Flow state in a Vampire Survivors style game

Samuel Spithorst (5956684)
Stefan Hoekzema (6896383)
Vos Wesseling (6885373)
Jeppe Vroegindeweij (6955266)
Lars de Kwant (6958680)

1 INTRODUCTION

We aim to model player flow in a Vampire Survivors [8] inspired game. In this game, the player controls a character that moves around an arena while enemies spawn periodically. The objective is to survive incoming waves of enemies for as long as possible. To aid in this goal, the player can collect experience points from defeating enemies, which can be used to discover and upgrade weapons that increase damage output. The difficulty level can be adjusted by modifying enemy strength and frequency.

Flow [2] refers to a state of complete absorption and optimal experience that arises when one is fully engaged in an intrinsically rewarding activity. To achieve flow, there must be a balance between the challenge of the activity and one's skills. If the challenge exceeds skills, one feels anxious. If skills exceed the challenge, one feels bored. When plotted, the area where challenge and ability are in balance can be visualized as the flow zone (Figure 1).

Our goal is to assess whether players are in a state of flow by determining if the game's challenge matches their skill level. To do so, we collect gameplay data together with a periodic self-reported assessments of experience and frustration. These inputs will be fed into a machine learning model that will be trained to estimate the relative difference between the player's skill and the game's difficulty at any given moment. The model outputs will indicate whether the player is in flow (challenge balanced with skill), anxious (challenge exceeds skill), or bored (skill exceeds challenge).

By modeling flow based on gameplay metrics, we can dynamically adjust game parameters to keep players in an engaging state of flow throughout. This work will provide insights into adapting games to player skills and experience.

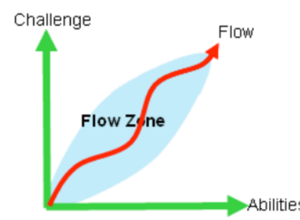


Figure 1: The Flow Zone, a state where the game's challenge matches the player's skill

2 RELATED WORK

Player modeling is all about detecting, modeling, predicting and expressing human player characteristics that are are manifested through cognitive, affective and behavioral patterns [12]. A **model** is in the form of a mathematical representation, like a rule set, a vector of parameters, or a set of probabilities [13]. Player modeling approaches can be split into top-down and bottom-up approaches, also known as model-based and model-free.

Model-free approaches refer to the data-driven construction of an unknown mapping (model) between a player input and a player state [13].

A **model-based** approach [12] is based on a theoretical framework. As such, researchers follow the modus operandi of the humanities and social sciences, which hypothesize models to explain phenomena. Yannakakis and Togelius (2018) [13] states that there are three main disciplines we can borrow theoretical frameworks from and build models of player experience: psychology and affective sciences, neuroscience, and finally, game studies and game research. In the psychology and affective sciences field is the concept of flow [2], which our research is about.

GameFlow [10, 11] is a model for evaluating player enjoyment in games, that consists of eight elements: concentration, challenge, skills, control, clear goals, feedback, immersion, and social interaction. The model was used to evaluating one high rated and one low rated real-time strategy (RTS) game. The goal was to find out what makes RTS games enjoyable and to find the relative importance of each GameFlow element. They found out that their model only works for some game genres and that it is not suited for strategy games.

Nacke and Lindley (2008) [7] had the goal to detect any possible correlations between measurable valence and arousal features and self- reported subjective experience. They had participants play an first person shooter game, while being measured with electroencephalography, electrocardiography, electromyography, galvanic skin response and eye tracking equipment. After each level, participants also had to fill in a questionnaire. They found that physiological responses can be an indicator of psychological states of gameplay experience.

These have been examples of research on flow in RTS and FPS games. Player flow modeling research for other game genres have also been done. Like [3] for horror games and [6] for online games. However, for many other game genres, there has not yet been any player modeling done with the focus on flow. One niche genre is a vampire survivors[8]-like, which can be categorized as a mix of the bullet hell and rogue-like genres, both of which also have little to no flow modeling research done on. This paper will hopefully contribute a bit to the overall knowledge on flow modeling for games with our approach for a vampire surivors-like game.

3 METHODS

3.1 Game implementation

As stated in the introduction our game is Vampire Survivors [8] inspired. Vampire Survivors is classified as a rogue-like, bullet hell game, where the player has to survive waves of incoming enemies. Another important gameplay feature is the levelling system, which is the primary driver to gain and upgrade items, as well as get passive boosts. So these are the core features that we use in the implementation of our game. We made our game in Unity version 2022.3.10f1.

Starting with the player character of our game, a cat (Figure 5). The player controls two facets of gameplay: the movement of the character and the choice of level-up reward. Thus the items that are used by the player operate independently and can only be influenced by player movement. We hope that this limited amount of mechanics allows the player to focus on survival. The player won't have to focus on their weapons as much, thus allowing them to focus on avoiding the enemies.

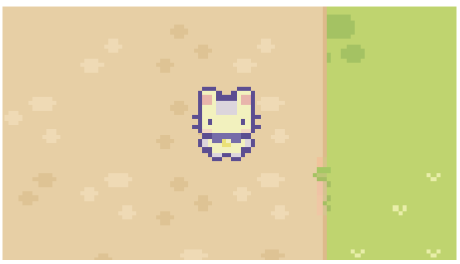
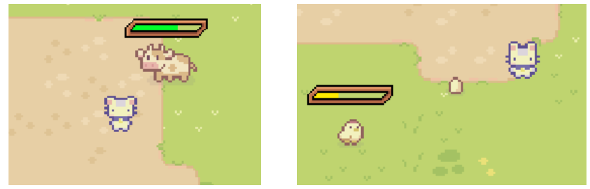


Figure 2: Player Character

Our game knows two enemies, a cow (Figure 3a) and a chicken (Figure 3b), each with different behaviours. The cow acts like a zombie and walks straight for the player to attack. The chicken is a ranged enemy that approaches the player to a certain distance, after which it throws eggs at the player. The chicken also has a run mechanic, where it will try to flee from the player if you get too close. The enemy spawning mechanism is done in waves and spawns enemies just outside the player's viewport up to a given maximum and checks this number on an interval to keep the number of enemies up. A wave thus has three descriptors: the size of the wave, a spawn buffer (the time between spawn checks) and a spawn chance, which gives the chance that a spawned enemy is a chicken or cow. We hope that this combination of enemies will push the player to play aggressively. In the sense that the player will actively seek upgrades and thus kill the enemies while moving around. Staying in a position will not only get the player surrounded, but the chickens will hunt the player down from a safe distance. The cows will however give some shielding for the chickens. So the player will need to find a way to cull the number of chickens all the while avoiding getting surrounded.

In our game, the level-up system and item system are intertwined. Each enemy drops some experience points, which can be picked up by the player. The player then levels up when a certain threshold of experience is reached. The player is then presented with a choice to strengthen their character. There are three types of presented choices: Taking a new weapon, upgrading a weapon and upgrading the player character.



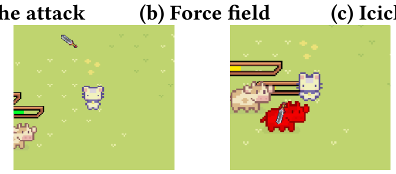
(a) Cow enemy **(b) Chicken enemy**

Figure 3: Enemies

For our game, we implemented four weapons: A scythe (Figure 4a), which attacks in front of the player, a throwing knife (Figures 4d and 4e), which after an interval, is thrown up and falls down in an arc, an icicle (Figure 4c), which is thrown at the nearest enemy in a fixed interval and lastly a forcefield (Figure 4b), which explodes on enemy contact after which it takes time to recover. Each of these weapons has unique upgrades that impact their size, attack speed or damage.

For the player character upgrades we consider three types: A defensive boost, a damage boost and a speed boost. For the level-up choice, three random rewards are picked from the available upgrades for the player to choose from. The randomness will hopefully create a new and interesting combination of choices for each run. This will hopefully allow the player to feel a sense of creativity. We also hope that the player gets more attached to their player character since they have a major impact on how the character develops. This will add to the survival experience we hope to emulate.

(a) Scythe attack **(b) Force field** **(c) Icicle attack**



(d) Knife attack, throw up **(e) Knife attack, fall down**

Figure 4: Weapons

For the in-game UI elements, we have a player health bar and a timer. This allows the player to keep track of their health and the player can figure out some information about the current wave. The game only has a main menu in which the player can start or close the game. The questionnaire and level-up menus, pause the game and show the questions and choices overlayed on the game screen. The level-up choices are text buttons, with a basic explanation of the choice. The questionnaire has a text box to show the question and a number of buttons below it to answer on a Likert scale. After a choice is made, the game is unpaused and the player can continue. We think that this will allow the player to stay in the flow of the game, while also giving a natural break and point to make the game more challenging.

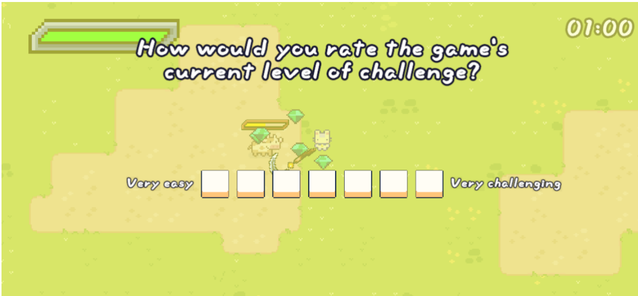


Figure 5: Questionaire prompt UI

3.2 Learning Model

We adopted a supervised learning model to estimate flow states during gameplay based on player in-game behaviors. The model was trained on gameplay metrics as proxies for player skill and challenge paired with brief self-reports of flow.

The input features consisted of gameplay events and performance metrics that reflect player skill and game difficulty, including enemies killed, damage dealt, accuracy, and time per wave. These objective metrics served as quantifiable proxies for the skill + difficulty component of flow.

The output targets were concise self-report questions administered after each wave gauging perceived challenge, either directly or indirectly via emotional state. These subjective labels indicated the flow state resulting from the skill-challenge balance.

By training a model to map from the input metrics to flow state targets, we aim to estimate flow in real-time based solely on observed gameplay behaviors, without requiring disruptive self-reporting during play.

The data was segmented into each wave, meaning each sample refers to a single wave.

Features (input) The following data was recorded, to act as a proxy of player skill and game difficulty.

- **Enemies Killed**: The number of enemies defeated by the player (in the last wave)
- **Damage Taken**: The amount of damage the player sustained (during the last wave)
- **Damage Taken per Second**: The rate at which the player's character sustains damage (during the last wave)
- **Damage Dealt**: The amount of damage the player inflicted damage to enemies (during the last wave)
- **Damage Dealt per Second**: The rate at which the player's character inflicted damage to enemies (during the last wave)
- **Experience Points Gained**: The amount of experience points picked-up during the last wave
- **Total Experience Points Gained**: The total amount of experience points picked-up
- **Weapon Levels**: What weapons the player has equipped (as of the end of the last wave), and what their upgrade level is. Unequipped weapons are marked as level -1 (equipped weapons start at level 0)
- **Weapon Damage**: The amount of damage done by each of the player's weapons (during the last wave)
- **Weapon Damage per Second**: The rate at which damage was done by each of the player's weapons (as of the end of the last wave)
- **Player Stats**: The player's stats (as of the moment of capturing the data), these include total health, movement speed, attack power, attack speed and defense points
- **Total waves cleared**: An integer representing the number of the currently completed waves

Labels (output) Validated scales exist to measure player experience [1] or flow [5] directly, but these scales have the downside of being fairly long questionnaires. Since we want to adaptively measure if the player is in flow, and thus measure the player's flow multiple times during their play session, these long questionnaires are not appropriate. Shorter versions exist (such as the Flow Short Scale [9]), but these remain too long for our use-case. Luckily, we do not have to use such a scale to measure 'flow', or 'player experience'. Since flow directly translates to perceived challenge [4], we can directly assess that. That said, perceived challenge is inherently subjective and thus requires the use self-reported measurements. Since we want to assess a player's flow state after every wave, it is beneficial to be assess this in a way that does not break the player's gameplay experience. Therefore we opt to use very brief integrated questionnaires, ensuring the player can quickly answer and return to their playing.

Validated scales exist to measure perceived challenge in games [4], but are still very long. Subsets of these could be used, but would still be either too long, or no longer representative of what they aim to measure. For our use-case it is necessary to ask as little questions as possible to avoid disrupting flow, therefore these existing scales are not a viable option. As such, we opted to design two questions ourselves: One question to assess challenge directly, and one to assess it indirectly. These questions go both ways, meaning they all follow the same spectrum (from boring to frustrating), as opposed to questions measuring separate parts of the scale (such as measuring boredom and frustration separately). This avoids confusing situations where the user is equally bored and frustrated, leaving the system without an indication of wether or not the game is too difficult or easy.

Question 1 - Assessing Challenge Directly: Players rate the challenge level on a 7 point likert scale ranging from "Very easy" to "Very challenging." This rating provides a direct indication of how the player perceives the game's difficulty in relation to their ability.

Question 2 - Assessing Challenge Indirectly: Players describe their emotional state during gameplay on a 7 point likert scale ranging from "Very frustrating" to "Very boring."

The labels here were selected to be in line with flow terminology, where the game being too challenging is referred to as frustrating and too little challenge being referred to as boring. This assessment offers insights into the player's emotional response to the game's challenge.

These self-report questions are administered at the start of every new wave, rather than at fixed time intervals. This ensures that each player's assessment occurs following the completion of specific in-game milestones and not in the midst of them playing the game, thus minimizing the disruption of their gameplay experience.

3.3 Data collection

To collect our data we asked participants to play our game until they either won, by surviving 5 minutes, or they gave up. When they died participants were given the opportunity to try again, for as many times as they want. The participants were chosen using convenience sampling, i.e. participants consist of fellow students, friends and family members. Before starting the game, participants signed a consent form (see appendix B) and given verbal instructions by the researchers on how the game works and what the controls are.

As stated before, every minute that the participant survived (also known as a wave) the game paused and they were presented two questions they had to answer. When this questionnaire is filled in all intermediary values of the features mentioned in section 3.2 are recorded along with the answers to the questionnaire. This data is written to a text file in CSV format where the headers are the features described in section 3.2 every row corresponds to a wave cleared. If a participant were to try again, the data from this new attempt is appended to the same CSV file starting at wave 1 again. From this data we only used the "best" attempt, i.e. the attempt were the participant cleared the highest number of waves.

Due to an oversight in the code wave 5 displays as wave 4 in the raw data. Fortunately we could easily fix this manually by changing any second successive fourth wave into a 5.

4 CREATING AND EVALUATING MODELS

The results contained 122 samples, of which 36 from clearing the first wave, 29 from clearing the second wave, 22 from clearing the third wave, 18 from clearing the fourth wave and 17 from clearing the final wave. This shows a skew towards the earlier waves.

Before proceeding with the construction and training of our models, we conducted an initial data analysis, which included an investigation into the correlation between two key self-report questions administered to players. The first question assesses challenge directly, asking participants to rate the challenge level from "Very easy" to "Very challenging". The second question asked participants to assess challenge indirectly, by rating their emotional state from "Very boring" to "Very frustrating". According to flow theory [2, 7, 10, 11], a high degree of frustration in the second question should correspond to a perception of high challenge in the first question. Our analysis revealed that indeed the two measures do correlate, though only moderately (pearson $r = 0.45$, $p < 0.00001$). This suggests that when players perceived the game as more challenging, they also reported feeling more frustrated. Conversely, when players perceived the game as easier, they reported feeling more bored.

Due to the moderate correlation between the two measures, we opted to train separate models for predicting direct perceived challenge, emotional state, and an averaged measure of both. This resulted in three supervised learning models:

- A model trained to predict direct perceived challenge based on the first self-report question
- A model trained to predict emotional state based on the second self-report question
- A model trained to predict the average of perceived challenge and emotional state

Table 1: Regression models for various targets, and their cross validated (5-fold) scores.

Target	Model	MAE	MSE	MdAE	R^2
Aggregated	DTR	1.27700	2.55083	1.10000	-1.36587
	GBC	0.98805	1.56761	0.85244	-0.43923
	LR	17263.77205	15307964281.03949	1.10127	-17587106854.36820
	RFR	0.91901	1.42335	0.76650	-0.30335
	SVR	0.88397	1.18154	0.76451	-0.01908
Challenge	DTR	2.08500	6.58233	1.90000	-1.82282
	GBC	1.58244	3.91236	1.36382	-0.69658
	LR	18233.12858	15590505260.51234	1.83328	-8046681358.43690
	RFR	1.50588	3.53033	1.38100	-0.49533
	SVR	1.44042	3.07358	1.41927	-0.23120
Emotion	DTR	0.78800	1.33133	0.70000	-0.72540
	GBC	0.69526	0.84082	0.53808	-0.04710
	LR	16294.74706	15528776925.52703	0.64764	-22368409705.59061
	RFR	0.64028	0.74746	0.53700	0.06148
	SVR	0.58865	0.64494	0.40042	0.21972

First, models were trained to predict the values treated as a continuous scale (from 1 to 7) using regression approaches. These approaches include Decision Tree Regression (DTR), Gradient Boosting Regression (GBR), Linear Regression (LR), Random Forest Regression (RFR), and Support Vector Regression (SVR). These models were trained and validated using 5-fold cross validation, the results of which can be seen in Table 1. These models performed poorly overall, with almost exclusively negative R^2 scores. The only models with a positive R^2 are models trained to predict the 'Emotion', with Support Vector Regression having the best fit ($R^2 = 0.22$), followed by Random Forest Regression ($R^2 = 0.06$). These two models have low MAE (0.59 for SVR, 0.64 for RFR), low MSE (0.64 for SVR, 0.75 for RFR) and low MdAE (0.40 for SVR, 0.54 for RFR), but unfortunately still have a relatively poor fit.

Overall, the regression models performed poorly at predicting the continuous target variables, with R^2 scores being negative for nearly all models. This indicates that the models failed to fit the training data well. The only exceptions were the RFR and SVR models trained to predict emotion ratings, which achieved slightly positive R^2 scores of 0.06 and 0.22 respectively. However, these values still represent relatively weak model fits. These two models performed best when looking at other evaluation metrics too, with the SVR model for predicting emotion producing the lowest mean absolute error (MAE) of 0.59, mean squared error (MSE) of 0.64, and median absolute error (MdAE) of 0.40. The RFR emotion model had the next best scores of 0.64, 0.75, and 0.54 for MAE, MSE, and MdAE respectively. While these errors are low in magnitude, the poor R^2 values indicate these models are not sufficiently capturing the variation in the true target variables. The linear regression model performed the worst and exhibits extreme scores, likely due to the scales not following a linear relationship with the input data.

Given the poor performance of the regression models, a classification-based approach was explored next. The continuous 1-7 challenge and emotion ratings were binned into discrete categories:

- Challenge: Easy (1-3), Neutral (4), Challenging (5-7)
- Emotion: Boring (1-3), Neutral (4), Frustrating (5-7)
- Aggregate: Easy/Boring (1-3), Neutral (4), Challenging/Frustrating (5-7)

By converting the continuous ratings into categorical classes, the modeling task becomes predicting which class each sample belongs to rather than regressing an exact rating value. Six classification algorithms were evaluated: Decision Tree Classifier (DTC), Gradient Boosting Classifier (GBC), K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest Classifier (RFC), and Support Vector Classifier (SVC). As with the regression models, 5-fold cross-validation was used for training and validation, the results of which can be found in table 2.

The best performing models per target were:

- For predicting perceived challenge, the Random Forest Classifier (RFC) achieved the highest scores across all evaluation metrics with an accuracy of 0.48, recall of 0.48, precision of 0.42, and F1 score of 0.44. The RFC model was able to moderately distinguish between easy, neutral, and challenging gameplay situations based on the gameplay metrics.
- For predicting player emotional state, the K-Nearest Neighbors (KNN) classifier attained the best performance with an accuracy of 0.56, recall of 0.56, precision of 0.51, and F1 score of 0.51. Though still not extremely high, these results were the highest across all of the models. The KNN model was able to moderately categorize gameplay data into boring, neutral or frustrating emotional experiences. The second best performing one for this category was RFC.
- When predicting the aggregated categories, the Logistic Regression (LR) model achieved the top accuracy of 0.47 and recall of 0.47. However, the RFC model had slightly higher precision of 0.43 and F1 score of 0.43 for this target. Overall, performance for the aggregated variable was comparable between the LR and RFC classifiers.

Despite KNN being the best for emotional state, we can see that Random Forest Classifier based models perform very well (relatively speaking) across the board.

Due to the nature of the classification task at hand here, where the model has to distinguish between one of three options, an accuracy, recall, precision and F1 score of over 0.33 is necessary - as that is the performance of a model that randomly guesses. Most models achieve this. However, the scores don't exceed 0.33 by a lot, and the highest performing one, KNN, still only has an accuracy of 0.56, recall of 0.56, precision of 0.51, and F1 score of 0.51.

Table 2: Classification models for various targets, and their cross validated (5-fold) scores. The highest scoring entry in each scoring metric is bolded.

Target	Model	Accuracy	Recall	Precision	F1
Aggregated (Cat.)	DTC	0.37600	0.37600	0.39242	0.37860
	GBC	0.38433	0.38433	0.36705	0.36742
	KNN	0.36167	0.36167	0.37659	0.35340
	LR	0.46767	0.46767	0.33455	0.36626
	RFC	0.45100	0.45100	0.42941	0.42734
	SVC	0.44267	0.44267	0.19619	0.27181
Challenge (Cat.)	DTC	0.39433	0.39433	0.40240	0.38611
	GBC	0.41833	0.41833	0.38254	0.38544
	KNN	0.34433	0.34433	0.31990	0.32897
	LR	0.46833	0.46833	0.40723	0.39383
	RFC	0.47667	0.47667	0.42349	0.43833
	SVC	0.47600	0.47600	0.22744	0.30756
Emotion (Cat.)	DTC	0.36933	0.36933	0.37680	0.36770
	GBC	0.43400	0.43400	0.40295	0.40596
	KNN	**0.55767**	**0.55767**	**0.50921**	**0.51474**
	LR	0.49967	0.49967	0.31390	0.36685
	RFC	0.53267	0.53267	0.45775	0.48040
	SVC	0.49133	0.49133	0.26386	0.34205

5 DISCUSSION

The classification models results, outperforming the regression approaches, as well as the baseline prediction of a random model. The RFC and KNN models in particular were able to moderately distinguish between categories for the perceived challenge and emotional state targets respectively.

These results suggest that mapping gameplay data to discrete categorical flow states may be more feasible than directly predicting precise ratings on a continuous scale. The moderate classification performance indicates potential for estimating players' overall flow status based on behaviors, although there is substantial room for improvement.

However, one problem of trying to predict player emotional state directly from gameplay data, is the possibility of multiple play styles. One participant might choose to use a completely different play style than another participant. For example, running away instead of actually fighting the enemies and leveling up your skills. This might cause drastically different gameplay data, even though they may have a similar feeling of challenge or emotion. For example, one participant might find challenge in dodging all the enemies, while another might find challenge in timing your attacks to fight the enemies.

The study also had some limiting factors. One of these factors is the fact that there were not many participants and thus the amount of data is sparse. Having more data points could lead to more accurate models, especially with recognizing the different playstyles. Another limitation is that most participants did not make it to the end of the game. This leads to a skew towards data collected in the earlier waves of the game, which may therefore bias the models. On the other hand, if a participant did not complete the first wave, there was no way to collect their data. This limitation caused the loss of some important data of participants that had a hard time with the game, and were perhaps far from a flow state. Some other limitations are related to the scales used to determine the participants current emotional state. One of these limitations was related to players not being able to express enjoyment of the game, as there was no clear option to mention that. Another issue was that frustration and boredom are not mutually exclusive, which was insinuated by the scales. This led to some confusion among certain participants. Finally, even though the scales were still found to be correlating to flow theory, they were not validated. Therefore, validating the scales or using validated ones might yield better results.

These limitations allow for improvements in future work. By collecting more data and also making sure there is an even amount of data for each wave, the models can be trained on more and higher quality data. This could lead to better performing models. Future work can also improve upon this paper by validating and extending the scales used for monitoring player emotional state.

6 CONCLUSION

In this paper, we presented two types of models to predict the flow state of a player. The regression models performed poorly on the collected data, with negative R^2 scores. Therefore a classification approach was attempted instead, which resulted in more modest results. The classification models performed better than a random model, however the accuracies are still only ranging from 0.34 to 0.56. The best performing model was the KNN model, which obtained an accuracy of 0.56 on the predictions of player emotion. In general the models performed better on the prediction of emotion than on the prediction of challenge. To conclude, the classification models outperformed the regression models on the collected dataset, and performed the best on the prediction of emotion.

REFERENCES

[1] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies* 135 (2020), 102370. https://doi.org/10.1016/j.ijhcs.2019.102370

[2] Jenova Chen. 2006. *Flow in games.* Ph. D. Dissertation. http://jenovachen.info/abstract

[3] Edirlei Soares de Lima, Bruno M.C. Silva, and Gabriel Teixeira Galam. 2022. Adaptive virtual reality horror games based on Machine learning and player modeling. *Entertainment Computing* 43 (2022), 100515. https://doi.org/10.1016/j.entcom.2022.100515

[4] Alena Denisova, Paul Cairns, Christian Guckelsberger, and David Zendle. 2020. Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS). *International Journal of Human-Computer Studies* 137 (2020), 102383. https://doi.org/10.1016/j.ijhcs.2019.102383

[5] Susan A. Jackson and Herbert W. Marsh. 1996. Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *Journal of Sport and Exercise Psychology* 18, 1 (March 1996), 17–35. https://doi.org/10.1123/jsep.18.1.17

[6] Chuang-Chun Liu. 2017. A model for exploring players flow experience in online games. *Information Technology & People* 30, 1 (2017), 139–162.

[7] Lennart Nacke and Craig A. Lindley. 2008. Flow and Immersion in First-Person Shooters: Measuring the Player's Gameplay Experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share* (Toronto, Ontario, Canada) *(Future Play '08)*. Association for Computing Machinery, New York, NY, USA, 81–88. https://doi.org/10.1145/1496984.1496998

[8] Poncle. 2021. *Vampire Survivors by Poncle.* itch.io. https://poncle.itch.io/vampire-survivors

[9] Falko Rheinberg, Regina Vollmeyer, Stephane Engster, and Radhika Sreeramoju. 2023. FSS - Flow Short Scale (English Version). (09 2023).

[10] Penny Sweetser, Daniel Johnson, and Peta Wyeth. 2012. Revisiting the GameFlow model with detailed heuristics. *Journal of Creative Technologies* 2012, 3 (2012), 1–16. https://eprints.qut.edu.au/58216/

[11] Penelope Sweetser and Peta Wyeth. 2005. GameFlow: A Model for Evaluating Player Enjoyment in Games. *Comput. Entertain.* 3, 3 (jul 2005), 3. https://doi.org/10.1145/1077246.1077253

[12] Georgios N Yannakakis, Pieter Spronck, Daniele Loiacono, and Elisabeth André. 2013. Player modeling. (2013).

[13] Georgios N Yannakakis and Julian Togelius. 2018. *Artificial intelligence and games.* Vol. 2. Springer.

A ETHICS AND PRIVACY QUICK SCAN

Response Summary:

Section 1. Research projects involving human participants

P1. Does your project involve human participants? This includes for example use of observation, (online) surveys, interviews, tests, focus groups, and workshops where human participants provide information or data to inform the research. If you are only using existing data sets or publicly available data (e.g. from Twitter, Reddit) without directly recruiting participants, please answer no.

- Yes

Recruitment

P2. Does your project involve participants younger than 18 years of age?

- No

P3. Does your project involve participants with learning or communication difficulties of a severity that may impact their ability to provide informed consent?

- No

P4. Is your project likely to involve participants engaging in illegal activities?

- No

P5. Does your project involve patients?

- No

P6. Does your project involve participants belonging to a vulnerable group, other than those listed above?

- No

P8. Does your project involve participants with whom you have, or are likely to have, a working or professional relationship: for instance, staff or students of the university, professional colleagues, or clients?

- Yes

P9. Is it made clear to potential participants that not participating will in no way impact them (e.g. it will not directly impact their grade in a class)?

- Yes

Informed consent

PC1. Do you have set procedures that you will use for obtaining informed consent from all participants, including (where appropriate) parental consent for children or consent from legally authorized representatives? (See suggestions for information sheets and consent forms on the website.)

- Yes

PC2. Will you tell participants that their participation is voluntary?

- Yes

PC3. Will you obtain explicit consent for participation?

- Yes

B INFORMATION AND CONSENT SHEET

Information Sheet

We are conducting research to model player flow in a game inspired by Vampire Survivors. In this game, players control characters in an arena, facing periodic enemy spawns and collecting experience points to upgrade weapons.

Flow is a state of complete absorption and optimal experience, achieved when the challenge of an activity matches one's skills. We aim to assess whether players are in a state of flow during the game. We will collect gameplay data, including health changes, enemies defeated, movement patterns, and item choices. Additionally, the game will prompt you to answer two 7 point likert scale questions in between every wave. This data will be used to train a machine learning model that estimates the player's flow state.

Our goal is to adapt the game dynamically to keep players in the flow state, enhancing their gaming experience.

Consent Form

The students conducting the experiment: Samuel Spithorst (5956684), Stefan Hoekzema (6896383), Vos Wesseling (6885373), Jeppe Vroegindeweij (6955266), Lars de Kwant (6958680)

For questions after this evaluation, you can contact Samuel Spithorst at s.f.spithorst@uu.nl

Please complete the form below by ticking the relevant boxes and signing on the line below. A copy of the completed form will be given to you for your own record.

- ☐ I confirm that the experiment has been explained to me and I have had the opportunity to ask questions which were answered satisfactorily.
- ☐ I am voluntarily taking part in this evaluation. I understand that I don't have to take part, and I can stop my participation at any time;
- ☐ I don't expect to receive any benefit or payment for my participation;
- ☐ I confirm that I am 18 years of age or over.
- ☐ I understand that the information/data acquired will be securely stored by the researchers, but that appropriately anonymized data may in the future be made available to others for research purposes only.
- ☐ Participating (or choosing not to) will in no way impact your grade in class
- ☐ I understand that I can request any of the data collected from/by me to be deleted.
- ☐ I agree to take part in this experiment.

Name of participant: Date: Signature:

Name of researcher: Date: Signature: