Balancing Video Games: A Player-Driven Instrument

Johannes Pfau
j.pfau@uu.nl
Utrecht University
Utrecht, Netherlands

Magy Seif El-Nasr
mseifeln@ucsc.edu
University of California, Santa Cruz
Santa Cruz, California, USA

« the DPS output of a class should be a function of rotational complexity, boons it can give, self-sustain/squishiness, range/safety, CC contribution, and so on »

– Participant 540

ABSTRACT

Video game balancing is a controversial and highly debated topic, especially between players of online games. Whether a game is sufficiently balanced greatly influences its reception, satisfaction, churn rates and success. In particular, different ideologies of balance can lead to worse player experiences than actual imbalances. This work succeeds a fine-grained investigation about the attitudes of the *Guild Wars 2* community regarding balancing factors, and introduces a player-driven quantitative tool to approximate closer configurations of balance that could optimize player experience and satisfaction. After an initial evaluation, theoretical constellation outputs of this tool improved players' perception of the balance between in-game build options – where aggregated opinions of ($n = 64$) players even showed benefits over individual opinions, indicating a potential "wisdom of the crowd" effect.

CCS CONCEPTS

• **Information systems** → *Massively multiplayer online games*; • **Human-centered computing** → *User centered design.*

KEYWORDS

Video game balancing, player-centric development

ACM Reference Format:
Johannes Pfau and Magy Seif El-Nasr. 2023. Balancing Video Games: A Player-Driven Instrument. In *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY Companion '23), October 10–13, 2023, Stratford, ON, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3573382.3616097

1 INTRODUCTION

Along with narrative, content and game aesthetics, balancing video games is one of the biggest factors of game design that shapes player experience, and, especially in the context of online multi-player games, the arguably most controversial one. Among all of the currently popular genres, games employ never-ending balance patches that most often continue years after the launch of a game (or never cease to exist). These can operate on different layers of in-game choices, such as the playable Champions[1] of the Multiplayer Online Battle Arena (MOBA) *League of Legends* [14], equippable weapons[2] of the First-Person Shooter (FPS) *Counter-Strike: Global Offensive* [36] or active and passive changes to classes, traits and skills[3] of the Massively multiplayer online role-playing game (MMORPG) *Guild Wars 2* [3]. One of the leading factors that exacerbate proper configurations of balance is the lack of a proper definition, as stressed by Becker and Görlich [5]. In their large-scale review of academic and industrial viewpoints, they conclude that even across fields, authors and practitioners are unable to agree on a shared understanding of game balancing. This thus entails that adequate balancing that maximizes player experience should be tailored to the intricacies and idiosyncrasies of the particular game - which is not only constrained to the game's genre, mechanics, dynamics and developers' intentions, but largely influenced by the actual player community. In preceding work on the balance of *Guild Wars 2* [26], we captured such fine-grained requirements of its player base, following Becker and Görlich's overarching categories of **difficulty**, **symmetry**, **viability** and **fairness** [5]. In particular, this work continues the extensively discussed approach of player-centric balancing by the implementation and evaluation of a community-driven balancing tool that captures player requirements in an adaptive and quantifiable manner and computes balance constellations based on these. Thus, this paper approaches to answer the following research question:

- **(RQ1)**: How can subjective understandings of balancing be translated into parameter constellations that realize players' attitudes?
- **(RQ2)**: Does directly quantifying players' opinions on a game's balance increase their perception of a well-balanced game?
- **(RQ3)**: Does the aggregation of community attitudes lead to a beneficial "wisdom of the crowd" effect that enhances balancing, or does integrating a wider diversity of opinions lead to controversial outcomes?

By developing an artifact that translates granular balance factors (as per previously collected dimensions) into parameter constellations, and evaluating if this player-driven approach leads to a more satisfactory balance outcome for a subset of the *Guild Wars 2* community, we contribute to the fields of games user research (investigating detailed player needs), we advance measures (and

This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

CHI PLAY Companion '23, October 10–13, 2023, Stratford, ON, Canada
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0029-3/23/10.
https://doi.org/10.1145/3573382.3616097

[1] https://www.leagueoflegends.com/en-us/news/tags/patch-notes/
[2] https://liquipedia.net/counterstrike/Patches
[3] https://en-forum.guildwars2.com/forum/6-game-update-notes/

interventions) for game evaluation (specifically balancing) and provide further evidence that player-centric development can elevate game experience.

2 RELATED WORK

For other application cases, player-driven approaches have already delivered promising solutions that exceeded the capabilities of complete in-house implementations. One example would be Da Silva et al.'s utilization of the collective power of a community to retrieve a widely nested narrative and story background merely from player input [9]. Concerning procedural content generation, Shaker et al. highlighted the capabilities of player-centric approaches, including personalization of in-game maps or experiences [33]. Partlan et al. utilized participatory design in order to assess requirements and develop design-driven features for co-creative game AI design tools [25]. Even completely player-driven game development cycles have been shown to result in novel experiences, dynamic design procedures and central game features that are inherently tailored to the actual target audience [18]. Canossa and Drachen argue that increasing the players' agency and influence on the development process can lead to enhanced experiences and immersion when introducing play-personas for customized gameplay [8] – which arguably extends similarly to continual balancing updates. Eventually, player-driven paradigms might even scale to the large magnitudes of communities that popular modern games accumulate, as Ma et al. indicate in their work on user innovation evaluation strategies, incorporating over 21,000 players that produced novel and sufficiently complex ideas and suggestions [21].

However, purely player-driven balancing decisions might still be warped by diverging opinions, missing empirical knowledge about the actual state of the game and the gap of experience (and requirements) between novice and expert players. The arguably most objective measures to counter lacks of knowledge and to condense how games actually play out are empirical data-driven methods [13, 37]. The majority of these data-driven approaches are stemming from and/or focusing on delivering insights for the game industry (e.g. data mining, classification or prediction) [10], targeting measures against churn [15], facilitating content generation [31], or easing the burden of testing [2] (among other areas). Most of the remaining approaches follow academic interests concerning similar topics or fundamental (psychological or technical) regularities, structures and concepts [11, 41], often employing visualizations to gather insights for researchers or analysts [7]. Prominent implementations target spatio-temporal movement [1, 23, 38, 39], decision making of individual or aggregate players [19, 24] or higher-level metrics and statistics [12]. Effectively, certain academic approaches already addressed the balancing of viable game options, such as automatic symmetric and intransitive player modeling approaches from Pfau et al. [27, 28], asymmetric Monte-Carlo balancing from Beau and Bakkes [4], Jaffe et al.'s maximization of fair and useful card game cards [16] or Leigh et al.'s reduction of dominant strategies through coevolution [17]. Even if most of these draw on simulation or calculation towards well-balanced game states and (to the best of our knowledge) no academic work included the players' perspective so far, it is reasonable to hypothesize that similar balancing solutions can follow or implement opinionative inputs.

Ultimately, we want to empower and harness *player-driven* balancing conceptions in accordance with *data-driven* methods. Making video game data transparent, explainable and applicable to its players is already one of the driving topics within the areas of game-related explainable AI and player modeling [20, 40, 43], and comparably holds in the context of balancing. For this reason, we developed, published and populated the player- and data-driven analytics tool *Guild Wars 2: Wingman*[29][4]. In the following sections, we briefly introduce the environment of *Guild Wars 2*, before outlining and evaluating the democratic player-driven balancing instrument of this work.

3 GAME ENVIRONMENT

The game used for the subsequent case study (*Guild Wars 2*) is a prototypical MMORPG featuring single- and multiplayer content in storylines, open world events, various PvP modes and endgame encounters such as raid bosses, fractals or strike missions. The latter make for a large share of players' time spent in game and as they are set in fixed scenarios with only few probabilistic factors and established group compositions, strategies and roles, they enable very comparable benchmarks [22, 35]. Based on these, large communities formed around discussions and optimizations to overcome these challenges from which balancing discrepancies can become apparent quickly. As *Guild Wars 2* is specifically designed to not feature power creep mechanics such as increasing level caps or item qualities over time, the vast majority of players of this endgame content participates on an identical or very similar character and equipment attribute level, and performance is mainly influenced by in-game proficiency, mastery of the classes, encounter knowledge, strategy and group composition, which further adds to the comparability of this approach with regards to balancing. Still, when accepting certain degrees of noise or acknowledging these factors of variance by clustering players into equipment tiers or approximately subtracting out these confounding variables, the balancing procedures as presented here are likely to produce similarly powerful insights for instanced (group) PvE content in general, such as raids, dungeons, trials or ultimate encounters in *World of Warcraft*[6], *The Elder Scrolls Online*[42] or *Final Fantasy XIV*[34]. *Guild Wars 2* features a variety of playable class options (from now on referred to as *professions*), where each of the nine core professions can be expanded by particular specializations that can add further capabilities or open up new roles for this profession. In theory, the professions of *Guild Wars 2* allow abundant combinations of character constellations, such as individually distributed equipment attributes, chosen passive character traits and active weapon type and skill choices. In that way, players can represent and play out different roles within their party, such as dealing direct damage, damage over time, offering support or different degrees of mixtures of these factors. However, for the sake of optimization and role compression within the group, most of the time, these builds are min-maxed towards the roles of full support (heal and buff application), offensive support (dps and buff application) or maximal dps (either direct damage or damage over time). The majority of players uses builds and equipment that maximizes their functionality in one of these roles with only situational variation, following

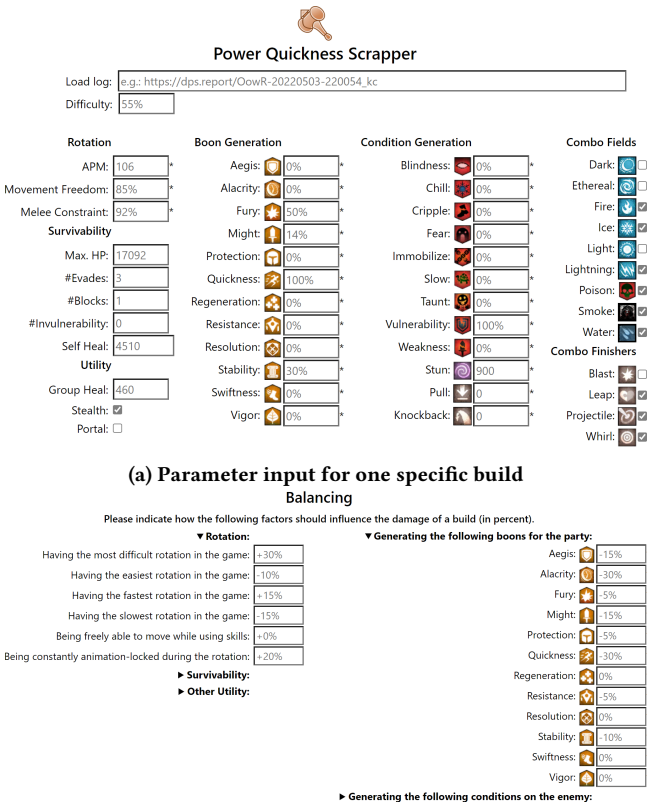
[4] https://gw2wingman.nevermindcreations.de/

community guides and recommendations. Thus, these builds make up the basis of elements to balance for our further purposes. All professions are (in theory) capable of fulfilling all of these roles, yet their viability and efficiently greatly diverge (and differ with respect to the combated encounters), which constantly raises balancing gaps between builds. For further terminology, the appendix lists explanations for all game-specific terms used across this work.

4 APPROACH

The immediate predecessor of this work collected requirements, facets and dimensions on how balancing can and should be carried out from the players' perspective: Players desire that more **difficult** builds should be rewarded by higher damage output (and, respectively, subpar performances should be punished with lower output), all builds should be similarly **viable** (and not consistently dominated by other options) and players should have a **fair** chance of achieving success with their build of choice, however, completely **symmetric** constellations of outcomes across builds are strongly undesired [26]. These requirements manifest in even more detailed, game-specific questions that can best be represented by the statement of one participant, reflecting the larger picture of the community's mindset: *"the DPS output of a class should be a function of rotational complexity, boons it can give, self-sustain/squishiness, range/safety, CC contribution, and so on"* (P540) [26]. In this context, a skill *rotation* is defined as the ideal sequence of actions that maximizes damage output, which can highly differ in complexity, length, speed and number of sub-rotations (loops) between builds [30, 35]. Based on these requirements, we present a democratic player-driven instrument that quantifies and implements desired balancing visions through interactive parameter tuning.

In this tool, presented on the web-based platform of *Guild Wars 2: Wingman* [29], users can give their input on a) existing in-game builds, their perceived difficulty and detailed parameters about their expected damage uptime and utility contribution (cf. Figure **??**) and b) their own (quantified) requirements on balancing endgame PvE content in general (cf. Figure **??**). In order to keep objective measures as accurate as possible, players can import logs for ideal rotations into the description of a build, which auto-completes star-marked fields such as actions per minute (apm), the relative amount of time this rotation is able to move while executing skills (as compared to being animation-locked), the relative amount of time it has to operate on melee range to maintain damage uptime and boon and condition generation. The most critical input is however the perceived difficulty of playing this build (as a percentage of the highest complexity imaginable). More importantly, users are asked to quantify their perception on how these factors should (positively or negatively) affect performance outcomes in the general view. Eventually, this tool interpolates the vector of parameters for each build (or a selection of these) between the specified balancing variables, calculates appropriate distributions of performances (median and variance), and produces a visualization about these distributions across the chosen builds (cf. Figure 2). With reference to this outcome, the example configuration of the *Power Quickness Scrapper* is allocated rather low expected dps performances on the median with minor variance due to its easier complexity, higher damage uptime and valuable contribution of the *Quickness*



(a) Parameter input for one specific build

(b) Parameter input for the individual stance on balancing

Figure 1: Input masks for (a) in-game complexity, speed, constraints, survivability and utility of particular builds and (b) how these parameters should impact performance balancing in general (excerpt).

boon to the party. This utility contribution is yet included in the visualization, proportional to the theoretically added group dps. This is even more emphasized in the *Heal Druid* in this case which completely focuses on supplying worthwhile boons and ensuring the survival of the whole group. In contrast, easier damage builds with slim to no utility like the *Power Berserker* are situated in the middle ground, while the very difficult *Power Catalyst* displays a large variance up to top performances and the difficult and only situationally efficient *Condition Mirage* displays a comparably large variance stemming from being weak at certain encounters, but very strong on others while even supplying some boons.

Users can modify their parameter constellation(s) at will, while having immediate updates in the interactive plot, and submit their perception to the pool of estimations once satisfied. The accompanying visualization shows the performance comparison based on own input by default, but can be switched to display the aggregated opinions of the entire participating community.

Along with the illustration of Figure 2 that directly incorporates **difficulty** such as rotation speed (apm) and perceived complexity as determining variance, **symmetry** and **viability** assessments are automatically calculated and displayed - by computing the mean

(B)	I am satisfied with this state of balancing.
(D)	Differently difficult builds produce different performance outcomes in this configuration.
(V)	This configuration matches my idea of how viable professions should be.
(S)	The performances of professions in this configuration are diverse.
(F)	Independent of their profession choice, players have a fair chance of achieving success within this configuration.

Table 1: Items of the quantitative evaluation form accompanying the instrument. All statements could be answered on 7-point Likert scales from "Strongly Disagree" to "Strongly Agree" and were presented after users a) customized their *own* configuration, b) were consulted with the whole *community's* aggregated configuration and c) were displayed empirical distributions from the June patch.

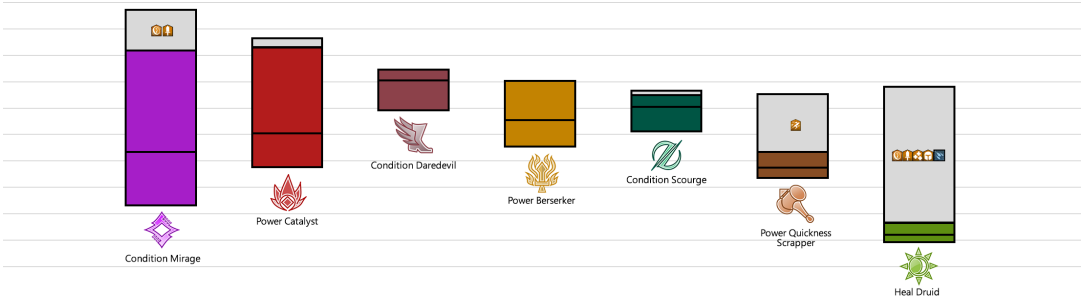


Figure 2: Balance visualization of the configured build and parameters in Figure 1, compared to theoretical performance distributions of alternative builds according to the same balance configuration. Colored boxplots indicate the damage output variance, where grey adds an estimate utility contribution.

squared error to perfect symmetry, and by prompting when certain builds are threatened to be *dominated* by another option (including damage as well as utility measures). Apart from this theoretical constellation, users are capable of comparing these proportions (directly highlighting differences) to the actual data-driven performances recorded by the underlying analytics platform at only the press of a button - using either optimal (training area) or empirical (actual boss) logs. This then does not only merge the customized class configuration with the empirical distributions (of *any* chosen patch), but also lists the approximate dps changes per build that would have to happen in order to realize the desired balance.

5 EVALUATION

To finally quantify whether the instrument actually satisfies the needs and opinions of the community, we evaluated the players' feedback over a two-weeks period. In theory, users can compare a) their own produced balance configuration to b) the aggregated constellation of all user inputs and c) empirical distributions of all patches for which we recorded data. For the sake of this paper, we only let them compare the two outputs to the patch discussed in previous work (June 28, 2022)[5]. Users of the tool were asked to give answers to five compact items for each configuration (cf. Table 1). On top of this subjective assessment, we quantified the balancing concepts, i.e. **difficulty** by the variance of a profession's dps distribution and **viability** by the number of professions that strictly dominated a build.

[5]https://wiki.guildwars2.com/wiki/Game_updates/2022-06-28

6 RESULTS

In total, ($n = 64$) users submitted a complete configuration of their balance requirements and responded to the following quantitative evaluation in this timeframe, which we investigated using Welch's t-tests. Figure 3 contrasts the balance perceptions of their own, the community outcome, and the June patch. Unsurprisingly, users were more satisfied with their own customized balance configuration ($M = 6.4, SD = 0.52$) than with the actual patch ($M = 1.86, SD = 0.82; p < 0.05$). Yet, in comparison to what the global community produced ($M = 6, SD = 0.88$), players did not find this significantly less convincing, which indicates that a larger set of opinions can still converge to a satisfying solution. When it comes to proper **difficulty**, own adjustments ($M = 4.9, SD = 0.64$) produced better fits than in the June patch ($M = 4.07, SD = 0.81; p < 0.05$), but the combined competence of the community resulted in even higher scores ($M = 5.57, SD = 0.09; p < 0.05$). This could imply a potential wisdom-of-the-crowd effect that appeared because more and more understanding of the particular classes were added. In previous work, players criticized that the June patch reduced the **viability** of the profession pool to a small set, thus we also observed to a significant drop in **viability** ($M = 2.47, SD = 0.95$) than for own ($M = 4.55, SD = 0.5$) or aggregated ($M = 4.04, SD = 0.81$) constellations. Yet, we did not find any significant differences between **symmetry** perceptions, presumably because the performances across the multitude of professions were not too symmetrical, but rather diverse even in the actual patch ($M = 5.41, SD = 0.91$). Finally, players deemed the **fairness** of achieving success to be higher for both own and community outcomes ($M = 5.59, SD = 0.5; M = 5.35, SD = 0.93$) when compared to the in-game patch

($M = 3.88, SD = 0.99; p < 0.05$). Beyond subjective impressions, we also calculated objective metrics following our preceding requirement study [26]. When computing the correlation between the subjective difficulty of a build and its dps variance, both single-player configurations ($r = 0.92$) as well as these from the whole community ($r = 0.89$) produced large associations in contrast to the live patch ($r = 0.34$). While this is partly because of the design of the calculation that is based on the players' demands, it still suggests that in live patches, proficiency is often not the decisive factor for the dps distribution of a class, which yet was one of the distinctly revealed community requirements. Regarding **viability**, each of the builds was dominating (or dominated by) ($M = 5.83, SD = 4.8$) other builds in the empirical data of the June patch, but only by ($M = 3.66, SD = 3.88$) in own configurations and even less ($M = 3.05, SD = 3.52$) in the community solution. Even if this measure would not eradicate dominating relations between builds, it comes closer to the desired notion of balancing.

7 DISCUSSION

This work outlines how game-specific balance concepts can be quantified a) per element to balance (in this case, builds) and b) with respect to their global impact on the performance outcomes. While the outputs of this tool claim to support the lower-level factors of **difficulty**, **symmetry** and **viability** measures numerically, they do not necessarily imply a subjective improvement over a balanced state. Thus, we eventually evaluated the player assessments of a) customized own configurations and b) the aggregated community solution against c) the empirical distributions of an actually published balance patch. Based on the presented methodology and results, we argue that player-driven balancing constellations can be achieved by unraveling the abstract concept of balancing into dimensions that are important for the community of the game, assessing the players' importance of these factors, and adapting the (damage and utility) outcome of available in-game builds towards constellations that minimize deviations from these requirements (RQ1). As hypothesized, such personalization of a balance state increases the satisfaction of individual players when applying their own perception of balance (RQ2), and requirements on **difficulty**, **viability** and **fairness** are fulfilled significantly better. Astonishingly enough, aggregated perceptions of a larger amount of players ($n = 64$) did not result in completely controversial or contradicting perceptions, as satisfaction, **viability**, **symmetry** and **fairness** scores were not significantly different from direct individual configurations, but perceptions of balanced **difficulty** even increased, which indicates that a potential "wisdom of the crowds" effect can emerge when including voices from a larger player base (RQ3). These outcomes suggests that empowering a player community to influence the state of balance with their own understandings and adjustments can lead to higher satisfaction and closer approximations of at least **difficulty**, **viability** and **fairness** requirements across playable options. This plays into Robert et al.'s findings on the motivations of player communities to aid game development (or even co-create) [32]: not primarily to create a more enjoyable game for themselves, but for the greater good of the very game.

8 LIMITATIONS & FUTURE WORK

The presented work undergoes a number of limitations, partly caused by the disparity of the conceptions and partly because of realistic restrictions within the methodology. Finding balanced configurations in a popular world-class video game that would suit requirements of hundreds of thousands to millions of active players without disappointing anyone is arguably impossible. To this respect, this initial exploratory evaluation of the tool could only reach a comparably minor part of the entire audience of players. Above that, those players interested in balancing (and in expressing their opinion) were more likely to respond to this study which might have skewed the results. Especially regarding RQ3, a larger-scaled evaluation is necessary see if incorporating player voices from all levels of proficiency also results in a "wisdom of the crowds" effect. Developers are moreover responsible to push novelty, enjoyment and shifts out of rigidly stuck constellations, in order to keep their game innovative, interesting and economically competitive. We neglected factors of flair, intrinsic motivation of play styles and further variables not inherently related to performance for now, but acknowledge that player choices and preferences are not completely rational (with respect to efficiency and strategy optimization). In order to showcase the applicability of this endeavor and to give a hands-on example that player communities are interested in and capable of being incorporated in balancing, we utilized *Guild Wars 2* as a fitting, popular and contemporary use case. The underlying balance concepts and the aggregation of a community's performance requirements however are arguably generalizable and similarly applicable for similar genres, single-player games or competitive PvP settings – which yields great potential for player-driven balancing. For future work, we mainly seek to extend and unravel the nuances of play styles, builds and implications for balancing. These will be incorporated in the presented tool to portray ideal perceptions of balance versus current empirical constellations down to the lowest possible detail. Even though we evaluated this tool in terms of balance perceptions regarding theoretical aggregated configurations, the implementation of such produced balancing decisions into actual gameplay and evaluation of the subsequent player experiences is an important open endeavor. In pursuing this path, we strive to tighten the connection between players and developers – as well as between industry and academia.

9 CONCLUSION

The balancing of in-game options is an interminable, controversial and considerable process for many genres, developers and players alike. Balancing adjustments are one of the major causes of update patches for online games, display never-ending experience optimization problems and highly impact how players play a game altogether. Scientific efforts to study balance are divided into several understandings of the term, even partly conflicting. When it comes to the definition of adjusting for equated or appropriate in-game options, related work is largely under-investigated and mostly regards simulation- or computation-based approaches to even out viability across choices. The role, perception and requirements of the player or even the actual player community have not been considered in academia so far, despite bearing considerable implications for

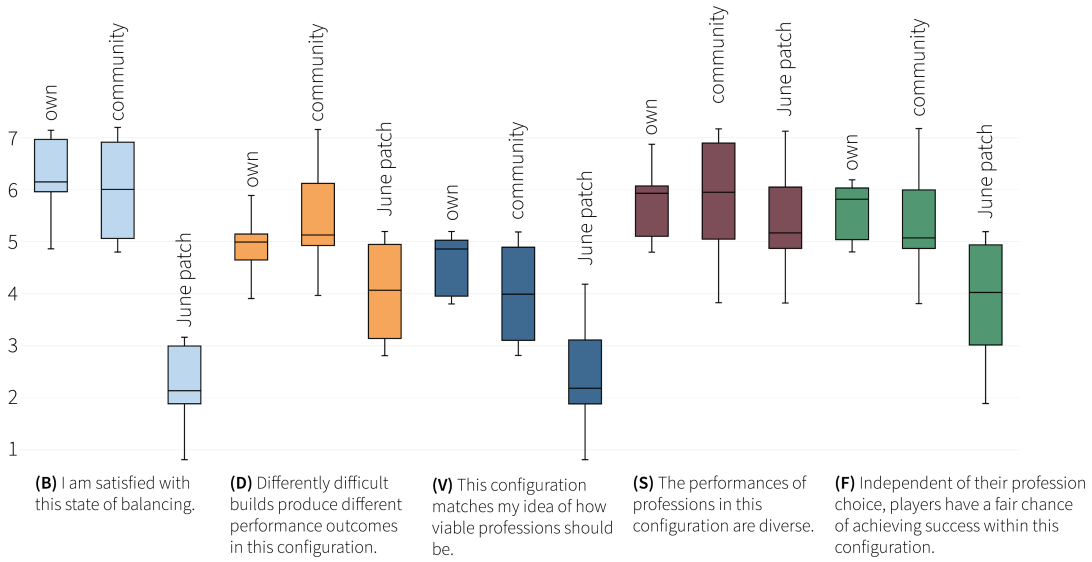


Figure 3: Distribution of responses regarding satisfaction, viability, symmetry, fairness and difficulty between the tool's outcomes of *own* constellations, the aggregation of all *community* members' constellations, and the *June 28 patch*. Answers to Table 1 ranged from 1 ("Strongly Disagree") to 7 ("Strongly Agree"), boxplots indicate means (–), standard deviations (boxes) and range (whiskers).

games user research, game design and human-computer interaction in general. For these reasons, we build on aggregated balance notions of the MMORPG *Guild Wars 2* (as a popular representative of online games undergoing constant rebalancing) and paved the way for finding community agreements on balancing through an interactive democratic tool which is likely to hold for comparable games and larger concepts of balancing.

ACKNOWLEDGMENTS

We thank all participants of the evaluation for their contribution on advancing player-driven balancing. Guild Wars 2 and all associated logos, designs, and composite marks are trademarks or registered trademarks of NCSOFT Corporation or ArenaNet, LLC, respectively. ©2021 NCSOFT Corporation, ©2021 ArenaNet, LLC. All rights reserved.

REFERENCES

[1] Sabbir Ahmad, Andy Bryant, Erica Kleinman, Zhaoqing Teng, Truong-Huy D Nguyen, and Magy Seif El-Nasr. 2019. Modeling individual and team behavior through spatio-temporal analysis. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 601–612.

[2] Aghyad Mohammad Albaghajati and Moataz Aly Kamaleldin Ahmed. 2020. Video game automated testing approaches: An assessment framework. *IEEE Transactions on Games* (2020).

[3] ArenaNet. 2012. *Guild Wars 2*. Game [PC]. ArenaNet, Bellevue, Washington, USA..

[4] Philipp Beau and Sander Bakkes. 2016. Automated game balancing of asymmetric video games. In *2016 IEEE conference on computational intelligence and games (CIG)*. IEEE, 1–8.

[5] Alexander Becker and Daniel Görlich. 2020. What is Game Balancing? - An Examination of Concepts. *ParadigmPlus* 1, 1 (2020), 22–41.

[6] Blizzard Entertainment. 2004. *World of Warcraft*. Game [PC]. Blizzard Entertainment, Irvine, California, USA..

[7] Brian Bowman, Niklas Elmqvist, and TJ Jankun-Kelly. 2012. Toward visualization for games: Theory, design space, and patterns. *IEEE transactions on visualization and computer graphics* 18, 11 (2012), 1956–1968.

[8] Alessandro Canossa and Anders Drachen. 2009. Patterns of Play: Play-Personas in User-Centred Game Development.. In *DiGRA Conference*.

[9] Sylker Teles Da Silva and Kiyoshi Tomimatsu. 2013. Game prototyping with community-driven narrative: Actor-network theory applied for Massively Multiplayer Online Games development. In *2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE)*. IEEE, 376–378.

[10] Anders Drachen and Alessandro Canossa. 2009. Towards gameplay analysis via gameplay metrics. In *Proceedings of the 13th international MindTrek conference: Everyday life in the ubiquitous era*. 202–209.

[11] Anders Drachen, Pejman Mirza-Babaei, and Lennart E Nacke. 2018. *Games user research*. Oxford University Press.

[12] Anders Drachen, Rafet Sifa, Christian Bauckhage, and Christian Thurau. 2012. Guns, swords and data: Clustering of player behavior in computer games in the wild. In *2012 IEEE conference on Computational Intelligence and Games (CIG)*. IEEE, 163–170.

[13] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. 2013. *Game analytics*. Springer.

[14] Riot Games. 2009. *League of Legends*. Game [PC]. Riot Games, Los Angeles, California, USA..

[15] Fabian Hadiji, Rafet Sifa, Anders Drachen, Christian Thurau, Kristian Kersting, and Christian Bauckhage. 2014. Predicting player churn in the wild. In *2014 IEEE Conference on Computational Intelligence and Games*. Ieee, 1–8.

[16] Alexander Jaffe, Alex Miller, Erik Andersen, Yun-En Liu, Anna Karlin, and Zoran Popovic. 2012. Evaluating competitive game balance with restricted play. In *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference*.

[17] Ryan Leigh, Justin Schonfeld, and Sushil J Louis. 2008. Using coevolution to understand and validate game balance in continuous games. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation*. 1563–1570.

[18] Pascal Lessel, Maximilian Altmeyer, and Nicolas Brauner. 2019. Crowdjump: Investigating a Player-Driven Platform Game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 149–159.

[19] Christian Sebastian Loh, I-Hung Li, and Yanyan Sheng. 2016. Comparison of similarity measures to differentiate players' actions and decision-making profiles in serious games analytics. *Computers in Human Behavior* 64 (2016), 562–574.

[20] Crisrael Lucero, Christianne Izumigawa, Kurt Frederiksen, Lena Nans, Rebecca Iden, and Douglas S Lange. 2020. Human-Autonomy Teaming and Explainable AI Capabilities in RTS Games. In *International Conference on Human-Computer Interaction*. Springer, 161–171.

[21] Jifeng Ma, Yaobin Lu, and Sumeet Gupta. 2019. User innovation evaluation: Empirical evidence from an online game community. *Decision Support Systems* 117 (2019), 113–123. https://doi.org/10.1016/j.dss.2018.11.003

[22] Sai Siddartha Maram, Johannes Pfau, Jennifer Villareale, Zhaoqing Teng, Jichen Zhu, and Magy Seif El-Nasr. 2023. Mining Player Behavior Patterns from Domain-Based Spatial Abstraction in Games. In *2023 IEEE Conference on Games (CoG)*. IEEE, 1–8.

[23] Dinara Moura, Magy Seif El-Nasr, and Christopher D Shaw. 2011. Visualizing and understanding players' behavior in video games: discovering patterns and supporting aggregation and comparison. In *ACM SIGGRAPH 2011 game papers*. 1–6.

[24] Truong-Huy D Nguyen, Magy Seif El-Nasr, and Alessandro Canossa. 2015. Glyph: Visualization Tool for Understanding Problem Solving Strategies in Puzzle Games. In *Proceedings of the 10th International Conference on the Foundations of Digital Games (FDG 2015)*. Foundations of Digital Games 2015, FDG 2015 ; Conference date: 22-06-2015 Through 22-06-2015.

[25] Nathan Partlan, Erica Kleinman, Jim Howe, Sabbir Ahmad, Stacy Marsella, and Magy Seif El-Nasr. 2021. Design-Driven Requirements for Computationally Co-Creative Game AI Design Tools. In *The 16th International Conference on the Foundations of Digital Games (FDG) 2021*. 1–12.

[26] Johannes Pfau and Magy Seif El-Nasr. 2023. On Video Game Balancing: Joining Player- and Data-Driven Analytics. arXiv:2308.07576 [cs.HC]

[27] Johannes Pfau, Antonios Liapis, Georg Volkmar, Georgios N Yannakakis, and Rainer Malaka. 2020. Dungeons & replicants: automated game balancing via deep player behavior modeling. In *2020 IEEE Conference on Games (CoG)*. IEEE, 431–438.

[28] Johannes Pfau, Antonios Liapis, Georgios N Yannakakis, and Rainer Malaka. 2022. Dungeons & Replicants II: Automated Game Balancing Across Multiple Difficulty Dimensions via Deep Player Behavior Modeling. *IEEE Transactions on Games* (2022).

[29] Johannes Pfau and Magy Seif El-Nasr. 2023. Player-Driven Game Analytics: The Case of Guild Wars 2. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 728, 14 pages. https://doi.org/10.1145/3544548.3581404

[30] Johannes Pfau, Jan David Smeddinck, and Rainer Malaka. 2020. Enemy within: Long-term motivation effects of deep player behavior models for dynamic difficulty adjustment. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–10.

[31] Sebastian Risi and Julian Togelius. 2019. Procedural content generation: from automatically generating game levels to increasing generality in machine learning. *arXiv preprint arXiv:1911.13071* (2019).

[32] Deborah Roberts, Mathew Hughes, and Kia Kertbo. 2014. Exploring consumers' motivations to engage in innovation through co-creation activities. *European Journal of Marketing* (2014).

[33] Noor Shaker, Georgios N Yannakakis, and Julian Togelius. 2012. Towards player-driven procedural content generation. In *Proceedings of the 9th conference on Computing Frontiers*. 237–240.

[34] Square Enix. 2013. *Final Fantasy XIV*. Game [PC,OSX,PS3]. Square Enix, Tokyo, Japan..

[35] Zhaoqing Teng, Johannes Pfau, Sai Siddartha Maram, and Magy Seif El-Nasr. 2022. Player Segmentation with INSPECT: Revealing Systematic Behavior Differences within MMORPG and Educational Game Case Studies. In *Extended Abstracts of the 2022 Annual Symposium on Computer-Human Interaction in Play*. 87–92.

[36] Valve. 2012. *Counter-Strike: Global Offensive*. Game [PC,PS,XBox]. Valve, Bellevue, Washington, USA..

[37] Günter Wallner. 2019. A brief overview of data mining and analytics in games. *Data analytics applications in gaming and entertainment* (2019), 1–14.

[38] Günter Wallner, Nour Halabi, and Pejman Mirza-Babaei. 2019. Aggregated visualization of playtesting data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[39] Günter Wallner and Simone Kriglstein. 2012. A spatiotemporal visualization approach for the analysis of gameplay data. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1115–1124.

[40] Lindsay Wells and Tomasz Bednarz. 2021. Explainable ai and reinforcement learning—a systematic review of current approaches and trends. *Frontiers in artificial intelligence* 4 (2021), 550030.

[41] Nick Yee. 2006. The psychology of massively multi-user online role-playing games: Motivations, emotional investment, relationships and problematic usage. In *Avatars at work and play*. Springer, 187–207.

[42] ZeniMax Online Studios. 2014. *The Elder Scrolls Online*. Game [PC,OSX,PS4,XboxOne]. ZeniMax Online Studios,Hunt Valley, Maryland, USA..

[43] Jichen Zhu and Magy Seif El-Nasr. 2021. Open Player Modeling: Empowering Players through Data Transparency. In *Proceedings of AIIDE Workshop on Experimental AI in Games (ExAG '21)*.

APPENDIX

A GAME TERMINOLOGY

Table 2: Terms especially used in *Guild Wars 2*, the genre of MMORPGs or their analytics

atomic	*(as in atomic actions)*: Logs of the game utilized in the data-driven evaluation part are recorded on the lowest-level possible, i.e. down to *skill* usage and character movement on a frame-by-frame logging basis.
boons	Temporary positive effects that increase character *stats* or yield *utility*, most often provided by *support builds*.
buff	Increase in damage, *utility* or general viability of a *profession* or *build* caused by balancing adjustments. Opposed to *nerf*.
build	The customizable configuration of equipment, *skills* and *traits* of a *profession*. Most builds target the optimization of *power* or *condition* damage output, maximizing *utility* or hybrid versions of these.
class	The overarching archetype for each character, referred to as *profession* in *Guild Wars 2*.
condition	Temporary negative effects that deal damage over time on an enemy or weaken their *stats*. Some *builds* are optimized to deal condition damage in contrast to direct *power* damage.
damage uptime	The ability of a *build* (or player) to consistently deliver damage, mainly influenced by factors such as survivability, range of *skills*, freedom of movement while executing *skills*, adaptability of the ideal *rotation* to live combat, and dependence on other factors such as the size of the enemy's hitbox, attack delay or movement patterns.
dps	*(damage per second)*: The theoretical or empirical damage a *build* or *player* afflicts onto their target(s).
endgame	In *Guild Wars 2*, *PvE* endgame content is mainly carried out in instanced dungeons for five players *(fractals of the mists)* or ten players *(raids and strike missions)*. As it features no power creep or item spiral and most players follow *builds* and *rotations* from community guides, combat logs for single bosses are highly comparable between groups and players and differences in efficiency are mainly attributed to the proficiency of players.
log	For the platform used in this work, single bosses or encounters are recorded in *atomic* detail, representing the full combat replay and *dps*, heal, *boon*, *condition* among other statistics at every single point in time for up to ten players [29].
nerf	Decrease in damage, *utility* or general viability of a *profession* or *build* caused by balancing adjustments. Opposed to *buff*.
performance	The quantified outcome of a player at a given situation, e.g. for one boss fight. Mostly expressed as *dps* values.
power	*(as in power damage)*: Direct damage as opposed to *condition* damage (damage over time). Some *builds* are optimized to deal power damage.
profession	*Guild Wars 2's* notion of character *classes*. It features nine core professions that can be extended with one of three *specializations* each for more *build* diversity.

Received 2023-06-22; accepted 2023-08-03

PvE	*(Player versus Environment)*: *Guild Wars 2* features single- and collaborative multi-player modes. This work focuses on the group-based *endgame* content.
PvP	*(Player versus Player)*: *Guild Wars 2* features small- and large-scale competitive PvP modes, yet to keep the assessment as concise as possible, we focused on balancing *PvE* in this work.
rotation	The sequence of *skills* players execute, often looping for ideal rotations (optimizing *dps*) within a *build*.
skill	*(as in executable skills)*: Single actions players activate to deal damage and/or provide *utility* by pressing the corresponding button.
skill	*(as in player skill)*: The proficiency of a player (on a specific *build* or *profession*), quantifiable in the amounts of *dps* or *utility* they can provide.
stats	In-game character attributes that influence damage, *utility* or survivability potentials of a *build*.
specialization	*Professions* can be added one of three specializations (27 in total) that affect the mechanics, damage and/or *utility* potentials of a *build*.
support	As opposed to *power* or *condition* damage *builds*, support roles mainly provide *utility*.
trait	Customizable passive perks of a *build* that mostly increase damage or *utility* potentials, as opposed to active customizations (*skills*).
utility	Beneficial value a *build* can provide for itself and/or other players in the group (apart from *dps*), such as *heal*, *buffs*, movement *skills*, *conditions*, crowd-control or the ability to resurrect fallen players.