensure that the text stimuli were comparable across conditions and as authentic as possible. For Movie Critic reviews, we modified reviews of obscure films to ensure that participants’ response to review text was not influenced by their opinions of the review subject. To create the positive and negative versions, we used affectively-charged *antonyms*. As an example, the statement

”I love this game” in the positive condition would be ”I hate this game” in the negative condition. The content and subject matter between conditions was maintained as much as possible. For example, if the positive condition discussed the camera positioning, so did the negative condition.

		Game Critic		Movie Critic		User Comments	
		+	-	+	-	+	-
1	+	7.0	0.8	7.6	2.9	8.5	2.3
	-	1.0	8.0	1.6	6.6	1.0	7.9
2	+	7.6	2.0	8.2	2.3	8.9	2.3
	-	0.5	7.0	1.5	7.6	0.8	7.5

Table 1: Valence scores for affective text. Columns show the source and valence condition (negative or positive text); rows show the positive and negative LIWC scores for two stimuli.

For the Game and Movie Critic reviews, we used multiple corroborating reviews to alleviate participant skepticism surrounding a single review. To increase participant buy-in, they were informed that the text had been taken from a professional review website. In all cases, the review text was about 1000 words. To ensure that the affective tone (positive or negative) of our text was comparable across conditions, we used the Linguistic Inquiry Word Count (LIWC) text analysis tool [Pennebaker et al. 2011] to measure the valence of review text for each game (see Table 1).

3.3 Participants and Procedure

Seventy participants (29 female), aged 18–36 (M=24) years began by completing a demographic survey, and then were randomly assigned to a group, balancing for sex. Players then completed the experiment using a custom system to deploy the games and gather the survey responses.

Players first provided their pre-text baseline valence rating using the self-assessment manikin (SAM) [Lang et al. 1980], a pictorial 9-point scale commonly used to self-report valence. Players then read the text corresponding to their test group (or read nothing in the control group). They were required to answer questions related to the text to ensure they systematically processed the information, and did not rely solely on heuristic cues [Chaiken 2003]. Players then provided their valence rating again, played the game for 15 minutes, and provided a final valence rating. Half of the participants played Tag first, the other half played FPA first. To evaluate their experience, players wrote a short game review and provided a game rating on a 100-point scale. This procedure was repeated for the second game. Before providing their ratings, players were never shown numerical scores to ensure that there were no anchoring effects [Tversky and Kahneman 1974] (when exposure to a numerical rating would influence a player’s own numerical rating) as seen in [Jenkins et al. 2010]. To help players choose scores, and to aim for consistency across players in their interpretation of the ratings scale, a score guideline was provided (0–19: I hated it; 20–49: I didn’t like this; 50–74: It was ok, but I’m rather indifferent; 75–89: It was good, but room for improvement; 90–100: I loved it).

3.4 Data Analyses and Dependent Measures

Our dependent measures are user ratings on a scale of 1–100 and user valence ratings from the SAM. These do not meet the assumptions of normality and equal variance measured using the Shapiro-Wilk test for normality. As such, we used the Mann-Whitney Test for all between-groups tests, and the Wilcoxon Signed Ranks Test for within-player comparisons [Huck and Cormier 1996]. Only data

from the first game played are presented, because initial analysis showed that players consistently discovered the experimental manipulation after playing the first game and answering the surveys, biasing the data from the second game. Consistent with previous work [Chaiken 2003; Jenkins et al. 2010; Chaiken and Maheswaran 1994], we discarded the data from the biased game trial.

4 Results

Does the valence of text affect the experience? Players who read positive text rated the game significantly higher than players who read negative text ($Z = 3.6, p \approx .000, M^+ = 72.9, SE^+ = 2.6$) ($M^- = 51.3, SE^- = 4.4$).

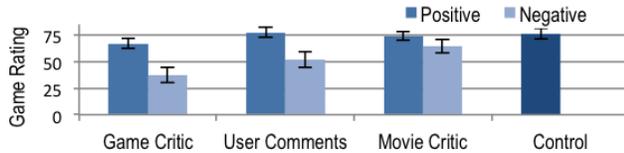


Figure 1: Means ($\pm SE$) for user game rating scores by source of affective text and valence of affective text.

Does the relevance of the affective text to the task matter? Tests for each text source (game critic, user comments, and movie critic) showed that reading positive text resulted in higher ratings than reading negative text for the players reading game critic reviews ($Z = 2.8, p = .004$) and user comments ($Z = 2.5, p = .011$) for the game that they subsequently played, but not for players reading movie critic reviews ($Z = 1.0, p = .315$). See Figure 1.

Does the authority (expert or peer) of the source matter? The experience ratings for players reading negative game critic (expert) and user (peer) text were not significantly different ($Z = -1.5, p = .127$). Neither were ratings for players reading positive game critic and user text ($Z = -1.7, p = .081$). See Figure 1.

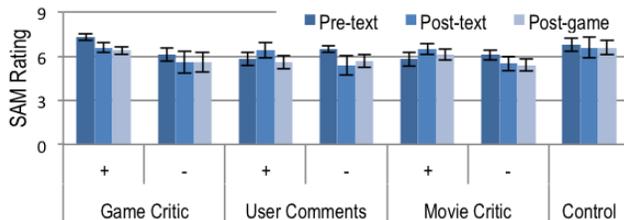


Figure 2: Means ($\pm SE$) of valence ratings (9-point scale, higher is more pleasant) for pre-text, post-text, and post-game by source and valence (+, -) of affective text.

Can the valence results be explained by a change in mood as a result of reading positive or negative text? We were interested in determining whether reading the affective text had a direct impact on player experience, or whether it caused a change in players' overall affective state (mood), which then had an impact on player experience. We had asked participants to rate their mood both prior to and directly after reading the affective review text. We compared these two ratings for each of the six groups (3 Source \times 2 Valence) and found that none of the differences were significant except for those who read positive movie text ($Z = 2.1, p = .034$). As Figure 2 shows, reading the positive movie review text improved players' self-rated valence. All other differences failed to reach significance

($p_{game^-} = .666, p_{user^-} = .103, p_{movie^-} = .336, p_{game^+} = .085, p_{user^+} = .230$).

How do our results compare to reading no review text? We included a control group that read no review text before playing the game to see whether the positive reviews were improving play experience or the negative reviews were worsening play experience. Players reading negative affective text rated the game lower than players reading no text at all ($Z = 3.0, p = .002$), but players reading positive affective text rated the game no differently than players reading no text at all ($Z = 1.2, p = .259$). See Figure 1.

5 Discussion

There are several results to take away from this study:

1. Reading negative review text negatively influences player experience.
2. To affect player experience, the text must relate to the game (e.g., a review of the game played).
3. The authority of the source does not matter; the peer and expert sources did not differ significantly in their influence on experience.
4. There was no significant difference in mood for players before and after reading game reviews and user comments.
5. The positive review text provides no positive benefit over no review text at all.

We next discuss the meaning of our findings and how they compare to prior work, their industrial relevance, and future research directions.

5.1 Valence Effects and Review Source

In our experiment we showed that the valence of review text significantly affected player experience. We further show that the negative text had a greater effect on player experience, and that the authority of the related review source (critic versus peer) does not matter. Interestingly, our results suggest that the observed effect on player experience was not primarily mediated by changes in player valence. If this were the case, reading the affective text would produce a change in mood, which would then produce a resulting change in player experience. Our results show that although the differences in player ratings were significant between the groups who read positive or negative reviews, the reviews did not produce a consistent and corresponding affective state change. It could be that the SAM ratings data was too noisy to find significant differences, but Figure 2 shows no consistent pattern, thus we are quite confident that the ratings differences cannot be attributed to changes in overall mood.

Our findings also suggest a primary contributing factor is relevance of the text to the game being played. The game reviews and user comments about the game produced a consistent and significant difference in player game ratings, while the movie reviews did not.

5.2 Explanations of the Observed Effects

We believe that systematic processing of the review text influenced players' opinions of and experience with the game; however, previous literature has suggested that changes in experience could be explained by heuristic cues, anchoring, or affective priming. We address each of these in turn.

The *bandwagon heuristic* suggests that if others think that something is good, then I will agree, while the *authority heuristic* sug-

gests that if an expert thinks something is good, then I will agree. Previous literature on the effects of experts and peers [Sundar 2007; Sundar et al. 2009] has shown the authority heuristic can be superseded by a bandwagon heuristic. Our results showed no differences in player experience after reading expert reviews or user comments. It is possible that systematic processing of the negative review text caused the prioritization and interpretations of negative content to overshadow any authority or bandwagon heuristic effects. In addition, we did not observe a difference between the control condition and any of the positive conditions, suggesting that when information is systematically processed, heuristic cues are less effective.

Affective priming uses a stimulus, unrelated to the task or object to be viewed, to change a participant's affective state or mood [Yi 1990]. In our movie review condition, we did see a positive impact of a positive review on players' moods, but not a negative impact of a negative review. However, there was no effect on player experience, thus our results cannot be explained by affective priming. Because the review text had to be related to the game to show effects, we can assume that task-relevance has a much stronger effect on player experience than affective priming.

Anchoring is a biasing effect where an individual's ability to estimate is biased toward a previously-known value. In previous work, players exposed to a negative rating of a game rated that game lower themselves [Jenkins et al. 2010]. This is clearly an effect of anchoring, so the previous work does not show that player experience was actually affected. In our study, we did not present the players with any numerical ratings, so the content and tone of the review text was causing the difference in player experience. Thus, our results cannot be attributed to anchoring.

We conclude it is cognitive processing and interpretation of negative text that has the greatest effect on player experience, not anchoring, affective priming, or heuristic cues. However, additional work examining the methods and effects of information processing, is required to confirm this.

5.3 Effects on Player Experience

Although we observed changes in game ratings as a result of reading affective reviews, the question remains whether the ratings differences were a result of a change in actual player experience or a reflection of the player's cognitive rationalization of their experience with the review text. In the former situation, a player who read a negative game review would actually not enjoy the experience as much as a player who read a positive review. In the latter, both players would have a similar experience; however, the one who read a negative review would interpret their experience as having been more negative.

To consider this, we can look at how players' moods changed over the course of gameplay. A significant correlation between review score and post-play SAM ratings (Spearman's $\rho = .442, p \approx .000$) show that the player's self-rated affective state and their rating of the game correspond. So although reading the review text failed to produce an observable change in players' affective state, playing the game after reading the review text did. The more positive that participants were feeling after playing, the higher they rated the game. This result leads us to think that actual experience was changed, but to better explore this question we must measure player experience during play. We plan to investigate this issue by measuring player experience at the physiological level [Mandryk 2008], which would tell us if the biasing effects from affective reviews are mirrored physiologically during play or if the mind tells a different story than the body.

5.4 Practical and Industrial Considerations

Compared to no affective text, negative reviews and user comments had a detrimental effect on player experience, whereas positive reviews produced little to no relative improvement. These results have practical significance for game designers, online community managers, and marketers.

Unlike other products, games provide entertainment experiences that not only require an investment of money but also of personal leisure time, which makes an informed purchase decision important for players. Game reviews facilitate this decision by providing player experience reports, giving players an idea of what to expect from the game they want to buy. It is standard practice for developers and publishers to show a game to media outlets prior to its release. If the game is well received, it can generate hype, which can translate into sales. Our results suggest that reading a positive review does not provide benefit in terms of player experience over reading no review, but that a negative review can significantly harm player experience, which would likely generate a negative hype effect. Therefore, negative reviews should be avoided, for example through a critic-proofing approach to playability testing [Livingston et al. 2010].

In an online community this negative effect can snowball as players who are influenced by critic reviews may in turn post negative comments online, influencing an even larger player population. As this cycle continues with new players reading and being influenced by the negative press, the bottom line of a game in terms of sales and future sequels could be negatively affected.

5.5 Limitations of the Study

We took care in our study design to eliminate influencing effects (e.g., anchoring), to remove any confounds from the stimuli (e.g., content, length, number of authors), and to test all conditions that would explain the nature of the effects (e.g., relevant text, authority or peer). Still, there are a few limitations in our work. The SAM results for the control condition show that players who read no text were happier than players in any of the other conditions. Reading and summarizing the text was not as fun as getting straight to the game, which is not surprising, but requires that we take care when interpreting differences between the control condition and other forms of play. In addition, our play sessions were conducted immediately following the reading of reviews and for a single, short duration. More work needs to investigate whether the effects hold for longer or repeated play sessions, for play sessions that occur long after reading the reviews, and for games a player is anticipating or is previously familiar with.

6 Conclusion

Game reviews and player ratings have an effect on the commercial success of games, but it has not been clear how game reviews affect player experience. We investigated how positive and negative text influences players' affective states and their play experiences. Our results show: 1) Players who read the negative text rated the game significantly lower than players who read positive text; 2) The text must be relevant and related to the game to produce significant player experience effects; 3) The authority of the text (expert or peer) did not influence the biasing effect; 4) The effects could not be explained by changes in mood as a result of reading affective text; and 5) Negative affective text has a larger biasing effect than positive affective text, likely due to the more careful processing of the negative information. Important for our understanding and analysis of player experience, our results have particular rele-

vance to the game industry where it is common practice to release games to critics and the media prior to the public release date.

Acknowledgements

This research was funded by NSERC and the GRAND NCE. Thanks to David McDine for literature suggestions and our past reviewers who provided helpful comments on previous versions of this document. We are also grateful to all participants for taking part in this study.

References

- BAUMEISTER, R., BRATSLAVSKY, E., FINKENAUER, C., AND VOHS, K. 2001. Bad is stronger than good. *Review of General Psychology* 5, 4, 323–370.
- CHAIKEN, S., AND MAHESWARAN, D. 1994. Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology* 66, 3 (mar), 460–473.
- CHAIKEN, S. 2003. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. In *Social Psychology: A General Reader*. Psychology Press, 461–473.
- FISKE, S. 1980. Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology* 38, 6, 889–906.
- HUCK, S. W., AND CORMIER, W. H. 1996. *Reading statistics and research*. New York: HarperCollins.
- IVORY, J. D. 2006. Still a man’s game: Gender representation in online reviews of video games. *Mass Communication & Society* 9, 1, 103–114.
- JENKINS, R., LEE, M., ARCHAMBAULT, R., SHANE, H., AND GREG, S. 2010. The influence of professional critic reviews. EEDAR/SMU behavioral study. Tech. rep., Southern Methodist University, Guildhall. Electronic Entertainment Design and Research, jul.
- LANG, P., SIDOWSKI, J., JOHNSON, J., AND WILLIAMS, T. 1980. Behavioral treatment and bio-behavioral assessment: Computer applications. In *Technology in Mental Health Care Delivery Systems*. Ablex Pub. Corp., Norwood, NJ, 119–137.
- LARSEN, J. M. 2008. Evaluating user experience - how game reviewers do it. In *Evaluating User Experiences in Games, Workshop CHI’08*.
- LIVINGSTON, I. J., MANDRYK, R. L., AND STANLEY, K. G. 2010. Critic-proofing: How using critic reviews and game genres can refine heuristic evaluations. In *Proceedings of FuturePlay 2010*, ACM, 48–55.
- MANDRYK, R. L. 2008. Physiological measures for game evaluation. In *Game Usability: Advice from the Experts for Advancing the Player Experience*. Morgan Kaufmann.
- PENNEBAKER, J. W., CHUNG, C. K., IRELAND, M., GONZALES, A., AND BOOTH, R. J., 2011. LIWC. Retrieved January 2, 2011, from <http://www.liwc.net>.
- PRATO, F., AND JOHN, O. 1991. Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology* 61, 3, 380–391.
- REINSTEIN, D. A., AND SNYDER, C. M. 2005. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The Journal of industrial economics* 53, 1, 27–51.
- SUNDAR, S. S., KNOBLOCH-WESTERWICK, S., AND HASTALL, M. R. 2007. News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology* 58, 3, 366–378.
- SUNDAR, S. S., XU, Q., AND OELDORF-HIRSCH, A. 2009. Authority vs. peer: how interface cues influence users. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, ACM, 4231–4236.
- SUNDAR, S. S. 2007. The MAIN model: A heuristic approach to understanding technology effects on credibility. *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*, 73–100.
- TVERSKY, A., AND KAHNEMAN, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185, 4754, 1124–1131.
- YI, Y. 1990. Cognitive and affective priming effects of the context for print advertisements. *Journal of Advertising* 19, 2, 40–48.
- ZAGAL, J. P., LADD, A., AND JOHNSON, T. 2009. Characterizing and understanding game reviews. In *Proceedings of the 4th International Conference on Foundations of Digital Games*, ACM, Orlando, Florida, ACM, 215–222.

