A Data-Driven Approach to Deduce Players' Preferences from In-Game Interactions in Gameful Systems

[1]Enrica Loria
[1] Annapaola Marconi
[1]Fondazione Bruno Kessler (FBK)
Via Sommarive, 18, 38123 Povo (TN), Italy
eloria@fbk.eu

Abstract

The way players interact with the system during the gameplay is a valuable source of data. While playing, they encounter a plethora of choices and, when they select an option rather than another, they intrinsically express a preference. This knowledge can be exploited in the analysis phase of the game to fine-tune the game towards each user, either manually or through algorithms. However, misinterpreting the players' actions could be very harmful. It could lead to misleading choices and could result in players' abandonment rather than in a better experience. In this preliminary work, we present a methodology to infer player preferences by analyzing the choices made during the gameplay. We gathered data from an on-the-field gameful system and applied the paired comparison technique to compute a preference score for the options available. We, then, tested the reliability of the method by computing players' consistency in their choices, and, partially, by predicting their last choice based on their previous selections.

Introduction

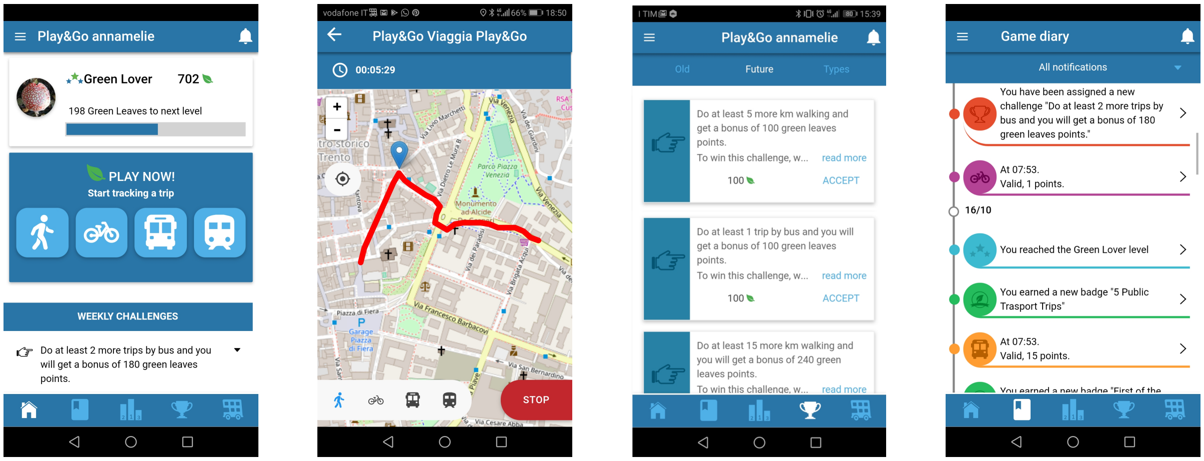
The term gamification or gameful system refers to the implementation of mechanics and elements usually found in games in other application domains (Huotari and Hamari 2012; 2017), persuasive technologies being an example. In the last decade, in which gamification has been particularly prolific, gameful systems have proven to mostly have a positive effect (Xi and Hamari 2019). However, the constant observation made by researchers is that, as in games, players answer differently to the same game mechanics (Orji, Mandryk, and Vassileva 2017). Thus, tailoring the content to the user that is interacting with it enhances the power of the instrument itself (Charles et al. 2005; Jacobs 2016), resulting in a higher retention and engagement level. Such tailoring, or customization, could both refer to the adaptation of difficulty level - i.e., the concept of flow (Csikszentmihalyi and Csikszentmihalyi 1992) - and to the user preferences in terms of game elements deployed.

Many theoretical models to classify players have been suggested (e.g., (Caillois 2001; Marczewski 2015; Yee 2016)) to assist the process of generating customized content better suited for each user. To this purpose also data-driven approaches (Hooshyar, Yousefi, and Lim 2018) have been studied, which exploit information retrieved transparently from players in-game behaviors - i.e., gameplay datalogs. Those datalogs often contain a huge amount of data and must be elaborated to be interpretable (Hooshyar, Yousefi, and Lim 2018). Such elaboration often lead to the definition of player profiles, and thus, to behavioral analysis.

Behavioral analysis and player modeling can be a powerful tool in the process of dynamically adapting the game, which allows the generation of game content of heterogeneous nature (Hendrikx et al. 2013). Keeping the players entertained (Khajah et al. 2016) can be achieved by adjusting the difficulty of the levels to avoid frustration and/or boredom (Lora et al. 2016) and tuning the system to the players' knowledge (Zook et al. 2012).

In this work, we aim at contributing to the Game User Research community by suggesting a method to deduce players' preferences by analyzing their choices during the gameplay. To this extent, we retrieved and studied data from a long-running on-the-field gamification campaign, in which players engaged in weekly challenges. In particular, they could set specific parameters for their challenges by choosing among two or three options every week. We used a paired comparison protocol to compute players' preferences score, which has proven to be more reliable than direct rating (Perez-Ortiz and Mantiuk 2017; Shah et al. 2016). Then, we used the computed preferences to predict players' last in-game choice. We argue that such a method could be useful in building a standard way to analyze player in-game behavior to infer preferences, in terms of game elements. The outcomes could be used by designers in the analysis of their games, by understanding which game elements have been more appreciated and from whom. Moreover, this could be useful for researchers that cannot (or do not want to) rely on self-assessments. They can use such deduced preferences to evaluate their algorithms recommending tailored content.

The remaining of the paper is structured as follows. We introduce some relevant works from the literature and describe



(a) (b) (c) (d)

Figure 1: **Screenshots.** Basic interactions in Play&Go.

the use case scenario employed for the analysis. Then, we present a description of the study conducted and preliminary analysis. Conclusion and the plan/ideas for future works follow.

Related Works

Game Analytics (El-Nasr, Drachen, and Canossa 2016) procedures aim at studying players' in-game behaviors to achieve a better understanding of the users interacting with the system. Among the most widely exploited source of information, there are gameplay data (El-Nasr, Drachen, and Canossa 2016). It has the advantages of being highly knowledgeable, but yet transparent to the user and easily provided to the stakeholders involved in the design process - e.g., through automatic logging system (Kim et al. 2008). Players' experience can be modeled following both theory-driven or data-driven approaches (Yannakakis and Togelius 2011). Despite theory-driven approaches supply a rationale to the choices of the model, they lack in adaptability, which counts as a key feature in contexts such as games, that are versatile by definition. As a consequence, data-driven approaches are often preferred. Being an objective measure, data-driven methods reduce the possibility to introduce biases in the analysis (Sangin 2018) and favors the implementation of gameplay metrics aimed at the improvement of the game itself (Kim et al. 2008).
Player behaviors extracted from gameplay data (e.g., (Loria and Marconi 2018), have been investigated for different purposes. Examples of applications go from the study of players general models (Canossa 2013) to the detection of specific patterns for instructing non-playing characters from real players' actions (Weber, Mateas, and Jhala 2011). Other researchers aimed at predicting the strategies that the players were using (Hsieh and Sun 2008) and reducing the probability of players churn (Thurau, Bauckhage, and Sagerer 2003; Mahlmann et al. 2010), resulting in retention's increase. Such retention can be fostered by providing players with content that they enjoy.
Several studies in this field show that users can be more or less receptive to different game elements (Busch et al. 2015; Kaptein et al. 2012; Monterrat, Lavoué, and George 2017). These studies suggest that user personality and preferences have a great influence on the effect that game elements have on user motivation. Generating appropriate game elements can lead to higher levels of user motivation, whereas inappropriate game elements can demotivate users. The tailoring of the game content can be informed by modeling players' behaviors. Zook et al. (Zook and Riedl 2012), for example, presented a tensor factorization approach to model and predict players' performance on skill-based tasks to dynamically adjust the difficulty of a game, providing motivating context for the adjustments provided. Zhang et al. (Zhang, Brown, and Shankar 2016) proposed a fully data-driven approach to construct behavioral personas from raw clicks gathered in telemetry data from our product use. Another example of application is the work of Martinez et al. (Martínez, Hullett, and Yannakakis 2010): a neuro-evolutionary preference learning embedding a player modeling module for the prediction of player preferences. However, in their framework players are requested to explicitly report their preferences on variants of the game via questionnaires. Then, computational models are built on the preference data. The approach of employing a survey-based analysis to identify users' preferred game elements and mechanics has also been used in other studies (e.g., (Tondello, Mora, and Nacke 2017)). However, the main critique moved towards questionnaires, also referred to as subjective PEM, is their tendency to be biased due to human error (El-Nasr, Drachen, and Canossa 2016). In contrast, the paired comparison protocol is more reliable than direct rating (Shah et al. 2016; Perez-Ortiz and Mantiuk 2017) and, most importantly, can be used when explicit self-declarations cannot be retrieved.
Gamification has been increasingly used to enhance user motivation (Ferro, Walz, and Greuter 2013; Looyestyn et al. 2017). Recent work on tailored gamification (Jia et

al. 2016) has provided valuable results that identify links between user types and relevant or motivating game elements. The advantages of tailored gamification have been proved in several fields, such as health (Orji, Mandryk, and Vassileva 2017) and sport (Lopez and Tucker 2018) and learning (Kickmeier-Rust, Hillemann, and Albert 2014; de Marcos, Garcia-Lopez, and Garcia-Cabot 2016; Legaki et al. 2019). Adaptive gamification is a young and growing field, especially for what concerns dynamic - in contrast to static - adaptation (Stuart et al. 2019). Thus, the literature is still limited and would benefit from new works, also inspired by methodologies already applied to games.

Use Case Scenario

The gameful system analyzed is a mobile gameful system that pushes its users to behave sustainably during their daily journeys. The game, Play&Go, awards its players points when they track their movement, according to the level of sustainability of the selected mean of transportation (walking is better than using the bus). Every week, a challenge is assigned to each player, intending to keep her engaged (Nakamura and Csikszentmihalyi 2014) in the game while fostering improvement in performance. The challenge is characterized by a parameter (walk kilometer, bike kilometer, bus trips, train trips or green leaves points), a target value, and a bonus obtained in case of a win. The target and the bonus define the difficulty of the challenge, which is calculated on the player's previous performance. The difficulty of the challenge is structured in 3 levels (from 1 to 3) and is transparent to the user. An example of a challenge instance is: "Track at least 5 kilometers by bike to win a bonus of 100 Green Leaves points".

For inexperienced players (Level 1 players), the challenge is automatically assigned by a recommendation system (RS), which sets the difficulty and the parameter according to the (limited) history of the players' tracings. When players reach Level 2, they can program their challenge week by week, by choosing among two options, provided by the recommendation system. When the players reach Level 3, the options to choose from become three. Since the challenges are produced by an RS, every player has a different set of challenges to choose from. The challenges proposed may vary on the parameter upon which the challenge is evaluated, and/or the level of difficulty of the challenge itself. Players have to submit their choice before the following game week starts or a default challenge is assigned.

During the 6-month gamification campaign conducted, about 500 players tracked at least 1 trip in the game. In total, 64k trips and 250k kilometers were traced. Of those trips, the vast majority were tracked by foot, followed by train, bus, and bike. Therefore, transportation means we call zero impact were the favored ones.

Among the active players, 160 programmed at least one weekly challenge, by actively choosing one of the options presented. However, in our study, we selected players that made at least 3 choices, which resulted in a final sample of 115 players. This filtering was required because we needed them to make at least 2 choices to compute our scores and the third was required to perform the prediction analysis (see Study Design Section). For the players selected for the study, we counted a mean of 87 active days, with a standard deviation of 13 days; which is equal to about 12 weeks of gameplay. During the gameplay, those players walked on average 332 km (std = 100 km); biked for 230 Km on average (std = 110 km); and traveled on bus and train for on average 204 (std = 156) km, and 505 (std = 338) km respectively.

The limited number of users eligible for the study must not be interpreted as a lack of interest in the game. Together with the elevated number of kilometer tracked in the game, we gathered a general positive response of players towards the game. At the end of the edition of the game, we allowed players to evaluate their experience through a self-assessment. The majority - 66% - of our users declared to have changed their mobility behavior, thanks to the app, to a medium-to-high extent. On the overall, player experience was rated positively by 85% of the active users.

Instead, it should be noted that the programming mechanism is an instance of a *customization game mechanics*, which as many theoretical and practical player modeling studies state (e.g., (Marczewski 2015)), attracts only a portion of the population of players.

Study Design

In our use case scenario, the game elements that can be customized are the weekly challenges, for which a *parameter* has to be defined. This parameter represents the counter upon which the challenge will be evaluated: walk kilometers, bike kilometers, green leaves points, train trips, and bus trips. Although the counters available were 5, some players may have been exposed to fewer counters, during the challenge selection. This is because the options are produced by an RS, exploiting players' in-game history. Thus, if a player never traced a train trip, probably the RS will never produce a challenge with the train counter. In this study, for each player, we considered only the subset of parameters used to produce her challenges' suggestions.

Our study seeks to deduce the players' preferred counter(s), by evaluating the choices they made during the game. We used the paired comparison protocol, an important tool in multi-attribute decision making. As classically done in paired comparison studies (Perez-Ortiz and Mantiuk 2017), we build a squared matrix of dimension n for every player, with n the number of parameters available. In such a matrix, a value of x for the cell c_{ij} means that the counter i was chosen x times over the counter j. In our study we dealt with an imbalanced design; i.e., every observer has a different experimental design, instead of the complete set of possible comparisons. Although full designs are preferable, incomplete designs can reach competitive performance (Bozóki, Csató, and Temesi 2016).

To compute the preference scores, we can then consider the number of votes received by each counter, which may then be divided by the number of comparisons per counter for normalization purposes (Perez-Ortiz and Mantiuk 2017). At the end of the process, for every player, we obtained a vector w of dimension n, with the preference scores for every

counter.
In the literature we can find more complex statistical methods exist for producing such scores, they were not shown to give a more accurate outcome than the vote counts (Thurstone 1927).
As we mentioned in the *Use Case Scenario* Section, each challenge has also associated a difficulty level (from 1 to 3), according to the estimated effort for the considered players to win the challenge, computed on her performance. We conducted another analysis also considering the customizable parameter as the couple (counter, difficulty) - e.g., (walk kilometers, 1), (walk kilometers, 1), (bike kilometers, 3), (green leaves, 1). The intuition that led to the introduction of the difficulty of the challenge is that the level of effort required from the challenge can also be an important discriminant in the choice. Thus, the results should have been more accurate.

Preliminary Evaluation

To evaluate the reliability of each participant, we computed their consistency, by calculating the number of intransitive comparisons occurring, known as cyclic triads. In other terms, a cyclic triad occurs when a is preferred to b, b is preferred to c, but c is preferred to a. The coefficient of consistency (Kendall and Smith 1940), for every user, is computed with the following formula: $\zeta = 1 - \frac{24*c}{z^3-z}$ where z is the number of choices made and c is the number of cyclic triads. When there is perfect consistency - i.e. no circular triads - $\zeta = 1$. The value of ζ will get smaller as the number of circular triads increases, and thus the consistency worsens. A consistency higher of 0.75 is considered good (Leloup et al. 2010). This limit allows both human errors - e.g., players clicking the wrong button - and exploratory choices - e.g., selecting an option out of curiosity towards something new.
From the 115 players selected for this analysis, only 1 had a $\zeta = 0.35$, with the remaining having a $\zeta \geq 0.79$.
We also computed ζ for the analysis including the difficulty of challenges as a customization parameter - e.g., when the customization was evaluated on the pairs (counter, difficulty). This time, we had $\zeta \geq 0.92$

We also partially evaluated the outcomes by predicting the last choice made by players. For every player, we computed the preference scores w on the $N-1$ weeks in which they made a choice. The N^{th} week was used to make the prediction. We listed the 2, or 3, options that were presented to each user and, using her scores w we predicted the option that most likely selected (the one with the highest preference score). Then we compared, the forecasted with the actual choice. Results showed a 70% of accuracy, considering the 115 players. The accuracy increased at 71% when considering players that made more than 10 choices - 94 players. When also the difficulty was considered, the value raised to 82%. This outcome is very informative. It should be considered that introducing the difficulty as a parameter brought to an increase in the number of counters available (actual counter + level of difficulty) while using the same amount of data. Thus, the number of available choices made per player and per counter was reduced, since they were scattered among the new options. Nevertheless, the prediction was more than 10 percentile points more accurate.

Discussion

In this preliminary work, we presented a methodology to infer from in-game behaviors, player preferences. Our results show almost all players manifested a high consistency score, and thus, that the in-game choices were made with criterion. Besides, we managed to predict the final choice of the majority of our players, proving that this method is indeed effective for computing preferences scores.
Such a method could be applied in gameful systems in several ways. In the first instance, designers may be interested in knowing which game elements attracted the majority of her audience, and use this information to characterize her players. A more visionary application could be using such preference scores to assist or validate machine learning algorithms, aimed at producing tailored content.

Conclusion and Future Works

We conducted our study by analyzing data from an on-the-field gameful system, which boasts a conspicuous participation. The biggest flaw lies in the modest numbers of players that made enough in-game choices to be considered in the study. This issue is relevant when validating the reliability of the tool through prediction. However, the high scores in the consistency values prove that the choices have been made knowingly. Therefore, this measure can be considered as an indicator of preference.
In future works, the priority lies in expanding the dataset with more participants to have enough data to conduct an accurate prediction analysis. Equally important is also ensuring that a complete pairwise comparison study can be conducted, to further prove the reliability of the consistency score.

References

Bozóki, S.; Csató, L.; and Temesi, J. 2016. An application of incomplete pairwise comparison matrices for ranking top tennis players. *European Journal of Operational Research* 248(1):211–218.

Busch, M.; Mattheiss, E.; Orji, R.; Marczewski, A.; Hochleitner, W.; Lankes, M.; Nacke, L. E.; and Tscheligi, M. 2015. Personalization in serious and persuasive games and gamified interactions. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 811–816. ACM.

Caillois, R. 2001. *Man, play, and games*. University of Illinois Press.

Canossa, A. 2013. *Game Analytics: Maximizing the Value of Player Data*. Springer London. 255–283.

Charles, D.; Mcneill, M.; Mcalister, M.; Black, M.; Moore, A.; Stringer, K.; Kücklich, J.; and Kerr, A. 2005. Player-centred game design: Player modelling and adaptive digital games.

Csikszentmihalyi, M., and Csikszentmihalyi, I. S. 1992. *Optimal experience: Psychological studies of flow in consciousness*. Cambridge university press.

de Marcos, L.; Garcia-Lopez, E.; and Garcia-Cabot, A. 2016. On the effectiveness of game-like and social approaches in learning: Comparing educational gaming, gamification & social networking. *Computers & Education* 95:99–113.

El-Nasr, M. S.; Drachen, A.; and Canossa, A. 2016. *Game Analytics*. Springer.

Ferro, L. S.; Walz, S. P.; and Greuter, S. 2013. Towards personalised, gamified systems: an investigation into game design, personality and player typologies. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, 7. ACM.

Hendrikx, M.; Meijer, S.; Van Der Velden, J.; and Iosup, A. 2013. Procedural content generation for games: A survey. *ACM Trans. Multimedia Comput. Commun. Appl, February 2013* 9(1):1–22.

Hooshyar, D.; Yousefi, M.; and Lim, H. 2018. Data-driven approaches to game player modeling: a systematic literature review. *ACM Computing Surveys (CSUR)* 50(6):90.

Hsieh, J.-L., and Sun, C.-T. 2008. Building a player strategy model by analyzing replays of real-time strategy games. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, 3106–3111. IEEE.

Huotari, K., and Hamari, J. 2012. Defining gamification: a service marketing perspective. In *MindTrek 16*, 17–22. ACM.

Huotari, K., and Hamari, J. 2017. A definition for gamification: anchoring gamification in the service marketing literature. *Electronic Markets* 27(1):21–31.

Jacobs, R. S. 2016. Play to win over: Audiences and effects of persuasive games. In *Etmaal van de Communicatiewetenschap 2016: Closer: Connecting through intimate communication technologies*.

Jia, Y.; Xu, B.; Karanam, Y.; and Voida, S. 2016. Personality-targeted gamification: a survey study on personality traits and motivational affordances. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2001–2013. ACM.

Kaptein, M.; De Ruyter, B.; Markopoulos, P.; and Aarts, E. 2012. Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2(2):10.

Kendall, M. G., and Smith, B. B. 1940. On the method of paired comparisons. *Biometrika* 31(3/4):324–345.

Khajah, M. M.; Roads, B. D.; Lindsey, R. V.; Liu, Y.-E.; and Mozer, M. C. 2016. Designing engaging games using bayesian optimization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5571–5582.

Kickmeier-Rust, M. D.; Hillemann, E.-C.; and Albert, D. 2014. Gamification and Smart Feedback: Experiences with a Primary School Level Math App. *International Journal of Game-Based Learning (IJGBL)* 4(3):35–46.

Kim, J. H.; Gunn, D. V.; Schuh, E.; Phillips, B.; Pagulayan, R. J.; and Wixon, D. 2008. Tracking real-time user experience (true): a comprehensive instrumentation solution for complex systems. In *CHI'08*, 443–452. ACM.

Legaki, N. Z.; Xi, N.; Hamari, J.; and Assimakopoulos, V. 2019. Gamification of The Future: An Experiment on Gamifying Education of Forecasting.

Leloup, F. B.; Pointer, M. R.; Dutré, P.; and Hanselaer, P. 2010. Geometry of illumination, luminance contrast, and gloss perception. *JOSA A* 27(9):2046–2054.

Looyestyn, J.; Kernot, J.; Boshoff, K.; Ryan, J.; Edney, S.; and Maher, C. 2017. Does gamification increase engagement with online programs? a systematic review. *PloS one* 12(3):e0173403.

Lopez, C., and Tucker, C. 2018. Towards Personalized Adaptive Gamification: A Machine Learning Model for Predicting Performance. *IEEE Transactions on Games* 1–1.

Lora, D.; Sánchez-Ruiz, A. A.; González-Calero, P. A.; and Gómez-Martín, M. A. 2016. Dynamic difficulty adjustment in tetris. In *FLAIRS Conference*, 335–339.

Loria, E., and Marconi, A. 2018. Player types and player behaviors: Analyzing correlations in an on-the-field gamified system. In *CHIPlay'18 Extended Abstracts*, 531–538. ACM.

Mahlmann, T.; Drachen, A.; Togelius, J.; Canossa, A.; and Yannakakis, G. N. 2010. Predicting player behavior in tomb raider: Underworld. In *Computational Intelligence and Games (CIG), 2010 IEEE Symposium on*, 178–185. IEEE.

Marczewski, A. 2015. *Even Ninja Monkeys Like to Play*. Blurb.

Martínez, H. P.; Hullett, K.; and Yannakakis, G. N. 2010. Extending neuro-evolutionary preference learning through player modeling. In *Proceedings of the 2010 IEEE conference on computational intelligence and games*, 313–320. IEEE.

Monterrat, B.; Lavoué, E.; and George, S. 2017. Adaptation of gaming features for motivating learners. *Simulation & Gaming* 48(5):625–656.

Nakamura, J., and Csikszentmihalyi, M. 2014. The concept of flow. In *Flow and the foundations of positive psychology*. Springer. 239–263.

Orji, R.; Mandryk, R. L.; and Vassileva, J. 2017. Improving the efficacy of games for change using personalization models. *TOCHI* 24(5):32.

Perez-Ortiz, M., and Mantiuk, R. K. 2017. A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint arXiv:1712.03686*.

Sangin, M. 2018. *Games user research*. Oxford University Press.

Shah, N. B.; Balakrishnan, S.; Bradley, J.; Parekh, A.; Ramchandran, K.; and Wainwright, M. J. 2016. Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. *The Journal of Machine Learning Research* 17(1):2049–2095.

Stuart, H.; Serna, A.; Marty, J.-C.; and Lavou, E. 2019. Adaptive gamification in education: A literature review of current trends and developments. In *European Conference on Technology Enhanced Learning (EC-)*.

Thurau, C.; Bauckhage, C.; and Sagerer, G. 2003. Combining self organizing maps and multilayer perceptrons to learn bot-behaviour for a commercial game. In *GAME-ON*, 119.

Thurstone, L. L. 1927. A law of comparative judgment. *Psychological review* 34(4):273.

Tondello, G. F.; Mora, A.; and Nacke, L. E. 2017. Elements of gameful design emerging from user preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 129–142. ACM.

Weber, B. G.; Mateas, M.; and Jhala, A. 2011. Using data mining to model player experience. In *FDG Workshop on Evaluating Player Experience in Games*.

Xi, N., and Hamari, J. 2019. Does gamification satisfy needs? a study on the relationship between gamification features and intrinsic need satisfaction. *International Journal of Information Management* 46:210–221.

Yannakakis, G. N., and Togelius, J. 2011. Experience-driven procedural content generation. *IEEE Transactions on Affective Computing* 2(3):147–161.

Yee, N. 2016. The gamer motivation profile: What we learned from 250,000 gamers. In *CHIPlay'16*, 2–2. ACM.

Zhang, X.; Brown, H.-F.; and Shankar, A. 2016. Data-driven personas: Constructing archetypal users with clickstreams and user telemetry. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, 5350–5359. ACM.

Zook, A. E., and Riedl, M. O. 2012. A temporal data-driven player model for dynamic difficulty adjustment. In *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Zook, A.; Lee-Urban, S.; Riedl, M. O.; Holden, H. K.; Sottilare, R. A.; and Brawner, K. W. 2012. Automated scenario generation: toward tailored and optimized military training in virtual environments. In *Proceedings of the international conference on the foundations of digital games*, 164–171. ACM.