Improving Player Balancing in Racing Games

Jared Cechanowicz[1], Carl Gutwin[1], Scott Bateman[2], Regan L. Mandryk[1], and Ian Stavness[1]

[1]University of Saskatchewan
Saskatoon, Canada, S7N 5C9
firstname.lastname@usask.ca

[2]University of Prince Edward Island
Charlottetown, Canada, C1A 4P3
sbateman@upei.ca

ABSTRACT

In competitive games where players' skill levels are mismatched, the play experience can be unsatisfying for both stronger and weaker players. *Player balancing* provides assistance for less-skilled players in order to make games more competitive and engaging. Although player balancing can be seen in many real-world games, there is little work on the design and effectiveness of these techniques outside of shooting games. In this paper we provide new knowledge about player balancing in the popular and competitive racing genre. We studied issues of noticeability and balancing effectiveness in a prototype racing game, and tested the effects of several balancing techniques on performance and play experience. The techniques significantly improved the balance of player performance, were preferred by both experts and novices, increased novices' feelings of competitiveness, and did not detract from experts' experience. Our results provide new understanding of the design and use of player balancing for racing games, and provide novel techniques that can also be applied to other genres.

Author Keywords

Player balancing; driving games; dynamic balancing.

INTRODUCTION

There are many gameplay situations in which players' skill levels are mismatched. For example, when parents play with their children, or when friends play socially, one person can be considerably better at the game than others. In a competitive game, the mismatch means that play experience can be much less satisfying – for both the stronger and the weaker player. For skilled players, the game is uninteresting because there is little challenge; for weaker players, the game is frustrating because there is little chance of winning.

Player balancing seeks to address this problem by adjusting game parameters to boost performance of weaker players. Player balancing is different from typical game balancing in that it does not attempt to equalize the opportunities provided to each player (e.g., by giving players capabilities of equal strength); rather, the motivation behind player balancing is to provide a more engaging and fun experience for all players. In fact, player balancing works by *un*balancing the game in the traditional sense, in that it gives weaker players greater capabilities or resources.

Forms of player balancing have been seen in real-world games (e.g., handicaps in golf or "head starts" in running games) and in some video games (e.g., the "Bullet Bill" powerup in MarioKart, which allows players at the back of the race to catch up, or the "Fatboy" mod in Unreal Tournament, which makes successful players wider and easier to hit). These balancing techniques are effective, but are also often highly obvious – which may cause all players to feel that the game is unfair, reducing players' engagement.

Previous research has also examined player balancing, but only in a limited number of situations (e.g., [5,11,16,20]). For example, recent studies have shown that player balancing can be effectively applied in 2D and 3D shooting games [5,20]. However, this previous work also showed that applying these techniques in games can be difficult because there are many factors that can reduce the effectiveness of a balancing technique in a particular game or game genre. For example, applying balancing techniques in a more complex 3D game [20] proved to be more difficult than in a simple 2D game [5], because of game elements such as choice of weapons or targets that shoot back. In addition, player awareness of the assistance can reduce its efficacy [3].

Although this previous work shows the potential for player balancing, there is still relatively little known about how player-balancing techniques can be applied in real games. In particular, there are no studies that look at player balancing in game domains that do not involve targeting – and other genres introduce many factors that may change the way that player balancing should be designed.

In this research, we add to understanding about player balancing by investigating the design and application of balancing techniques in an untested genre – racing games. Driving and racing games are a popular genre that is highly competitive but very unlike the shooting games in which player balancing has previously been studied. Some current racing games already contain balancing techniques (such as Mario Kart's "Bullet Bill" powerup), suggesting that there is a real-world interest in player balancing for racing games. In designing balancing techniques for racing games, we focused on three goals: first, to *reduce noticeability*, since the obviousness of current balancing techniques for racing games can detract from the play experience; second, to *cover a large*

CHI PLAY'14, October 19–21, 2014, Toronto, ON, Canada.

ACM 978-1-4503-3014-5/14/10…$15.00.
http://dx.doi.org/10.1145/2658537.2658701

range, to be able to balance players with substantially different skill levels; and third, to *identify effects on other gameplay elements*, since there could be elements that are important to the play experience other than the main game outcome (wins and losses).

To explore these goals in the context of the racing genre, we carried out a series of studies with a custom-built realistic driving game (based on the Racing Game code provided with the XNA environment [18]). We also built a driver simulation system that allowed us to carry out many initial comparisons of different factors, and to have human participants race against opponents with consistent skill levels. Within our prototype, we developed several *player-balancing techniques* – which are composed of a set of balancing mechanism and an adaptation algorithm. We introduced four innovations in order to meet the design goals:

- *Using multiple mechanisms for balance*. We tested the value of balancing more than one game mechanism simultaneously (steering, speed, and acceleration), both to increase the adaptive capability of the system and to reduce the obviousness of any one manipulation.
- *Using assistance and hindrance*. We examined whether additional balancing power (and reduced noticeability) could be obtained by making the skilled player's game mechanics more difficult, rather than just making the weaker player's mechanics easier. For example, the system can push experts' cars away from the center of the road, at the same time as it attracts novices' cars.
- *Using different techniques for dynamic adaptation*. We developed four dynamic-adaptation techniques that use different approaches for calculating when to re-balance the game (two continuous distance thresholds, one fixed distance threshold, and one time-based sliding window).
- *Using a simulated driver.* We built a feedback-control based simulated opponent to perform automated tests of our adaptation algorithms and to provide consistent opponents in our studies with human participants.

We tested the effects of our balancing techniques both on several performance variables (e.g., wins, lead changes, distance and time between players), and also on game-experience variables that provide a richer view of a player's subjective experience (e.g., feelings of competence and autonomy, and subjective assessment of enjoyment, engagement, fun, and fairness).

Our studies provide six main findings:

- Using multiple balancing mechanisms simultaneously worked best both for increasing the balancing range, and for decreasing the noticeability-to-effect ratio;
- Using hindrance for experts substantially increased balancing range – and in addition, negative adaptations were less noticeable to expert players;
- Our techniques were able to balance several gameplay metrics. Gameplay metrics include measures other than game outcome (i.e., wins or losses) – e.g., distance between players or number of lead changes;
- Different adaptation algorithms led to different kinds of balance in the game (some techniques led to more lead changes, others to smaller distances), but also had different degrees of noticeability and perceived fairness;
- The balancing techniques significantly improved the subjective play experience for novices (e.g., they felt that they had a better chance of winning), and did not detract from the subjective play experience for experts;
- The technique that led to more lead reversals was ranked first by both novices and experts, suggesting that there is particular value in manipulations that lead to balance in gameplay metrics (such as the number of lead changes).

Our studies are the first broad examination of the design and effectiveness of player balancing techniques in racing games. We provide four main contributions. First, we show that player balancing techniques can work well in this genre. Second, we show that balancing multiple mechanisms simultaneously (including both assistance and hindrance) shows promise for increasing the range of the system and reducing noticeability on the player's own performance. Third, we show that different approaches to dynamic balancing have different effects on game results and perceived gameplay. Fourth, we show that there are several gameplay metrics that capture important elements of the overall play experience, over and above typical win/loss outcomes. Our work broadens the understanding of player balancing and suggests new ways to improve games' ability to support shared play.

RELATED WORK

Game Balancing

Flow is a concept that describes optimal experience, and it occurs when people are engaged and interested in their tasks without feeling overly anxious [8]. Game designers and developers try to *balance* games to provide players with the right mix of challenge and achievement [10], which can promote the experience of flow. Designing games to be balanced is extremely important – if a player finds part of a game too easy or too difficult, they may become disengaged or frustrated [17]. Balancing is an essential part of games because part of people's willingness to play games is based on whether or not they find a game enjoyable [17,21].

In order for people to have fun together in multiplayer games, each player should have an equal chance to win [1], and it is people's ability to compete that often motivates them to participate [21]. However, an important challenge for multiplayer games is that performance is often based upon previous experience, complex in-game skills (e.g., producing button combinations) and physical abilities (e.g., reaction time). These differences in abilities and experience can lead to substantial differences in performance between two players, which can reduce competition and fun [17]. Therefore, game designers are also interested in how balancing competition (i.e., player balancing) can be achieved. This has been recognized in the game industry, where play testing and simulations help uncover such game elements.

Player Balancing

We identified four main ways that player balancing can be achieved in video games.

1. Matchmaking: Grouping players by ability is one way in which multiplayer online games balance player competition. Some games make use of complex systems that work to identify groups of players with similar skill levels [17]. These systems work in a manner similar to ladders for squash, and are important parts of Halo [12] and StarCraft 2. Two problems with matchmaking are that it requires an accurate idea of player skill, which may not be available [22], and that it requires a pool of available players of different skill levels.

2. Asymmetric Roles: Many games allow players to perform different tasks or roles that suit their level of skill or their abilities. For example, in sports, players can have very different responsibilities (e.g., offence or defence). In games like World of Warcraft, a successful raid party is composed of several different roles (e.g., tanks and healers).

3. Difficulty Adjustment: Difficulty adjustment is a common way to match a game's level of challenge to a player's ability. For example, many games have explicit difficulty levels that can be selected before the game starts. Adaptive difficulty adjustment is a less explicit approach that changes game elements or controls based on player performance [14]. While this can provide appropriate levels of challenge, adaptive difficulty adjustments must be designed carefully. For example, obvious techniques such as the "rubber-band" effect used to speed up trailing players in Mario Kart can lead to a sense of frustration for experienced players who can feel cheated [17]. Further, scaling game difficulty based on player performance or progress can lead to a player feeling that their achievements are meaningless [6].

4. Assists: Assists adjust a player's ability to perform basic in-game actions by simplifying the input required to correctly perform an action. For example, aim assistance helps a player aim to a target. Recent studies have shown that aim assistance can improve player performance in both 2D [5] and 3D games [20]. One study showed that players with less skill were able to improve and better compete with more skilled partners, and players reported having more fun with the game when the competition was tighter [5]. Further, these studies have shown that assists balance games, and yet players have difficulty noticing them, or may not notice them at all. These studies also show, however, that as games become more complex, other game elements can overshadow the effectiveness of aim assistance – e.g., one FPS study suggested that player movement, weapon choice, presence of friendlies, and enemy aggressiveness can change aim assistance [20].

Video Racing Game Balancing and Assists

A wide range of research has been carried out with racing games. For example, studies have examined the effects of input device [1,15], view [1] and track width [1] on player satisfaction and performance; or the extent to which players modify cars or tracks, or use game mode options [13].

Some modern racing games provide user-controlled assists. In Forza 4, a player can selectively activate assists for stability control, traction control, braking, steering, shifting, damage, and optimal path guidance. Recently, Debeauvais et al. [9] analysed the use of these assists, showing that certain assists were used by most players and never turned off (e.g., limiting the amount of damage from collisions), and some players actively try enabling and disabling assists as their performance improves; however, they did not investigate the use of assists for balancing in racing games.

Other racing games provide player balancing techniques, such as the Mario Kart examples discussed above. Mario Kart is primarily a social game, however, and few examples of player balancing are available in more "serious" titles.

PLAYER BALANCING FOR RACING GAMES

Racing games are competitive multi-player games that involve different play mechanisms than 2D or 3D shooters. Racing games typically involve two or more players racing on a specified track – each player controls a car, and the system provides a realistic simulation of accelerating, braking, and cornering. Racing games often use standard game controllers, with a joystick for steering and buttons for controlling accelerator and brake pedals. The view for these games provides a third-person view of the car (sometimes a first-person view is used or is an alternative) and the track, with feedback about car state (e.g., a speedometer and gear indicator) and the race state (e.g., indicators of the player's current position). Figure 1 shows our example racing game that follows these conventions – in this game, the two players' views are shown in a split-screen format.

Figure 1. Two-player racing game used in the study. Local participant at right; simulated opponent at left.

Factors affecting play experience in racing

We analysed existing racing games to determine gameplay metrics that could be used in player balancing techniques.

- *Wins and losses.* The most obvious result of a race is who wins – this is the traditional outcome that is a natural target for balancing; however, winning a race is made up of a driver's performance on several factors, and it may be difficult to balance wins without introducing obvious and artificial assists that detract from the play experience.

- *Distance between cars*. Regardless of the final outcome, players have a sense of being "in the race" and of competing by being close during the contest – if a player is too far back, they feel that there is no chance of winning.
- *Other car within sight*. For both the leader and the follower, being able to see the other car can increase the sense of competitiveness and the chance of challenging for the lead. Conversely, if the leader is out of sight, a race devolves to a single-player game for both players.
- *Lead changes*. Passing the current race leader can increase excitement and feelings of competition for both players – again, regardless of the eventual outcome of the race.

We designed balancing techniques to target these important game elements. Our techniques are made up of three parts: 1) *a set of balancing mechanisms* (i.e., the parts of the game that are manipulated to change player performance), 2) a set of *levels* that determine the strength of the mechanisms, and 3) a set of *adaptation algorithms* that determine when, and by how much, to apply the mechanisms during the race.

Assist-based Balancing Mechanisms

We implemented three assist balancing mechanisms that affect different aspects of a driving game: the car's speed, the car's acceleration, and the player's steering accuracy. Each mechanism can be manipulated in either direction to provide assistance to novices, or hindrance to experts.

Speed. Speed is adjusted by adding or subtracting a fixed amount from the player's current speed, based on the level of assistance or hindrance (from -5 to +5), and scaled proportionally to the top speed of the car. For example, when moving at the top (unassisted) speed, a level of +5 adds 5 m.p.h. to the car's speed, and a level of -5 subtracts 5 m.p.h.

Acceleration. Acceleration is adjusted by manipulating the airFriction parameter of the car, using an empirically-determined constant multiplied by the assistance level (from -5 to +5). A level of -5 meant that the car took about 1.5x the baseline time to reach top speed, and with +5, about 0.75x.

Steering. To adjust steering, the system compares the car's current heading to the direction of the road, and calculates the steering error – a number from 0 (no error) to 1 (90 degrees off course). The player's current steering amount is then adjusted by an amount proportional to the error and to the level (-5 to +5), either towards the optimal path (assistance) or away from it (hindrance). This means that the system does not steer for the player, it only adjusts the player's own steering actions (pilot testing showed that this feature was important for reducing noticeability).

We combined the three balancing mechanisms into a single approach that contained ten levels of possible adjustment to players' performance: five of assistance, and five of hindrance (i.e., from +5 to -5).

Adaptation Algorithms

We developed four schemes for applying the balancing mechanisms, plus a control condition with no balancing.

Realtime100. This algorithm sets the level of the balancing based on the distance between the players. Every 100 meters represents a level of assistance or hindrance – for example, if the player is 1-100 meters behind, they receive level +1 assistance; if they are 101-200 meters behind, they receive level +2 assistance. Similarly, if the player is 1-100 meters ahead, they receive level -1 hindrance.

Realtime40. Uses the same principle as Realtime100, but changes the balancing level with every 40-meter difference in distance (rather than 100 meters), thus it is more aggressive in adapting performance – players reach the maximum level of assistance or hindrance when there is 200 meters between players (rather than 500).

Rolling. Averages the level of assistance or hindrance from the Realtime100 scheme, using a rolling window of 50 seconds. The time window means that balancing takes longer to have an effect, but any advantage or disadvantage also lasts longer; e.g., a player who is behind will not receive assistance immediately, but after 50 seconds, their assistance level will remain, even if they pass the other player.

MaxDistance. A player's level of assistance is that used for the maximum distance that they were ever behind or ahead (using Realtime100), but all balancing is turned off when the gap between players is less than 50 meters. This scheme therefore balances the play to bring weaker players up to stronger players, but does not interfere once they are close.

Driving Game Apparatus

We implemented our algorithms in C# in a realistic two-person driving simulation based on the *Racing Game* prototype included with Microsoft's XNA environment [18] (Figure 1).

Simulating Opponents

To test our adaptation algorithms and to provide consistent opponents in our main study, we simulated the opponent in the game by building an automatic driving controller. Simulated opponents can provide a more reliable level of performance across different races and different participants, and also reduced the overall number of participants needed for the study, allowing us to gather data about more conditions.

Car speed and steering were automatically controlled in order for the simulated driver to follow a target path around the track. The controller used a proportional-derivative feedback control law [2] so that the simulated driver would automatically respond to disturbances, such as collisions, and continue racing. The performance of the simulated driver could be hindered by adding noise to the feedback so that the controller's actions were not perfect, and consequently the simulated driver's lap time increased.

Past work has used a predetermined script of driving actions or behaviors [7]. Our controller-based simulated driver uses feedback of in-game events to respond appropriately in the context of a specific race. Importantly, balancing mechanisms can be directly incorporated into the controlled driver, which is not possible with a pre-scripted driver.

Simulated drivers were used in a pilot study to tune the parameters in the balancing mechanisms. Races between two simulated drivers with different levels of performance were conducted over a wide range of balancing mechanism parameters in order to estimate a range of parameters that resulted in the best-balanced races. This simulation-based tuning approach allows for large numbers of simulated races to be performed in a fraction of the time that would be required if done with human participants. The study showed that multiple mechanisms are more effective in adjusting performance than single mechanisms, and that the combination of all three performed best of all. We therefore used all three mechanisms in the main study.

Simulated drivers were also used in the main study to act as the opponent for our human participants. The target path around the racetrack for the simulated drivers was determined empirically by recording the path taken by expert and novice human drivers prior to conducting our experiment. Four different paths (source traces) were used for each of the expert and novice simulated drivers, which allowed for minor lap-to-lap variation, making the simulated drivers seem more realistic to the human players.

Pilot Test of Noticeability on Players' Own Performance

We carried out a pilot study to assess the noticeability of the three balancing mechanisms on a player's own performance, at different global levels of assistance or hindrance. Four participants from our lab carried out solo races with the different levels of assistance and hindrance, and then guessed at the level. The system applied either one mechanism or all three.

Results of the pilot showed that although all mechanisms had a roughly linear effect on performance (both with negative and positive modifications), noticeability was greater when individual techniques were used. The idea of "spreading around" the assist/hindrance appeared to reduce the noticeability in any one area. In addition, hindrances were consistently less noticeable than assists. We further discuss the issue of noticeability later in the paper.

STUDY: TESTING AND COMPARING THE TECHNIQUES

Our final study compared the effects of the four adaptation algorithms (which determined when and how the balancing mechanisms were applied) plus a control condition.

Experimental Conditions

Each condition combined the three balancing mechanisms (speed, acceleration, and steering) with one of the five adaptation algorithms (Control, Realtime100, Realtime40, Rolling, and MaxDist) in our racing game. Each adaptive algorithm (other than Control) used the mechanisms in the same way; therefore there were five experiment conditions, comprised of the four adaptation algorithms and the control.

Procedure

Participants completed a consent form and demographics questionnaire, and were introduced to the controls for the driving simulator. Participants self-identified as a novice or expert in racing games, and the experimenter subjectively confirmed these self-assessments (no participants were found to have over- or under-estimated their skill level). Participants were introduced to the two-player version of the game (see Figure 1), and told to complete the course as quickly as possible. Participants then completed 10 two-person races, two for each adaptation condition (order was rotated using a Latin Square, blocked by condition and data from the two races were aggregated for central tendency).

To provide opponents whose skill remained consistent across participants and conditions, all opponents were simulated drivers: we simulated a novice driver if the participant was an expert, and an expert driver if the participant was a novice. The simulated opponent was implemented as previously described, and one of the four described source traces was used for each condition (source trace order was counterbalanced). Participants were told that their opponents for the races were other humans – either novices or experts – who would be playing across the network from another location. We also stated that each set of races could involve a set of (unspecified) manipulations to the cars.

After each race, participants completed an experience survey. The questionnaire included subjective questions about enjoyment, competitiveness, and perception of fairness, and also included PENS questions [19] that examine the concepts of competence, autonomy, and relatedness from self-determination theory. After the session, participants ranked all conditions in terms of overall preference.

Participants and Apparatus

Thirty participants were recruited from a local university (7 female; mean age 24 years); 15 were experts, and 15 were novices (based on self-assessment as described above).

The study was conducted on a Windows 7 PC with a 1920x1080 screen. The software recorded all experimental data including race times and player positions.

Design and Research Questions

The study used a 5×2 mixed-factorial repeated-measures design, with between-participants factor *Expertise* (Expert or Novice) and within-participants factor *Adaptation* (Control, Realtime100, Realtime40, Rolling, MaxDist). We examined dependent measures relating to performance (wins, time differences, distance between cars, number of lead changes), and measures relating to experience (perception of enjoyment, competitiveness, and fairness; and measures of competence, autonomy, and relatedness).

We explored five questions in the study:

Q1. *Does balancing work?* Will the gap between expert and novice performance be reduced in the balancing conditions compared to the control condition?

Q2. *Does balancing improve novice experience?* Are novices' ratings of subjective experience higher in balancing conditions than in the control condition?

Q3. *Does balancing detract from expert experience?* Are experts' ratings of subjective experience worse in the balancing conditions?

Q4. *Is balancing preferred overall?* Do players prefer the balanced conditions over the control condition?

Q5. *How do the adaptation schemes differ?* What are the differences in the adaptation conditions' effects on performance and experience measures?

Results

Our results are organized below by analysis of performance measures and experience measures; we return to the research questions after the presentation of results.

Performance Measures

The subsections below consider win/loss rate, race time, mean distance between players over the course of the race, lead changes, and the amount of time that players were close enough to see one another (within 80 meters). Recall that experts and novices both played against simulated opponents (of the opposite skill level), and thus all novice-expert comparisons are comprised of independent race data.

In all of the performance analyses where experts and novices could be compared, experts were significantly better (all $p < 0.05$). This was expected, so we do not report these results, in order to focus on the effects of the balancing techniques. In all charts, error bars show +/- one standard error.

Wins and Losses

As expected, experts won the large majority of races: human experts won 94.6% of their races; human novices, only 7.3%. The win-loss results are shown in Figure 2; only two balancing techniques led to any wins for human novices, and three techniques led to losses for human experts.

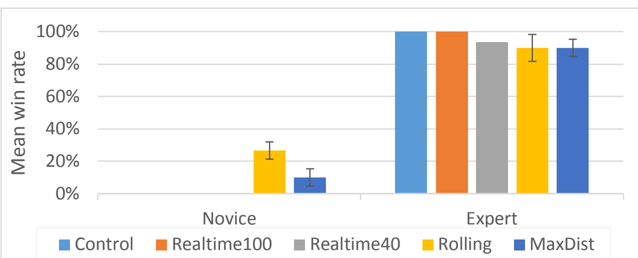

Figure 2. Mean win percentage by expertise and adaptation.

RM-ANOVA showed no overall main effect of Adaptation ($F_{4,112}$=2.24, p=.069) on wins, but did show an interaction between Adaptation and Expertise ($F_{4,112}$=7.31, p<.0001). For experts, there was no effect ($F_{4,56}$=1.59, p=.189), but for novices, the effect of Adaptation was significant ($F_{4,56}$=7.66, p<.0001).

A follow-up Tukey HSD test on the novice results showed that *Rolling* and *MaxDist* had significantly more wins than the other three balancing techniques, and also that *Rolling* was significantly higher than *MaxDist* (all p<0.01).

Difference in Race Time

Overall, human experts finished their races 6.6 seconds ahead of their simulated opponents, and human novices finished 11.2 seconds behind. The results for the different adaptation conditions are shown in Figure 3.

Repeated-measures ANOVA showed an overall main effect of Adaptation ($F_{4,112}$=148.3, p<.0001), and an interaction between Adaptation and Expertise ($F_{4,112}$=15.33, p<.0001). A follow-up Tukey HSD test on the main effect showed that all of the balancing conditions had smaller time differences than the control condition (all p<0.01) and that *MaxDist* had a smaller time difference than *Realtime100* (p<0.05). The interaction is illustrated in Figure 3: there was less of a time difference between human experts and their simulated opponents than for human novices and their opponents.

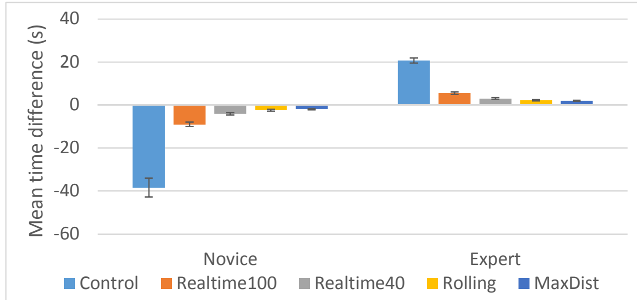

Figure 3. Mean time difference between players.

Mean Distance Between Players Over Entire Race

We recorded the distance between the players throughout the race. On average, human experts were 92 meters ahead of their simulated opponents, and human novices were 112 meters behind their simulated opponents (see Figure 4).

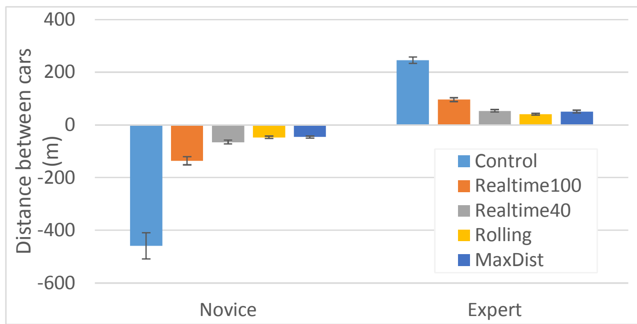

Figure 4. Mean distance between players over the race.

RM-ANOVA showed an overall main effect of Adaptation ($F_{4,112}$=155.9, p<.0001), and an interaction between Adaptation and Expertise ($F_{4,112}$=18.66, p<.0001). As before, human experts were not as far ahead of their simulated opponents as the human novices were behind their opponents. A follow-up Tukey HSD test on the main effect showed that all of the balancing conditions had smaller distance differences than the control condition (all p<0.01) and that *MaxDist* and *Rolling* had smaller differences than *Realtime100* (p<0.05).

Lead Changes

The number of times the lead changed during the race is shown in Figure 5. Overall, the races with human novices had 4.6 changes, and the races with human experts had 3.1.

RM-ANOVA showed an overall main effect of Adaptation ($F_{4,112}$=45.84, p<.0001), but no interaction between Adaptation and Expertise ($F_{4,112}$=1.19, p<.318). Tukey HSD showed that *Rolling* and *MaxDist* had significantly more lead changes than the other three conditions, and that *Rolling* also had more than *MaxDist* (all p<0.01).

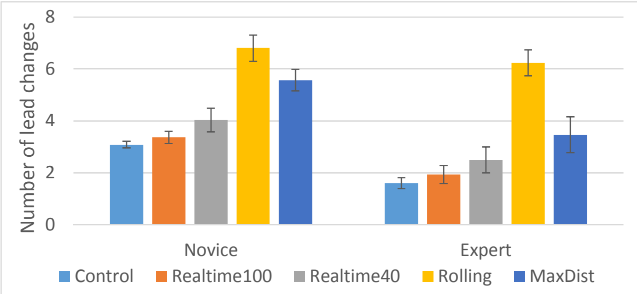

Figure 5. Mean number of lead changes.

Time that Players were Within 80m

To analyse how the different conditions kept players within sight of one another, we recorded the number of seconds that the distance between the two cars was 80m or less (an estimate of the distance at which the front car could be seen, based on informal tests). Results are shown in Figure 6.

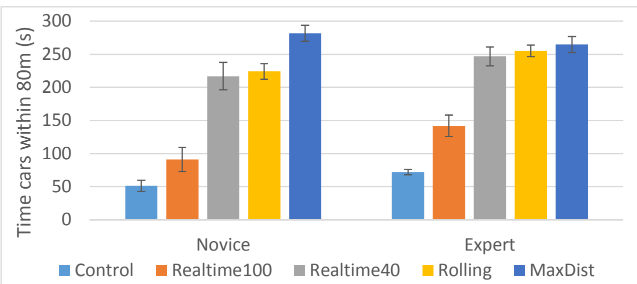

Figure 6. Mean time players were within 80m.

Repeated-measures ANOVA showed an overall main effect of Adaptation ($F_{4,112}$=162.1, p<.0001), and an interaction between Adaptation and Expertise ($F_{4,112}$=3.01, p<.05). Tukey HSD on the main effect showed that all the balancing conditions led to significantly longer times where players were within 80m (all p<0.01), that *Realtime100* led to less within-sight time than the other balancing conditions (p<0.01), and that *MaxDist* had a longer time than *Realtime40* (p<0.05). The interaction can be seen in Figure 6; there was no difference between *MaxDist* and *Rolling* or *RealTime40* for Experts, but there was for Novices.

Experience Measures

Subjective Questions

Participants completed a questionnaire after each adaptation condition. A summary of responses to the subjective experience questions on this survey is shown in Figure 7. (Means are shown in the charts, but all analyses discussed below used non-parametric tests on the rank data).

We used Friedman ranks tests to look for effects of Adaptation on the four experience measures. We analysed experts and novices separately; Table 1 shows the results.

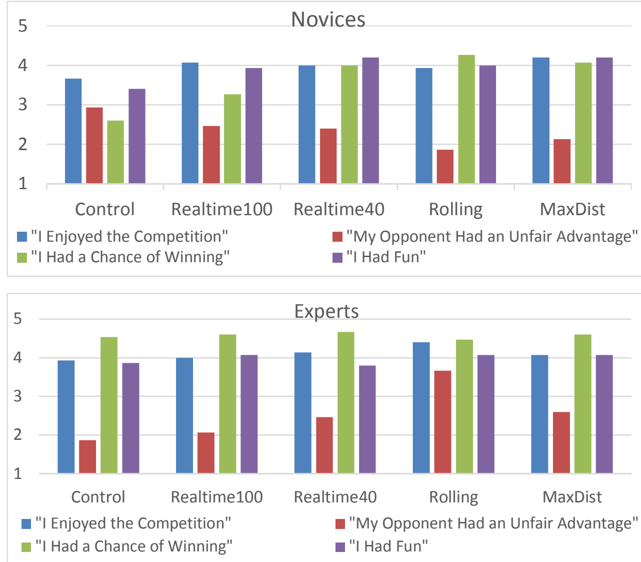

Figure 7. Mean subjective experience responses (5-point Likert scale: 1 = "strongly disagree", 5 = "strongly agree").

We used Friedman ranks tests to look for effects of Adaptation on the four experience measures. We analysed experts and novices separately; Table 1 shows the results.

	Experts		Novices	
	χ^2_r	***p***	χ^2_r	***p***
I enjoyed the competition	1.69	.79	2.60	0.63
I had a chance of winning	0.53	0.97	25.2	**<0.01**
My opponent had an unfair advantage	18.4	**<0.01**	8.04	0.09
I had fun	1.20	0.88	6.64	0.16

Table 1. Friedman rank tests of subjective responses.

For the two measures that showed significant effects of Adaptation, we carried out follow-up pairwise Wilcoxon tests. For the novices' responses to "I had a chance of winning", followup showed that *Realtime40*, *Rolling*, and *MaxDist* had significantly higher ratings than *Control* (p<0.05). For experts' responses to "My opponent had an unfair advantage", followup showed that *Rolling* had higher ratings than *Control* and *Realtime100* (p<0.05).

PENS Measures

We gathered perceived competence, autonomy, and relatedness using a validated scale after each condition. RM-ANOVA revealed a main effect of Adaptation on competence ($F_{4,112}$=3.9, p<.005) and relatedness ($F_{4,112}$=4.6, p<.002). Players rated their competence as lower after *Control* than after *Realtime40*, *Rolling*, or *MaxDist* (all p<0.02); however, a significant interaction with Expertise ($F_{4,112}$=4.3, p<.003) shows that these results held true for Novices, but that there were no differences for Experts. A significant interaction of Adaptation and Expertise on Autonomy ($F_{4,112}$=2.7, p<.004) revealed that Novices rated their autonomy lower after *Control* than after *Rolling* or *MaxDist* (all p<0.03), whereas there was no difference for Experts. Both experts and novice players rated their relatedness after *Control* lower than after *Realtime40*, *Rolling*, or *MaxDist* (all p<0.05) with no interaction with Expertise (p>0.06).

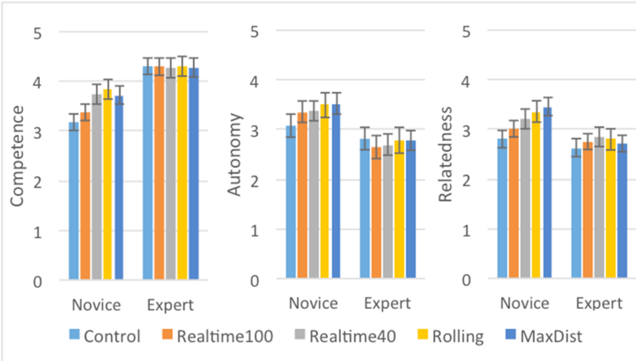

Figure 8. Mean perceived competence, autonomy, and relatedness by expertise and adaptation (1-5, higher is better).

Overall Preferences

At the end of the session, participants ranked the five conditions in order of overall preference (5 = least preferred, 1 = most preferred). Mean rankings are shown in Figure 9. As shown in the figure, there were some differences between novices' and experts' preferences (e.g., experts ranked *Control* higher than novices), but both groups ranked *Rolling* as the best condition overall, both in terms of average rank and number of first-rank votes (10 of 15 novices and 6 of 15 experts ranked *Rolling* in first place).

We looked for effects of Adaptation on preference with a Friedman rank test. There was a significant effect for novices (χ^2_r=19.36, p<0.001) but not for experts (χ^2_r=6.03, p=0.20).

Follow-up pairwise Wilcoxon tests for novices' data showed that *Rolling* was preferred over *Control* and *Realtime40*, and that *MaxDist* was preferred over *Control* (all p<0.05).

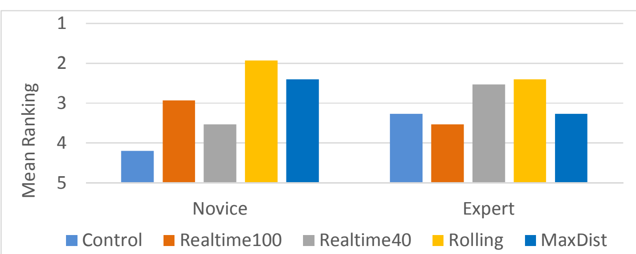

Figure 9. Mean rankings by expertise and adaptation

Summary of Results by Research Question

Q0: Does balancing multiple mechanisms simultaneously increase range and reduce noticeability?

- Pilot studies showed that combining all three mechanisms was best at increasing range (using simulated drivers) and reducing noticeability (using human participants).

Q1. Does balancing work?

- All of the adaptive algorithms had positive effects on at least some of the race performance variables, although experts still won the large majority of races.
- Various adaptations improved all of the elements of performance that we tested (win/loss rates, time differential, distance between cars, and lead changes).
- The *Rolling* and *MaxDist* adaptations were best at reducing performance differences between experts and novices.

Q2. Does balancing improve novice experience?

- Three adaptations (*Rolling*, *MaxDist*, and *Realtime40*) changed novices' experience of the game's competitiveness, and improved experienced competence and relatedness to the opponent. *Rolling* and *MaxDist* also raised the experience of autonomy.
- No effects were found for subjective enjoyment or fun.

Q3. Does balancing detract from expert experience?

- One adaptation (*Rolling*) changed experts' perception of fairness; this technique led to increased feelings that the other player had an unfair advantage.
- No effects were found for questions about enjoyment, fun, or competitiveness, and experts did not rate their competence lower for adaptions that hindered them.
- Three adaptations (*Rolling*, *MaxDist*, and *Realtime40*) improved experts' feelings of relatedness to the opponent.

Q4. Is balancing preferred overall?

- The *Rolling* technique was the top-ranked technique by both novices and experts.
- Novices ranked all of the adaptations higher than the control condition.
- Novices significantly preferred the *Rolling* and *MaxDist* conditions over the other conditions.

Q5. How do the adaptation schemes differ?

- The more aggressive adaptations (*Rolling* and *MaxDist*) had greater effects on performance variables – and were the only techniques that led to wins for human novices.
- The more aggressive techniques also led to an increase in experts' feelings that the opponent had an unfair advantage; but also represented the most preferred technique overall by novices and experts (*Rolling*).

DISCUSSION

In the following sections we discuss explanations for our results and underlying ideas that are applicable in other settings and for other games.

Explanations for Results

Why did the balancing techniques succeed?

Player balancing techniques work by artificially assisting the weaker player (or by hindering the stronger player), but the presence of a balancing technique does not guarantee that the outcome of play will be balanced, since there are many aspects of a game that can complicate the application of balancing [19]. In our study, there are three main reasons why the player balancing techniques worked well.

Continuous balancing. Our techniques involved game activities that happen continuously (e.g., steering and speed) rather than discrete events such as the shooting events studied in previous studies [4,19]. This is valuable for player balancing because it provides more opportunity to assist (or hinder) the player – as a result, the manipulation can be smaller and potentially less noticeable.

Increasing the adaptation range. Player balancing schemes can fail to work effectively if they cannot provide enough assistance (i.e., experts are too good). Our techniques substantially increased the range of abilities that could be balanced through two innovations: manipulating three game mechanisms instead of one, and using both positive and negative manipulation. These innovations made a major difference in the range of differences that could be balanced, and did not increase noticeability.

Effects on important in-game events. Two of our techniques (*Rolling* and *MaxDist*) led to larger changes in the game state, which meant that there were more changes to elements of the race that are important to players. For example, lead changes or keeping the players within sight of one another appears to add more excitement and competitiveness to the game than other manipulations.

Why were Rolling and MaxDist preferred over Realtime?

The *Rolling* and *MaxDist* techniques led to the largest changes in game variables, were preferred by participants, and led to increased feelings of relatedness. The main differences between these techniques and the two *Realtime* techniques were the aggressiveness and duration of the balancing manipulations, and these differences indicate that these qualities may be important for improving play experience. The *Realtime* techniques can be considered as "asymptotic" methods, in that the level of assistance or hindrance is proportional to the distance between the players. This means that as the novice player catches up to the expert, assistance is gradually reduced – resulting in Sisyphean situations where the novice can get close to the expert but never reach them.

In contrast, *Rolling* and *MaxDist* do not follow an asymptote, but rather use a "pendulum" approach that maintains a stronger manipulation for a time period (*Rolling*) or until a distance threshold is reached (*MaxDist*). Although this approach makes it more obvious that players are being helped (reducing feelings of autonomy for novices), the effects on game events (e.g., lead reversals, occasional wins by novices) made for a more competitive experiences for all.

This result may underlie the design of existing techniques like the Bullet Bill powerup – that is, it is more fun (for all players) if the last-place driver occasionally moves into the lead than for them to remain in last place but by a smaller margin. Techniques like *Rolling* represent an important advance over the obviousness of Bullet-Bill-style assists, however, since *Rolling* is applied dynamically and invisibly (rather than as an explicit powerup) and only affects the player's driving actions, rather than taking over control of the car.

Generalization and Application of the Results

Application to other game genres

Several of the successful ideas used here can be applied to player balancing in other games. First, there are many genres that use the basic idea of racing as one element in the larger game (e.g., military simulations often involve travel by vehicle, and many games involve attempts to reach an objective before others). In these situations, the balancing techniques used here could be directly applied (although they will only balance the race-based parts of the game).

Second, balancing through simultaneous manipulation of multiple mechanisms could be used in several game types. For example, balancing in a first-person shooter game could involve factors such as health, hitbox size, movement speed, or visibility to other players; in an RTS game, balancing could affect factors such as the appearance rate of resources, or a unit's speed, power, or build time.

Third, the idea of continuous balancing could be applied even in genres where many of the important game elements are discrete events (e.g., shots in FPS games). In these situations, continuous balancing could be used on game variables such as avatar movement, the control-to-display ratio of aiming actions, or healing speed.

Fourth, assistance and hindrance can also be applied to other genres. Pairing hindrance with continuous game activities may be more effective than with discrete events, because whereas experienced players may notice even small degradations in the discrete behaviour of the system (e.g., shooting accuracy), it may be more difficult for them to predict the exact behaviour of a continuous variable.

Noticeability and Fairness

Our assessment of noticeability was limited to people's perception of how different techniques affected their own performance (in the pilot), and their perception of whether the other person had been given an advantage (main study). There are several other aspects of both of these issues that can be explored, and our investigation is only a starting point. Nevertheless, our findings suggest several issues that designers should consider regarding the obviousness of balancing techniques in real-world racing games.

Noticeability of effects on other players may have been higher in our study than in the real world. We stated that balancing techniques could be in use; it is possible that experts over-estimated the help that their opponent received – experts felt like their opponent had a small advantage even in the control (i.e., no balancing) condition (Figure 7).

It may be that noticeability has more to do with the players' previous knowledge of each others' abilities than with in-game changes caused by the techniques. Players in our study likely built up this kind of knowledge over the five races (even though the opponent was simulated, it had similar base ability throughout), and so the adaptations may have been more obvious than they would be in the real world. Suggesting experimentally that the participants are racing unique players each race could tease out the differential effects of algorithm and familiarity. The possible relationship between player familiarity and technique noticeability suggests that in situations where players do not know their opponent's abilities, it will be more difficult for them to notice balancing.

Future Work

We will extend this research in several ways. First, we will further test our techniques by examining their use in real-world social play situations – e.g., a race tournament to determine whether pairs of players with skill differences react like the expert-novice pairings studied here. Second, we will further explore effects of our techniques on different types of noticeability. Third, we will test the underlying ideas of multiple-variable balancing and positive-and-negative effects in a different genre such as FPS or RTS. Fourth, we will further examine the relationship between balancing and in-game play experiences that are not part of ultimate game outcomes.

CONCLUSION

Player balancing is a solution to the problem of competitive play situations in which players have different skill levels. Although player balancing can be seen in many games, few studies look at the design and effectiveness of these techniques – particularly outside of shooting games – and the research that has been done suggests that real games introduce many elements that can complicate balancing. In this paper, we provide initial knowledge about player balancing in the popular and competitive racing genre.

We carried out investigations to explore noticeability and balancing effectiveness, and tested the effects of several balancing techniques on measures of game performance and play experience. A comparative study found that balancing techniques can be very successful in a realistic racing game – our techniques significantly reduced differences in several aspects of race performance, were preferred by both experts and novices, increased novices' feelings of competitiveness, and did not detract from the play experience for experts. Our results provide initial understanding of how player balancing techniques can be designed and used in multi-player racing games, and suggest ways that the principles underlying our techniques can be applied to other genres.

REFERENCES

1. Adams, E. *Fundamentals of Game Design*. New Riders, 2010.
2. Araki, M. PID control. *Control systems, robotics and automation*, *2*. 2002.
3. Baldwin, A., Johnson, D., & Wyeth, P.A. The effect of multi-player dynamic difficulty adjustment on the player experience of video games. In *Proc. CHI EA*, 2014, 1489-1494.
4. Bateman, S., Doucette, A., Xiao, R.B., Gutwin, C., Mandryk, R.L., & Cockburn, A. Effects of View, Input Device, and Track Width on Video Game Driving. In *Proc. GI, 2011*, 207-214.
5. Bateman, S., Mandryk, R.L., Stach, T. and Gutwin, C. Target assistance for subtly balancing competitive play. In *Proc. CHI 2011*, ACM Press (2011), 2355-2364.
6. Bostan, B., and Öğüt, S. Game challenges and difficulty levels: lessons learned From RPGs. In *International Simulation and Gaming Association Conference*, 2009.
7. Chan, H. *Vehicular racing simulation: a MEL scripting approach* (Phd dissertation, Texas A&M Univ.), 2014.
8. Csíkszentmihályi, M. Flow: *The Psychology of Optimal Experience.* Harper and Row, 1990.
9. Debeauvais, T., Zimmermann, T., Nagappan, N., Carter, K., Cooper, R., Greenawalt, D., & Solberg, T. Off With Their Assists: An Empirical Study of Driving Skill in Forza Motorsports 4. In *Proc. FDG 2014.*
10. Falstein, N. Understanding fun – the Theory of Natural Funativity. In *Introduction to Game Development*, Rabin, S. (ed.), Charles River Media, 2005, 71-98.
11. Gerling, K., Miller, M., Mandryk, R., Birk, M., Smeddinck, J. Effects of Balancing for Physical Abilities on Player Performance, Experience and Self-Esteem in Exergames In *Proc. CHI 2014*, 2201-2210.
12. Herbrich, R., Minka, T., and Graepel, T. TrueSkill(TM): A Bayesian skill rating system. In *Advances in Neural Information Processing Systems*, 20, 569-576, 2007.
13. Hullett, K., Nagappan, N., Schuh, E., and and Hopson, J. Empirical analysis of user data in game software development. In *Proc. ESEM 2012*, 89-98.
14. Hunicke, R. The case for dynamic difficulty adjustment in games. In *Proceedings of Advances in Computer Entertainment Technology*, vol. 265, 429-433, 2005.
15. Kavakli, M., Thorne, J.R. A Usability Study of Input Devices on Measuring User Performance in Computer Games. In *Proc. of ICITA* 2002.
16. Mueller, F., Vetere, F., Gibbs, M.R., Edge, D., Aga-manolis, S., Sheridan, J.G., and Heer, J. Balancing Exertion Experiences. In *Proc. of CHI 2012*, 1853-1862.
17. Newheiser, Mark. Playing Fair: A Look at Competition in Gaming. *Strange Horizons*. Published March 9 2009, www.strangehorizons.com/2009/20090309/newheiser-a.shtml. Accessed May 6, 2014.
18. Nitschke, B. *XNA Racing Game.* www.exdream.com/XnaRacingGame, accessed May 6, 2014.
19. Ryan, R., Rigby, C., & Przybylski, A., Motivational pull of video games: A self-determination theory approach. Motivation and Emotion, 2006, 30, 347-365.
20. Vicencio-Moreira, R., Mandryk, R.L., Gutwin, C., and Bateman, S. The effectiveness (or lack thereof) of aim-assist techniques in first-person shooter games. In *Proc. CHI 2014*. ACM Press (2014), 937-946.
21. Vorderer, P., Hartmann, T., and Klimmt, C. Explaining the enjoyment of playing video games: the role of competition. *Proc. Entertainment Computing 2003*, 1-9.
22. Youssef, A. and Cossell, S. Thoughts on adjusting perceived difficulty in games. In *Proc. Australasian Conference on Interactive Entertainment*, 1-4, 2009.