Can a few words influence one's game playtime

Lee Hao Rong Javier

School of Computer Science

National University of Singapore

javierlee@u.nus.edu

Abstract

There have been significant strides in understanding and interpreting human languages, yet, intent detection and natural language understanding remain rapidly evolving fields that continue to face challenges. Game reviews, on the other hand, are often ambiguous, complex, and laden with game-specific jargon, complicating their analysis. This study focuses on finding whether emotional intent expressed in a game review has a measurable effect on a user's decision to continue playing. This is done using causal inference methods and causal diagrams to identify and validate potential causal links between game reviews and user's game playtime. The findings shed light on the psychological motivations behind review writing and their effects on individual behaviour in the context of gaming.

Highlights

Aristotelian rhetorical triangle, causal analytics, natural language processing, data mining, intent classification

1. Introduction

Steam is a digital platform owned by Valve Corporation, a video game publisher and digital distribution company. As a popular destination for playing, discussing, and creating games for gamers worldwide, Steam captures around 75% of the market share in 2013. The total Steam purchases amounted to approximately US$4.3 billion by 2017 and accounted for around 18% of global PC game sales. As of 2021, the platform boasted over 34,000 games with over 132 million monthly active users (Backlinko, 2024).

This study hypothesises that users leave a review through two categories of motivation factors - pull and/or push factors. A pull factor is defined as a form of intrinsic motivation that compels users to leave a review, given the right conditions are met. A user influenced by pull factors is often an "always-taker" — they are consistently motivated to leave reviews when their internal motivations align with their experiences. In contrast, a push factor is a form of extrinsic motivation that forces a player to leave a review (so given an unknown treatment at an unknown extent, the player will take this unknown treatment and leave a review).

From this premise, this study models the Conditional Average Treatment Effect as the effect of both an individual's pull and push factors on his decision to continue playing. This model can be simplified even further if we aim to control for external factors that influence gaming experience across all users, so we can isolate the effect of pull factors on one's decision to continue playing given they have left a review. Specific assumptions for isolating and acknowledging these effect are provided in Appendix A1. Thus, the treatment group in this scenario is the effect of a specific pull factor a on game playtime, after the user has left a

review, and the control group is the opposite experience of pull factor a on game playtime, after the user has left a review.

But what are these pull factors that influence the kind of reviews left by a user? This study identifies them through their mode of communication and persuasion - whether a review has shown pathos (appeal through emotion), logos (appeal through rationality in argumentation) or ethos (appeal through user credibility). The existence or lack thereof these three tenets form the Aristotelian rhetorical triangle. Through the rhetorical triangle, this study hopes to better clarify the relationship between how a user's intentional communication to persuade others can end up affecting their own behaviour in terms of future gameplay.

2. Dataset Description

Due to limitations in time, this study focuses on a specific game - Ark Survival Evolved by Studio Wildcard. Choosing a specific game allows for better control of external influences too. Ark Survival Evolved is an open-world survival game where players gather resources, craft tools, and explore a vast island to interacting with other creatures. Ark Survival Evolved has a huge database of user reviews compared to other games, since they are the 15th most profitable Steam Game. At the time of writing, the second game, Arc II, is expected to release at the end of 2024, so knowing their existent player database would be beneficial and relevant towards future marketing strategies.

The current dataset used includes all reviews from the Steam Game Reviews Dataset from when the game is first released on August 29, 2017 till June 30, 2021, when the collection of data stopped. In total, there are 154700 reviews by users which constitute towards roughly 1% of the total population of users who purchased the game (Video Game Insights, 2024). While Ark Survival Evolved has been popular since its release, the Genesis Expansions are credited with revitalising the game and player base due to renewed interest to the game (Metacritic, 2024). User reviews suggest that this could be attributed to improved performance and stability and the drastic change in the game from an open-world survival game to a more mission-based progression (Kaggle, 2021). Further description of the dataset and features can be found in Appendix A2 and B1.

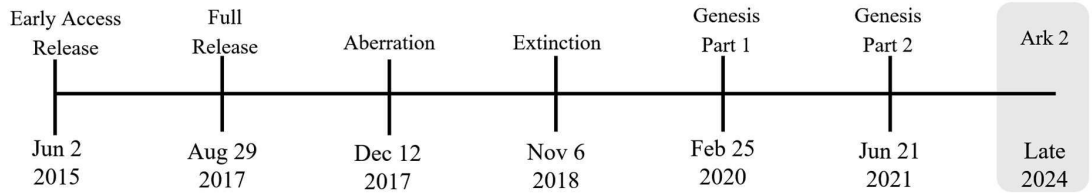


Figure 1: Timeline of major updates in Ark Survival Evolved.

3. Feature Selection

To effectively analyse the rhetorical triangle, it is essential to consider the key components of Pathos, Ethos and Logos. Since Pathos appeals to emotions, it is measured through the Sentiment Intensity Analyzer (SIA) from the Natural Language Toolkit (NLTK), as well as TextBlob's subjectivity and polarity metrics. Additionally, DistilBERT that was trained on the Twitter Emotions dataset is used to extract key emotional indicators, such as joy, anger, sadness, fear, surprise and love (Tunstall et al., 2022), was also used through getting an average 'emotion' score per sentence for an entire review, as well as getting the maximum 'emotion' score per sentence for an entire review. The labels of these items include an

emotion followed by either ‘average’, suggesting the former or ‘max’, suggesting the latter. On the other hand, Ethos, is measured through how relevant and credible a user’s review is. This is identified through the presence of topic clusters relevant towards games, and the whether the intention of the review is to be enlightening or sarcastic and funny. To find whether user’s review is relevant to a topic cluster, the Pearson correlation coefficient of the review embeddings is compared with the embeddings of the topics discussed. These topic clusters are further elucidated in Appendix B3. Finally, Logos is measured through the Dependency Parse Tree structure of each sentence in the entire review. This provides us with a view of complexity of sentence structure, as well as the persuasiveness of a user through Part-of-speech Tagging. All features included will be fully covered in Appendix B2.

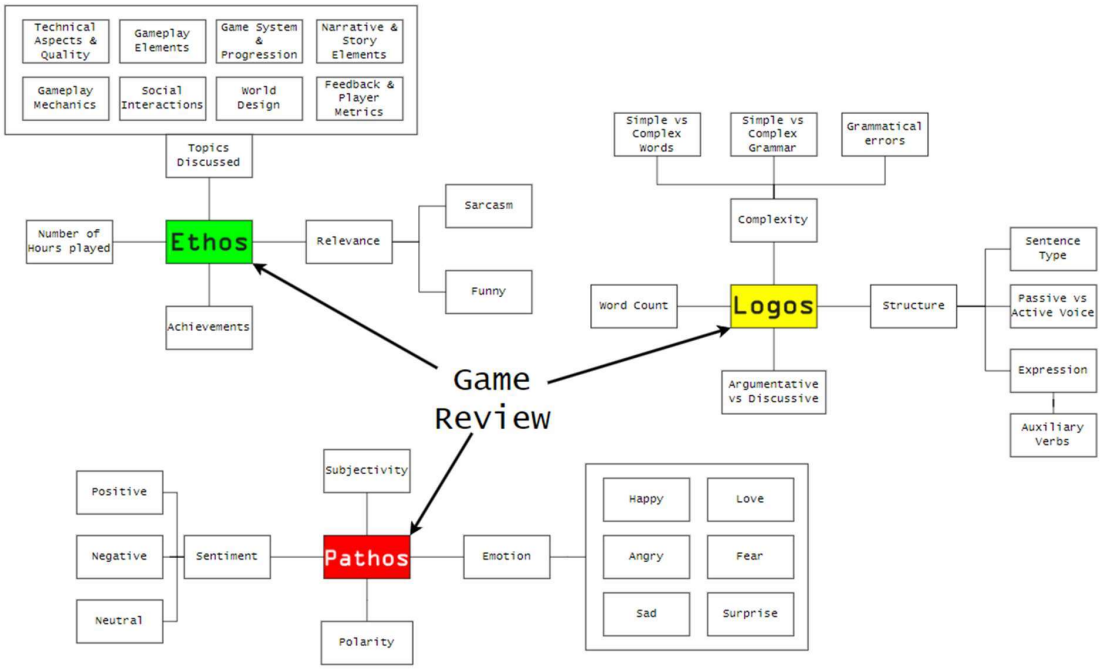


Figure 2: Summary of features extracted from user’s game review is further broken into their respective rhetorical triangle

4. Analysis

A causal diagram (Pearl, 1995) can thus be constructed based on the rhetorical triangle framework. The association diagram can be found in Appendix B4.

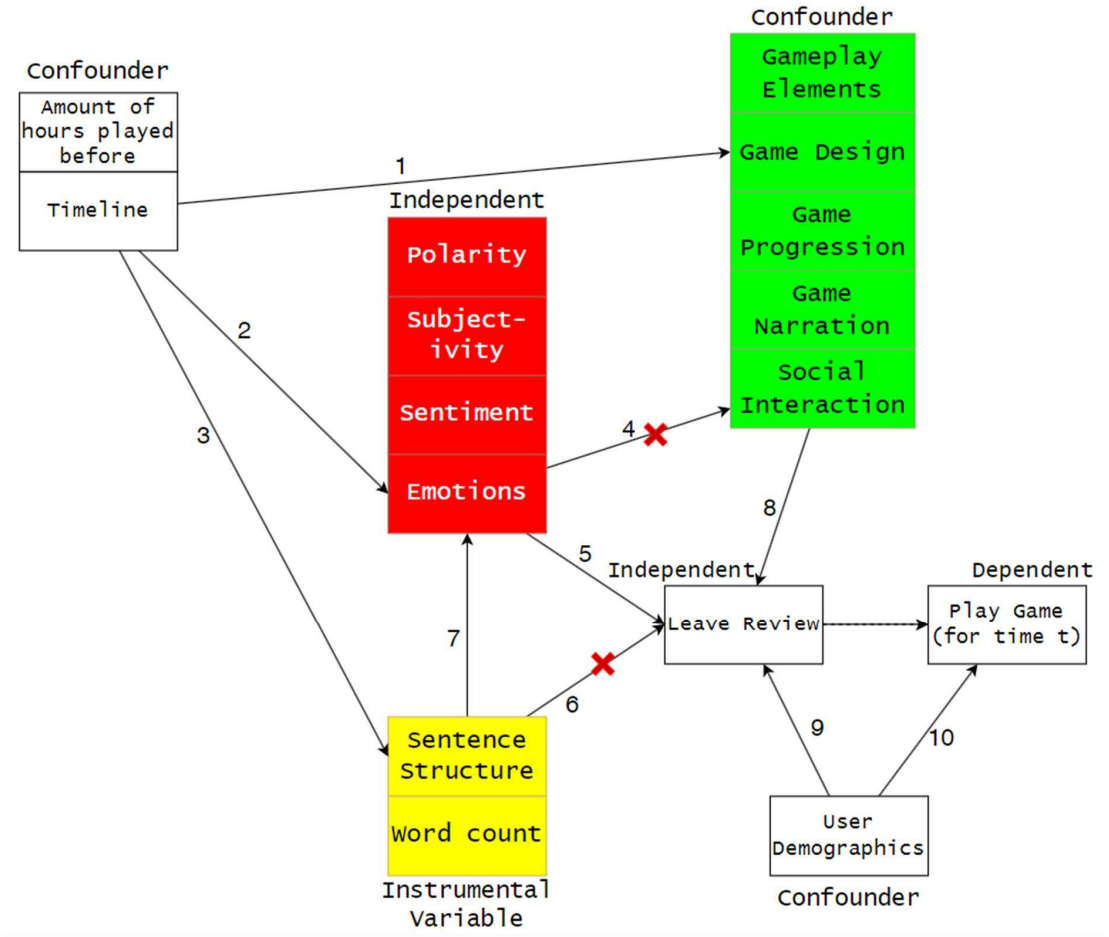


Figure 3: Causal Diagram

4.1. Effect of playtime before review and period which user left review, on topic clusters

Type of Bias: Confounding Bias

Additional covariates to be controlled: Interaction term between amount of hours played before review and time user left review on topic clusters.

To reduce heteroskedasticity of data, a power transformation of timeline (lambda = 0.5 since timeline is a count data) and a power transformation of amount of hours played before (lambda = 0 to reduce skewness of positively skewed data) is done to reduce non-linearity.

A user who plays the game for longer period should be able to, on average, point out more game related topics, thus enhancing his credibility. Empirically, this hypothesis is statistically significant ($p < 0.001$) in the areas of Gameplay Elements ($y = 0.14e^x+0.27$), Game Design ($y = 0.13e^x+0.24$), Game Progression ($y = 0.11e^x+0.27$), Game Narration ($y = 0.11e^x+0.23$) and Social Interaction ($y = 0.15e^x+0.24$).

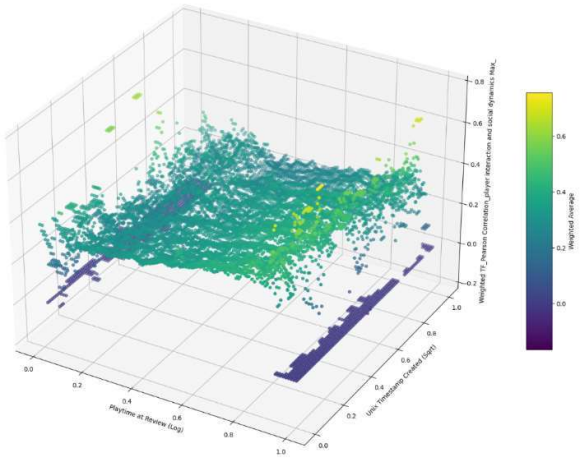


Figure 4: Plotting a response curve between timeline and amount of hours played before with social interaction topic cluster (coloured) shows a saddle like graph. Refer to Appendix C1 for more details.

A user who starts leaving game reviews later should on average, have reviews less related to game related topics. Empirically, this hypothesis is statistically significant ($p < 0.001$) in the areas of Gameplay Elements ($y = -0.09x^2+0.35$), Game Design ($y = -0.09x^2+0.32$), Game Progression ($y = -0.11x^2+0.34$), Game Narration ($y = -0.06x^2+0.28$) and Social Interaction ($y = -0.09x^2+0.33$) given some reviewers may see many reviews having talked about the same topics and may want to provide insights of other value compared to previous reviews. On this premise, playtime before review and timeline act as confounders toward the estimation of the effect of ethos on playing a game after leaving a review.

4.2 Effect of playtime before review and period which user left review, on emotions: polarity, subjectivity, sentiment and emotion categories: joy, sad, anger, love, fear and surprise.

Type of Bias: Confounding Bias

Additional covariates to be controlled: Interaction term between amount of hours played before review and time user left review on emotions. Interaction term between amount of hours played before review and joy_max$_+$, sadness_max$_+$, anger_max$_+$* .

Users who play a game for a longer period prior to leaving a review tend to leave a review that is more joyful. However, users who played the game for more than 2.5 years (top 30% of users) tend to leave a review that is less joyful than others as the graph is an inverted U-shape (Appendix C2.2, C2.13 and C2.14 bottom-middle). On the other hand, users who play a game for a longer period of time prior to leaving a review tend to leave a review that is less sad and less angry. However, users who played the game for more than 2.5 years (top 30% of users) tend to leave a review that expresses more sadness and anger than others as the graph is U-shaped (Appendix C2.5, C2.7 bottom-middle).

This phenomenon can be attributed to sunk cost fallacy. At the earlier stages of play, users get a sense of joy as the game provides entertainment, satisfaction and a sense of mastery. However, when players continue to play for a longer period of time, this sense of

* Let $_+$ represent a continuous linear spline model.

novelty and mastery slowly fades away, and users are more likely to focus on the flaws and aspects of the game that fail to meet their expectations. With higher expectations and a diminishing return for playing the game, it is reasonable for more experienced players to feel more anger and sadness. Perhaps, this could also be attributed to the amount of time spent on playing the game. Since playing a game for more than 2.5 years suggests that users have been through the best parts of the journey to the worst parts of the journey, negative bias can step in to vary a user's temperament in their reviews. As users tend to remember the bad experiences more vividly than positive or neutral experiences, the negative experiences may often leave a stronger impression especially if a user is exposed to the game for a long period of time. Furthermore, as time passes, good moments tend to be less distinctive and boring as users adapt to these good moments. The same experience that brings them joy may become more neutral due to content fatigue of the same experiences.

This emotional journey of players over time reflects a gradual shift from positive emotions at the start of their experience to negative emotions as users continue to play the same game, similar to an inverse U shape.

On the other hand, for users who played the game for less than 18 hours (approximately playtime_at_review_ln between 0 to 0.2, as represented in the top graphs of Appendix C2), there is high variance in emotional expression. This group tends to leave mixed, erratic, and varied reviews, possibly due to limited familiarity of the game.

4.3 Effect of playtime before review and period which user left review, on logic: sentence structure, word count

Type of Bias: Confounding Bias

Additional covariates to be controlled: Interaction term between amount of hours played before review on word count, existence of descriptive, exclamatory, conditional, compound, complex, compound-complex, cumulative, antithetical, affirmative, negative sentences, as well as existence of a passive voice, auxiliary verbs, sarcasm and asyndetons.

Overall, we would expect a user's appeal of logic to be constant when there is an increased playtime before leaving the review, however, this is not the case. From Appendix C3, the data suggests that users with more playtime before leaving a review tend to leave reviews with more words (C3.1), more usage of descriptive (C3.4), exclamatory (C3.7), conditional (C3.9), compound (C3.10), complex (C3.11), compound-complex (C3.12), cumulative (C3.16), antithetical (C3.21) sentences. There is more usage of polysyndetons (C3.25), passive voice (C3.27), auxiliary verbs (C3.29) too. All these point towards a higher logical complexity and an effort to build more elaborate and convincing arguments.

Users with more playtime before leaving a review also tend to leave reviews with lesser declarative sentences (C3.8), asyndetons (C3.26) and sarcasm (C3.30), suggesting that these users avoid making overly simplistic or one-sided or irrelevant claim but favour a more balanced approach. The combination of these factors suggests a higher cognitive maturity as experienced users are able to recognise the complexity of their experiences and thus avoid to reduce these experiences into simpler declarations.

Users with the least and most playtime tend to leave reviews with more negative words (C3.13) and less affirmation (C3.14) than the average user too since the graph is quadratic in nature.

On the other hand, there seems to be two distinct reviewer groups for most of the graphs in the bottom right corner of Appendix C3. The line of best fit appears to be a linear spline with a knot (discontinuity) around 0.8, which is also coincident with the release of Genesis Part 1, the extension and update of the game, after February 25, 2020. This suggests that updates that are positive can dramatically alter reviews.

4.4 Effect of emotions on topic clusters

Type of Bias: No Bias

Additional covariates to be controlled: None

In general, we see that emotions such as joy, sadness, anger, as well as positive sentiments and subjectivity are weakly positively correlated to the existence of topic clusters (Appendix C4.1, C4.2, C4.3). On the other hand, emotions as joy_max, sadness_max and anger_max is also able to capture a sizable amount of variance of the topic clusters (shown in Appendix C4.5) when calculating for the R squared.

4.5 Effect of emotions on users playtime after review

Users who express extreme happiness in their review is hypothesised to continue playing a game for a shorter period of time compared to other users who do not express extreme happiness in their review. From graphs in Appendix C5.2.1, C5.3.1 and C5.4.1, we see that playtime after review for the emotions - joy, sadness and anger respectively increases, then decreases after points 0.6, 0.1 and 0.1 respectively.

A user who only expresses one emotion (and that emotion being happiness) in their review can be attributed to be giving an argumentative perspective, while users who express multiple emotions in their review can be attributed to providing a more discursive and nuanced perspective. People who provide a more discursive perspective may have a deeper level of cognitive and complex engagement with the game. Such complex thought goes beyond the basic single emotions such as joy and may suggest a stronger personal connection that leads to more game playtime.

Furthermore, since users must invest in tailoring their game review to be discursive and not argumentative in nature suggests a form of dedication, investment and trade off towards said user using his time on other activities. From a psychological standpoint, users may then fall into the sunk cost fallacy, where they further feel the need to continue playing to validate the depth of involvement they have shown through their reviews. Furthermore, users may experience the endowment effect, where users overvalue the game now that they have invested a sizeable amount of time in leaving a discursive review.

Users who leave a more discursive review could also be more susceptible towards self-selection bias, where they implicitly see themselves to be part of the larger group and

community. By leaving a game review, it also suggests to a user that he has an illusion of control, where users believe that they have more influence over a game than they do.

On the other hand, negative emotions such as sadness_max and anger_max are positively U shaped. This implies that users who express extreme dissatisfaction are more likely to play the game for a longer period of time.

This could be associated with the illusion of control, as users who play the game longer may expect their feedback to be able to lead to changes in the game and this hope could be the driving factor that keeps them from leaving. Reinforcement of negative emotions can also occur, where players want to overcome their frustration or resolve their dissatisfaction especially when the game as challenging mechanics or unfair elements. The frustration may push them to keep playing to prove their skills.

As users increase the amount of hours played, it is expected that users who are emotionally attached to the game will increase the number of hours played prior to leaving a review, and this in turn brings them more utility and satisfaction, thus leading to them playing the game even longer. This suggests a positively reinforcing loop and it is hard to measure the cyclical effects of amount of hours played before leaving a review, and their emotions. As a result, logos (the appeal to rationality and logic) is used as an instrument variable to predict pathos.

4.6, 4.7 Effect of logos on users playtime after review and the appropriateness of logic as an instrumental variable

Since logical users use their rational thought to evaluate how much satisfaction or happiness (or lack thereof) they are getting from playing a game, users may implicitly reason that playing more of the game is good for well-being, and this in turn impacts their emotions. Thus, being logical does not directly influence the decision to continue playing a game but impact a user's emotions that drives gameplay. The usage of rationality as an instrumental variable breaks the endogenous cyclical relationship between happiness and future gameplay.

First Stage F-statistic is also greater than 100 for independent variables 'joy_avg', 'sadness_max' and 'anger_max', albeit a low correlation coefficient score shown in Appendix C7.2 C7.3 C7.4. We can then run 2SLS controlling for other covariates that influence their emotions to predict for each pathos variable. Using the predicted values of happiness for each individual that is influenced by rationality as an explanatory variable in regression of future game playtime, we get the results shown in Figure 5.

By instrumenting emotions with rationality, our model offers a much more robust causal effect without the bias from cyclical nature of the relationship between emotions and playing the game.

4.8 Effect of topic clusters on playing game after leaving a review

When users appeal to credibility by using game terms associated with the topic clusters, they present themselves as an expert or someone worthy of giving a comment. When these users perceive themselves as getting more recognised, such as the increase in votes on

their review, they may feel the need to live up to their established image and maintain credibility. This pressure is internal and self-imposed and not external at all.

On the other hand, by using game terms associated with topic clusters, the user may be more susceptible towards self-selection bias, where they implicitly select themselves to be part of the larger group and community. The need to maintain this social role can lead them to play more too. By leaving a game review with game terms associated with topic clusters also suggests to a user that he has an illusion of control, where users believe that they have more influence over a game than they do.

4.9, 4.10 Effect of user demographics on leaving a review and playing game after leaving a review

A user who owns a number of games or has left a number of reviews may be less attached to one specific game since his attention has to be spread between multiple games and may be more selective in his reviews and games he chose to play due to considerations about opportunity cost. This can influence the quality of review left as well as the amount of time invested in a game after the review.

However, users who own more games or leave more reviews can also be more experienced in either and this knowledge can increase the perceived credibility of the game review.

5 Results

Description	Df	R squared	Adj R squared	Location to find results
OLS	37	0.3700	0.3700	Pg9 + Appendix C3,C4,C6,C7.ipynb
OLS with interaction terms with playtime before review	126	0.3940	0.3930	Pg9 + Appendix C5.ipynb
OLS with interaction terms with playtime before review and interaction terms with timestamp	254	0.3980	0.3970	Pg9 + Appendix C5.ipynb
OLS with interaction terms with playtime before review and interaction terms with timestamp with forward selection of 0.1	68	0.3930	0.3930	Pg9 + Appendix C5.ipynb
OLS and using Logos as IV for Pathos	22	0.3236	0.3235	Pg9 + Appendix C3,C4,C6,C7.ipynb

Figure 5: Prediction for game playtime after leaving a review

6 Estimation of the counterfactual using matching

To estimate for the conditional average treatment effect on whether emotional intent expressed in a game review has a measurable effect on a user's decision to continue playing a game, greedy matching of mediators and confounders was done. A doubly robust estimator using the second last regression model in Figure 5 was used to estimate for the conditional average treatment effect too.

Emotion	Estimated Treatment Effect Before Matching (hr)	Estimated Treatment Effect After Matching (hr)
joy_avg > 0.6	1520.83	-285.61
angry_avg > 0.6	-2752.21	-8861.86
polarity > 0	4377.03	619.35
subjectivity > 0.5	1545.81	-316.39
sentiment > 0	3673.85	1422.66

Figure 6: Estimated Conditional Average Treatment Effect, given average total game playtime is 13779 hours ~ 1.5 years.

Prior to matching, the estimated treatment effect is mainly positive, suggesting that users who express strong amounts of joy and the other emotions in their review would continue to play the game for longer periods. However, after matching to control for confounding variables, the estimated treatment effect of all 5 emotional intents decreased substantially, suggesting that high expressions of joy did not lead to longer playtimes. This is in line with our hypothesis that users who display too extreme of an emotion may not play the game longer than users who display a more nuanced emotion. Rather, these intense emotional responses could reflect content fatigue compared to other users who give a more balanced response.

7 Future Improvements

There are many improvements that can be done such as incorporating GPT models to further improve the accuracy of feature extraction and intent classification. Currently, features that capture the appeal to logic are all binary in nature, thus the effectiveness of these features as an instrumental variable is very limited.

Further, the study is limited to the focus on one game. Including more case studies of course, can help determine if these observed relationship holds across game genres and further enhance the study's impact.

8 Conclusion

Through the findings of this paper, developers can use these insights to release updates aimed at renewing engagement at the 2 to 2.5 year mark to address the observed decline in positive sentiment and further prevent disengagement. Further, game design elements could be developed (as a form of targeted intervention) to ensure that users are nudged towards writing a more nuanced or positive review, as this might lead to higher user retention rates. Specific towards highly competitive games such as Ark Survival Evolved, developers can also target players who express sadness or anger through nudging players to take a break from the game so that they can maintain long term engagement while reducing burnout and negativity.

Ultimately, this study demonstrates that player behaviour evolves based on their emotional and logical experience. Designing a game with both engagement and well-being in mind can not only extend the lifespan of games but also cultivate stronger relationships between players, ensuring mutual benefit to both the gaming community and developers.

References

Backlinko. (14 June, 2024). Backlinko. Retrieved from Steam Usage and Catalog Stats, November 27, 2024, from https://backlinko.com/steam-users

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Kaggle - Steam Reviews Dataset. (2021, April 27). Kaggle. Retrieved November 27, 2024, from https://www.kaggle.com/datasets/forgemaster/steam-reviews-dataset

Metacritic - ARK: Survival Evolved - Genesis Reviews. (2024). Metacritic. Retrieved November 27, 2024, from https://www.metacritic.com/game/ark-survival-evolved-genesis/

Pearl, J. (1995). Causal diagrams for empirical research. Biometrika, 82(4), 669. https://doi.org/10.2307/2337329

Tunstall, L., Von Werra, L., & Wolf, T. (2022). Natural Language Processing with Transformers, Revised Edition. “O’Reilly Media, Inc.”

Video Game Insights – Games industry data and analysis. (2024). Video Game Insights. Retrieved November 27, 2024, from https://vginsights.com/game/ark-survival-evolved