

Come Meet Me at Ulduar: Progression Raiding in World of Warcraft

Jeffrey Bardzell

Indiana University
919 E Tenth St.

Bloomington, IN 47401
jbardzel@indiana.edu

Jeffrey Nichols

IBM Research

650 Harry Road San
Jose, CA 95120 USA
jwnichols@us.ibm.com

Tyler Pace

Indiana University
919 E Tenth St.

Bloomington, IN 47401
tympace@indiana.edu

Shaowen Bardzell

Indiana University
919 E Tenth St.

Bloomington, IN 47401
selu@indiana.edu

ABSTRACT

In spite of decades of research on virtual worlds, our understanding of one popular form of virtual world behavior—raiding—remains limited. Raiding is important because it entails intense, high-risk, and complex collaborative behaviors in computer-mediated environments. This paper contributes to CSCW literature by offering a longitudinal analysis of raiding behavior using system data manually collected from the game world itself, comparing two raiding teams as they worked through the same content. Supplemented with interviews and chat transcripts, this research sheds light on what actually happens during raids across four different temporal scales: seconds, hours, days, and months. It also distinguishes between behaviors that are imposed by the system design and those chosen by players. Finally, it derives two viable raiding styles from the data.

Author Keywords

HCI, World of Warcraft, raids, virtual teams, failure

ACM Classification Keywords

H5.m. Information interfaces and presentation: Miscellaneous.

General Terms

Design

INTRODUCTION

The rise of massively multiuser virtual environments, which had been building up since the 1990s with MUDs, MOOs, and EverQuest, hit a tipping point in the mid-2000s with World of Warcraft (WoW) and Second Life, which, at the height of their popularity, boasted over 24 million users between them. Research in anthropology, psychology, CSCW, education, HCI, and media studies has shed light on virtual world behavior, with implications for systems design, social psychology, virtual teams, pedagogy, and cultural theory. Such research has illuminated virtual world social behavior, from virtual rape [10] to the life and death of guilds [11]. Full-scale ethnographic research has present-

ed studies of virtual community diaspora [25], of virtual personhood [4], and of WoW as a platform for cross-cultural aesthetic experience [23].

As researchers continue to shed light on the social behaviors and their meanings that unfold in virtual worlds, one form of social behavior continues to generate interest: end-game raiding. As large-scale, complex group activities that involve 10-40 people working together in real time to solve extreme problems, raids have clear implications for CSCW, in research areas such as knowledge gathering, software development, or military applications, especially where large virtual teams, fast time constraints, and/or a high risk of failure are prevalent. A concept critical to players' understanding of raiding is *progression*. For most players, the most basic definition of progression refers to how much high-end content guilds have worked through (i.e., how many raid bosses have been defeated).

The central goal of this project is to obtain a better understanding of raid progression, with particular emphases on how raid behavior and events shed light on player decision-making, leadership, failure, and multi-user interaction design. We offer two main contributions, based primarily on log data obtained from two raid teams in WoW over five months. First, we offer basic metrics of raiding in terms of time, human involvement, distribution of successes and failures, etc., providing a foundational understanding of *progression raiding behavior*. Second, we offer an analysis of the *management choices* two raiding teams made during the study within in distinct temporal structures.

BACKGROUND: RAIDING IN WOW

The raid is a type of large, complex, real-time activity common in many virtual gaming worlds. In WoW, they are available as 10- and 25-player challenges. In a raid, players typically enter a separate area of the game, called an "instance," and they take on a series of difficult challenges, usually involving combat against boss enemies. Success brings rewards: when beaten, raid bosses drop "loot"; in fact, they drop the most coveted items in the game, items that can *only* be acquired through raids. For any given win, only a small fraction of players will get useful loot. Additionally, raids are quite difficult; the majority of them result in the boss killing all of the raid team, called "wiping." When a team wipes, no one gets loot from that boss.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW'12, February 11-15, 2012, Seattle, Washington, USA.

Copyright 2012 ACM 978-1-4503-1086-4/12/02...\$10.00.

Success in raids is generally predicated on teams having sufficient resources to take on these elite challenges. Because of the difficulty of building sufficiently resourced teams, players usually group together across sessions. One mechanism to support persistent grouping is the “guild.” As noted in earlier research [14, 30], guilds form around many aspects of the game, serving social as well as ludic needs. One major type of guild is the raid guild, which is a group of highly experienced players who take on the game’s most difficult challenges. Not all raid guilds are the same: the most “hardcore” guilds raid 5-7 nights per week, are often highly regimented [30], and compete with other hardcore raid guilds on the server for first kills and other achievements, which are prioritized higher than social goals. Far more common is what Golub [16] refers to as “medium-core” raiding guilds, which blend both achievement and social goals. Medium-core raiding guilds generally try to raid 2-3 nights per week and prioritize fun and achievement equally. Because medium-core raiding guilds are far more typical than hardcore guilds, this study focuses on them.

Progression is the main aspiration of most raiding guilds. One of the most basic progression metrics raiding guilds use is whether they have cleared a raid’s content before a new patch is released. *Content patches* are introduced by WoW designers about every 4-6 months. New content patches typically contain a new raiding instance with new rewards that are substantially better than anything previously available. For raid guilds that do not achieve clear content before a new patch, they typically move on to the new content and leave the old content behind.

RELATED WORK

The topic of user research in virtual worlds is hardly new to CSCW. Much of this work has come into CSCW collectively under the banner of “productive play.” This body of work has done much to correct commonsense but highly problematic distinctions between work and play. CSCW research on virtual worlds has touched upon a number of key themes. This includes contributions that explicitly confront the work/play dichotomy already mentioned [5, 7, 33]. The exploration of virtual worlds for enterprise use is another area [7, 18]. Others [29, 30, 32] have explored the motivations and nature of enjoyment in virtual worlds to suggest design strategies for future virtual world collaborative applications. The significance of avatars as a particular kind of embodiment is another theme of this research [22, 34]. The multiplayer nature of virtual worlds also means that much of the research goes beyond the individual. Sociability, teams, grouping, and collaboration are a dominant sub-topic, including [3, 9, 11-15, 21, 24]. Some of the contributions focus on the nature of virtual worlds as spaces or ecologies, especially the particular ways that these spaces foster play, productivity, and collaboration [2, 6, 20]. Kahai et al. [19] emphasize that the medium or affordances of virtual worlds shape leadership and collaboration, proposing that researchers should “seek to understand leadership emergence during virtual team collaboration in virtual

worlds.” An underlying concern in all of this research is the extent to which virtual collaborative behavior is shaped or conditioned by technical systems, a concern also reflected in existing WoW research [e.g., 11, 23, 30].

Raiding in massively multiplayer games has also been researched previously [8, 16, 23]. These works ethnographically characterize the authors’ experiences as members of a raid team. Common to all of this raid research, however, is that it relies on self-report, including the self-report of the researcher participating in the raids supported with interviews with fellow players. No published studies on raids, however, are based on data obtained from the game itself; neither are any raid studies based on the activities of more than one guild; nor has any of the raid research focused extensively on raid leadership. Several social networking studies, including [11, 14, 30] offer analyses of guilds, including raid guilds, using data derived from the game, but none of these focus specifically on raids or reveal much about actual raiding behavior. Williams et al. [30] account for this problem as follows: “a range of obstacles will confront any similar future work. Chief among these is securing access to the database of players and logs of player behavior.” Blizzard notoriously does not collaborate with researchers, so access to their databases remains out of reach. In the present study, we present results based on longitudinal raid log data collected manually from two guilds. This data enables us to present the first analysis of raiding behavior from more than one guild, progressing through the same game content, based on data from the game itself.

METHODOLOGY

As with prior primarily quantitative WoW research [e.g., 11, 14, 30], the present study is informed on three levels. First, the authors of this study have been active participants in WoW for years, each logging hundreds of hours and participating in several guilds; three of the authors are guild leaders (of three different guilds on three different servers). Our research questions, data collection methods, and data analysis and interpretation are all profoundly shaped by this participation. The second level is system-generated combat log data manually harvested for this study; this data is our primary source of evidence. The third level is qualitative, involving follow-up interviews conducted with guild officers and study of chat logs.

We collected data about two medium-core raid teams’ progress through the 25-person version of the most difficult raid zone at the time: “Ulduar.” Released as part of content patch 3.1 in April 2009, Ulduar contains 14 bosses. The first raid attempt for each guild began on April 14, 2009. As of August 25, 2009, when we cut off data for the purposes of the study, neither guild had killed more than 12 of the 14 bosses. We codename the two raid teams “Judgment” and the “Titans.” Officers from each of these raid teams are co-authors of this paper, and two authors were not in either raid team.

Data Collection

For each raid team, we manually collected combat logs and chat logs of the raids. *Combat logs* are system-generated text records that list every spell cast and weapon swing by every player in the raid group and every enemy they face; they are arguably the best source of behavioral data available in WoW. Combat logs are commonly collected by raiding guilds and fed into online analysis tools with the results made available in the raid guild’s online forums. We collected a total of 65 combat logs, 31 from Titans and 34 from Judgment, each representing a single raid session.

Chat logs are text records of all the chat that appears in the chat channels in the game during the raid. This includes messages written by players, scripted event dialogue, and a variety of status messages generated by the game. The extent of the player-written messages depends on how heavily the guilds use VoIP applications; additionally, raid leaders often discuss planning issues in private chat channels before announcing their decision to the rest of the raid through a VoIP channel. In this paper, we use chat logs as a secondary source of data, focusing on the raid leader chat for qualitative results and game-generated messages that provide a clearer indicator for some events than the combat logs. We collected the same number of chat logs as combat logs, again, one for each session.

The final data source came from two interviews of guild officers who are not part of the research team. Interviews were conducted by researchers outside of the guilds in question. The two semi-structured interviews lasted approximately one hour each and included questions concerning the officers’ impressions of the guild’s progression, including achievements and failures, planning, and risk management. More informal follow-up interviews happened afterwards.

Data Extraction and Analysis

From the combat and chat logs of each raid, we extracted a summary of each raid as follows:

- How long it took for the raid to form from the start of invites to the first engagement with an enemy.
- How many bosses the raid fought.
- Which bosses the raid fought, and in what order.
- For each boss fight, how long the fight lasted and whether the team defeated the boss or wiped.

Combat and chat logs contain a great deal of data irrelevant to our analysis. The sheer scale of their content makes them hard to read manually: a 4-minute boss fight might entail 45,000 lines of text. To interpret the logs, we developed a parser to extract just the data important to the above questions into a structured format. Our parser is a combination of a modified freely available parser for combat logs [26] and our own chat log parser.

In the course of this work, we separated our data into two sets: the *complete raid data* from all sessions and all fights attempted during the study period; and a second set, which

we nickname *Ulduar-25*, which includes only raids that met all the following criteria: Ulduar raid zone, 25-person parties, normal difficulty, bosses 1-12. We exclude data on bosses 13-14 because neither team ever defeated these during the study period. In the next section, we offer descriptive statistics from both variants of the data—complete and Ulduar-25—so readers understand both the overall scope of raiding in these teams and also what portion of the overall raiding was devoted to Ulduar-25. Later, where we explore managerial aspects of progression raid teams, we use only the Ulduar-25 data to facilitate comparison of the two teams, except where noted. Our analysis of the data was also informed by two of the authors, who attended every raid represented in the study. Their participation in the raids contributed to our interpretation of the system log data.

THE SCALE AND SCOPE OF RAIDING

The parsed combat logs reveal much regarding the scale and scope of these raid activities. In the following tables, we summarize several key descriptive statistics for the study, distinguishing between the complete and Ulduar-25 data sets. Table 1 summarizes several descriptive statistics regarding the time committed to raiding, by team.

	Titans	Judgment
Number of raid sessions	31	34
Avg. raid sessions/week	1.63	1.79
Avg. duration of sessions (h:mm)	3:55	3:13
Med. duration of sessions (h:mm)	4:04	3:19
Total time raiding (h:mm)	121:46	109:47

Table 1. Raid sessions and durations, by team (complete)

Both teams are quite similar on these measures of their overall level of commitment. Table 2 summarizes key descriptive statistics surrounding boss fight attempts for all raid sessions during the study period.

	Titans	Judgment
Total boss attempts (complete)	386	396
Total boss attempts (Uld-25)	302	355
Total boss wins (complete)	103	130
Total boss wins (Uld-25)	92	105
Win % (complete)	26.68%	32.83%
Win % (Uld-25)	30.46%	29.58%
% of total time in boss battles (complete)	23%	27%
Avg boss fights / hr of raiding (complete)	3.61	3.17
Avg boss fight duration (m:ss) (complete)	4:39	4:33

Table 2. Boss attempts and wins, by team

Again, the two teams are quite similar across these measures. That only roughly a quarter of the time spent raiding actually involved boss fights is a reminder that raiding is much more than boss fights: as with any collaboration, it is also a social event, with planning and preparation work, chatter, snack/bathroom breaks, etc. The data in Tables 1 and 2 offer empirical confirmation that these two guilds are “medium-core” guilds, comparable both in terms of time commitment and readiness for Ulduar content.

Table 3 summarizes raid participation in the two guilds. These numbers show who is in these raid teams (note, how-

ever, that this data tells us about characters, not human players). It is clear that the Titans have more team churn than Judgment: 50% more characters participated in at least one raid for the Titans than for Judgment. Additionally, the average number of raid sessions each character participated in for the Titans was half the average number of raid sessions for Judgment characters.

	Titans	Judgment
Num avatars raided at least once	121	80
Num raid leaders or assistants	15	14
Avg # raid sessions / avatar	5.99	12.06
Avg % of total raid sessions / avatar	24%	34%

Table 3. Avatar raid participation, by team (complete)

As these three tables show, raiding is a team activity that requires considerable commitment on the part of players. This commitment is immediately visible in the sheer number of hours they spend, but it is also visible in the patience they exhibit given the 70% failure rate.

FINDINGS: MANAGING PROGRESSION

Data from the combat logs exposes behavioral patterns that suggest player agency, i.e., player decisions that are not predetermined by the design of the game. In this section, we unpack this data by analyzing the raid data according to four temporal scales. In distinguishing among different temporal scales of progress trajectories, we are following in the footsteps of [27], whose analysis of collaboration in medical settings led them to understand the ways in which temporal scales are used to manage collaboration in hospitals. These temporal scales include individual patient illness trajectories (measured in days or weeks), nurse shift trajectories (measured in 12-hour units), and horizons (a “window of time” in which nurses optimize multitasking, measured in a few hours). As we reviewed our own data, we could see a similar set of progression trajectories taking place across four discrete temporal scales, as shown in Table 4.

Trajectory	Decisions Affect:	Time Scale
Raid zone	Team progress through a raid zone.	~4-6 mos
Raid zone lockout	The team’s progress through a raid zone lockout (defined below).	~1 wk
Raid session	The team’s progress within a single play session.	~3-4 hrs
Boss fight	The team’s strategy and tactics during a single boss fight	~5 mins

Table 4. Summary of raid progression trajectories

These temporal trajectories are not simply passive structures in which play happens; instead, raid teams actively make decisions that affect how they progress within these trajectories. For example, choosing to be among the first five guilds server-wide to complete the Ulduar raid (i.e., defeat all bosses) is a decision at the level of the raid zone trajectory: it reflects players’ intentions and commitments to one another over months. Deciding boss order within a raid zone, however, is typically done at either the raid zone lockout (explained below) or raid session trajectories. The

decision to fight a boss in the corner as opposed to fighting it in a circular motion around the room is a decision at the level of the individual boss trajectory.

The Raid Zone Trajectory

The raid zone trajectory is the scope in which a raid team tracks its progress across the months spent within a given raid zone. Raid zone trajectories correlate with game patch releases, since each patch typically includes a new raid instance. Bosses in a raid zone such as Ulduar are mostly presented in a loose sequence; in other words, players have only limited choices about how they proceed through the zone. Players are well aware of this trajectory and celebrate their progress with respect to other raid guilds via popular community websites [17, 31]. To show progress, for each guild we identified two encounters for each boss: the date of the first attempt against each boss, and the date of the first kill of each boss. In addition to showing both the attempt and kill orders, this pairing of data offers evidence about the *duration* in between first attempt and first kill for each boss. The average number of days between first attempt and first kill for the whole raid zone for Judgment was 13 (median=11), and for Titans was 7 (median=5).

More telling than the overall zone averages for days between first attempt and first kill is the *distribution* of these ranges for the two guilds. Figure 1 charts the different raid zone trajectories of the two raid teams. For each team, Figure 1 displays the number of days it took to kill a raid zone boss for the first time after it had been engaged for the first time. Regression models were applied to the trajectory data for both Judgment [solid line; $r^2 = .396$, $df = 1$, $p = .028$] and Titans [dashed line; $r^2 = .149$, $df = 1$, $p = .216$]. These models indicate that Judgment experienced a consistently increasing level of difficulty as the days required to first-kill early bosses were significantly lower than for later bosses. In contrast, the model for Titans reveals no such consistency in their progress. For Titans, the number of days required to kill a new boss was highly variable across all bosses leaving the Titans to experience inconsistent fluctuations in difficulty throughout the entire raid zone.

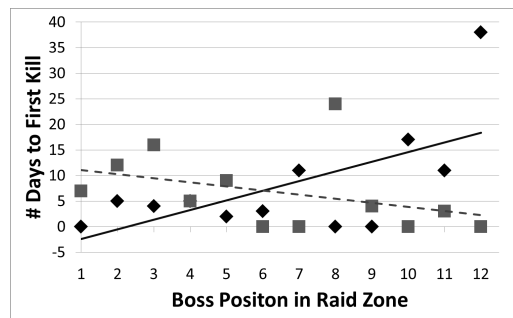


Figure 1. Days between boss first attempt and first kill, by guild [Judgment = Diamonds, Titans = Squares]

The difference in the guilds’ numbers are not due to happenstance; they reflect the decisions of the raid leaders. Not visible in Figure 1 is the fact that in spite of the apparent chaos of the Titans, they killed the 12th boss for the first

time in half the number of days Judgment needed to do the same. The reason is that whereas Judgment methodically progressed one boss at a time, the Titans tried and died and moved on to another—and the latter yielded faster results.

Other behavioral differences between the teams’ raiding styles are visible in Figures 2 and 3. For each ~3 hour session during the study period, these figures chart the overall number of attempts (including successful and failed attempts, shown with black versus gray bars, respectively). Comparing them, we see that Judgment experienced considerably more wipes in the early part of the raid zone than Titans did. This suggests that Titans may have been more prepared for the zone wide challenges than Judgment. Figures 2 and 3 also chart two trends: overall attempts over time and overall wins over time. For both of the teams, the wins trended up: as the weeks passed, they defeated more bosses, which is not surprising since raid zone lockouts require repeated killing of earlier bosses to access later ones. More interestingly, the overall attempt trends differed: whereas Judgment’s attempts trended down, Titans’ trended upwards. Thus, over the course of the study, Judgment gradually won more and lost less, while Titans won more and lost more.

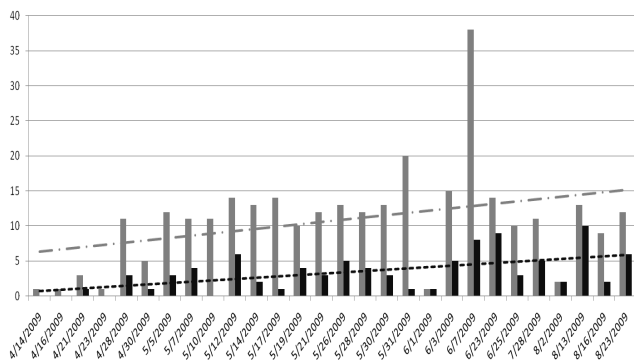


Figure 2. Attempts vs. wins over time, all bosses, by Titans

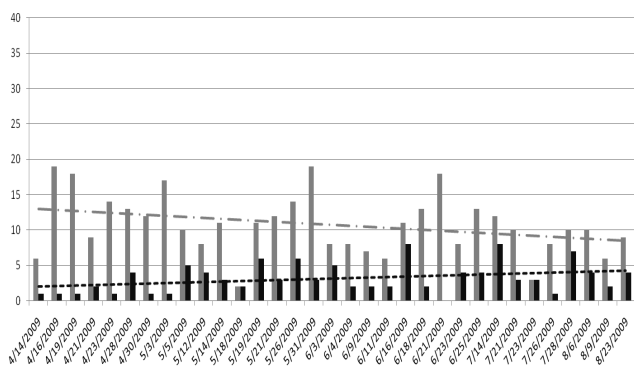


Figure 3. Attempts vs. wins over time, all bosses, by Judgment

The most common indicator of progress used by players is which boss their guild is on, e.g., 7 out of 14 bosses have been killed. On that measure, Titans is clearly superior to Judgment, since they progressed through the bosses more quickly. However, other metrics suggest that Judgment is doing better than Titans: Judgment’s attempt/win ratio improves over time, while Titans’ gets worse. Judgment’s

progress is more predictable and more closely aligns with the increasing difficulty over time curves found throughout WoW [11], while Titans’ progress is much less predictable. In these numbers we begin to see evidence of a tendency of each guild, which reveals itself in our data across temporal trajectories: though both guilds are “middle-core,” Judgment tends toward an incremental and methodical style, while Titans tend toward a tumultuous achievement style: we will return to this tendency more than once below.

The Raid Zone Lockout Trajectory

The *raid zone lockout* is a game mechanic that limits the duration of a raid instance to one week (Blizzard subsequently changed it), running Tuesday to Tuesday; this means that each Tuesday the instance “reset,” and bosses that had been killed the prior week reappeared at full health. The lockout has two key implications: no bosses could be killed more than once per week (limiting the loot that could be won from the boss), and every Tuesday the guild had to start over on the first boss. Most medium-core raid teams hold several raiding sessions during each lockout period, including both Titans and Judgment, who aspired to 2-3 sessions per week, but as Table 1 shows, actually averaged 1.63 and 1.79, respectively.

Decisions that affect this trajectory include raid zone decisions (e.g., which raid zone to tackle), 10- versus 25-person teams, which bosses are emphasized, normal vs. heroic difficulty, and in what order to attack the bosses. What is at stake with these decisions? Both the scholarly [16, 23] and popular [1, 28] literature on raiding is universal in offering a two-sided answer to this question. First, defeating raid bosses for the first time, in Nardi’s words, “were famously moments of performative ecstasy” [23]. Second, repeated failing has dangerous short- and long-term threats. Short-term threats are that raiders will quit a session or inexplicably go AFK (away from keyboard), a problem that repeatedly vexed officers throughout our game chat transcripts; one popular press book on raiding [1] devotes several pages to raid policies for managing AFK problems. Long-term problems include demoralization and social strife, both of which can lead to the dissolution of the raid guild itself; a guild officer for Judgment even offered a metric: if a raid is called off more than five times in a patch, “that’s bad news, probably guild death.”

With ecstasy as one outcome and guild dissolution the other, it’s hard to imagine how the stakes could be higher for raiders. The winning formula, according to the literature, seems to be achieving a sufficient number of first kills while managing negativity during droughts [1]. Our officer interviews and chat transcripts suggested that some of these issues were decisions handled within a given session, rather than at the weekly planning level. However, we found that the system logs nonetheless shed light on how raid teams managed these issues week to week.

One way to compare teams is to see how they treat *farm status bosses*. Over time, bosses transition from achieve-

ment status to farm status. *Achievement status* refers to bosses that are difficult to defeat, and when they are defeated, raiders feel ecstatic. *Farm status* refers to bosses that the guild is confident it can defeat, and it takes on because they must in order to fight harder bosses in the dungeon, to win more loot that will improve the raid team statistically, or to give a safe boost to morale. The transition from achievement to farm status can be seen in log data. For example, Figure 4 shows attempts (gray bars) and wins (black bars) for Judgment against Kologarn, a boss positioned early in the raid zone. The first five weeks saw 16 attempts and 3 wins, while the next 8 weeks saw 9 attempts and 8 wins.

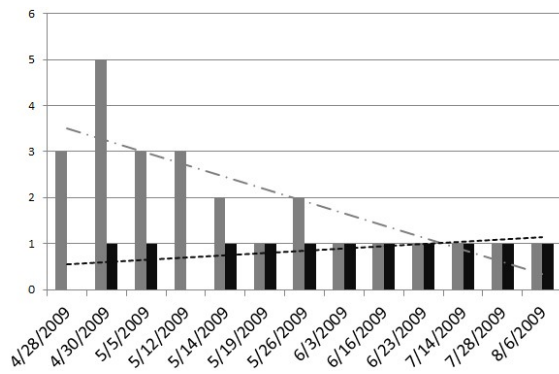


Figure 4. Attempts vs. successes against Kologarn, Judgment

In our officer interviews, we asked at what point they would consider a boss to be on farm status. Both independently replied with the same *farm status metric*: when a raid team has defeated a boss three times in a row in two or fewer tries each. Using their metric, we can say that Judgment converted Kologarn to farm status on May 26, 2009.

We then applied the officers’ farm status metric to both raid teams on all 12 bosses included in the Ulduar-25 data set, with the results shown in Table 5. Because Titans defeated the twelfth boss in half the time Judgment did, we predicted that Titans would likewise convert these bosses to farm status sooner. Yet our data shows otherwise: Judgment made slightly more use of farm bosses, converting two more bosses to that status and winning seven more fights against bosses already on this status. In spite of that, as a proportion of all fights, the two guilds were quite similar: 38.0% of all Titans wins were on farm status bosses, while 39.0% of Judgment wins were on farm status bosses. One possible explanation for the similarity is that the design of the raid zone itself—especially the weekly resets forcing teams to start over—enforces a certain percentage of farm status wins on teams.

	Titans	Judgment
Num bosses on farm status	7	9
Num wins against farm status bosses	34	41
Farm status wins as % of all wins	38.0%	39.0%

Table 5. Use of farm status bosses as a resource, by team

The opposite of farm status is achievement status, and here the two guilds’ respective approach to the 13th boss, Yogg

Saron, is illuminating. Yogg data was excluded from the Ulduar-25 data set used for most of the present analyses for two reasons: neither team defeated this boss during the study period, which complicated statistical comparison to other bosses that they had defeated; and Titans’ Yogg data was so different from the rest of the data set that it single-handedly distorted it. During the study period, Judgment took on Yogg (unsuccessfully) 2 times (0.5% of all their boss attempts during the study period). Titans took on Yogg 53 times, all wipes (14.9% of all attempts). Titans officers informed us after the study ended that they fought Yogg more than 200 times before they finally first-killed him, in what must have indeed been a “moment of performative ecstasy.” Titans’ persistence contrasted with Judgment’s disinclination even to attempt him suggests that the guilds had different raiding priorities—and that Titans have a particularly high tolerance for failure for certain achievements.

We were able to identify several other measures of guild failure tolerance. We considered the highest number of consecutive wipes on any boss as an indicator of the length of droughts experienced while learning the mechanics of the fight. A related number is the overall average number of pre-kill wipes per boss as a measure of how long it typically takes a raid team to learn the boss mechanics and defeat the boss. We also considered the average number of wipes *after* the boss had been first defeated as an indicator of how well the guild team maintained its knowledge/skills against a given boss. Finally, we considered the number of times each guild wiped more times on a subsequent kill of a given boss than it did for the first kill. The results (which, once again, are from the more limited Ulduar-25 data set and exclude Yogg) are listed in Table 6.

	Titans	Judgment
Highest num consecutive wipes on any single boss (excl. Yogg)	21	30
Ave. num pre-kill wipes per boss, all bosses	8.10	12.75
Ave. num post-kill wipes per boss, all bosses	2.55	2.23
Num of times attempts for second kill > attempts for first kill	3	1

Table 6. Tolerating failure on a given boss, by team

The first two rows clearly show that Judgment took longer to master bosses for the first time than Titans did, a finding compatible with Titans’ quicker first-kill pace through the raid zone. The second two rows tell a different story: Judgment was more likely to retain its gains than Titans, a finding consistent with the data suggesting that Judgment had more wins against farm status bosses. Another way of summarizing this is that the first kill was generally a more significant event for Judgment than for Titans, in the sense that it was harder for Judgment to achieve it and it also more strongly predicted smooth sailing afterwards.

System log data clearly does offer insight on how raid leaders managed the risk-reward ratio of progression raiding. It is compatible with existing literature in suggesting that

first-kill is a key milestone in raiding, offering empirical evidence suggesting how difficult it is to achieve first-kill, for how many subsequent attempts a boss remains difficult, and when a boss finally becomes easy for the raid team. The log data also exposes one factor that prior research appears to have overlooked, which is the specific extent to which the game design itself forces raid teams to take on farm-status bosses. We speculate that this design was intentionally implemented to help raid teams manage the risk-reward ratio without absolutely pre-determining it.

The Raid Session Trajectory

The emotional highs and lows that dominate the felt experience of raiding occur predominantly within the ~3-hour raid session. Managing the team’s success and emotions involves both practical decisions about which bosses to take on and when to give up, as well as communication skills to help the raid team emotionally cope with and also learn from failures. We saw evidence of such management in both our officer interviews and chat transcripts. The following chat transcript from Titans officers during a raid exemplifies this reasoning:

*Officer_1: [should we do heroic or] easy mode FL [Flame Leviathan, first boss in Ulduar]?
 Officer_1: get him down, go on to other bosses?
 Officer_2: yeah, we really need to move forward on normal [mode] bosses and make up for our lackluster performance last week*

In this next transcript excerpt, Judgment officers deal with negativity following a wipe:

*Officer_1: lets not give too much focus to the negative stuff
 Officer_1: it was only Player_1 [who was negative]
 Officer_2: lets go over the genral strat[egy] again quickly... just to some ppl know what’s expected
 Officer_1: and we just way over-responded to it*

One way to manage expectations is to preempt them. Titans’ officers in our interviews reveal one such strategy: “We don’t tell the guild what content we are going to raid. We try to trick them to get attendance to happen.” Titans officers also have policies for managing failure: “if we’re completely stuck for two hours, we go to a back-up boss.” Our qualitative evidence confirms prior research indicating the intensity—good and bad—of raiding and that officers are quite sensitive to it.

	Titans	Judgment
Num sessions without a win	4/28 (14%)	1/33 (3%)
Num sessions last boss a win	8/31 (25.8%)	15/38 (39.5%)
Easier raid zone boss wins/attempts	10/16	22/27
Num of sessions with visits to easier raid zones	7	12
Final boss (Yogg) wins/attempts	0/53	0/2

Table 7. Measures of session success, by team

System logs shed light on some of the raid team behaviors that come about as a result of these issues. For example, our

officers noted that ending a four-hour session on a failure hurt morale, and that two wins near the end of the session felt more successful than two wins near the beginning. A first-kill ending a session is obviously the best outcome of all. How raid teams prioritize the fights near the session end sheds light on leader predispositions to risk management. Table 7 shows how the two raid teams dealt with some of these issues.

This data tells a clear story: Judgment prioritized feel-good wins more than Titans did. More than that, the *positioning* of the wins within the session reveals each guild’s priorities. Titans overall win percent for the study duration was 30%, and yet their final boss fight win percentage was 25.8%, below their overall average: we infer that Titans more often than not decided to throw caution to the wind and go for it, hoping for that elusive last-minute first-kill. Judgment, on the other hand, had an overall 29% win average, and yet their final fight was nearly 40% successful. Likewise, Titans’ tolerance for nights without even a single win was higher by a factor of 14 than Judgment’s. This data supports the hypothesis that Titans officers had more of a high risk, high reward disposition than Judgment officers.

We wanted to develop a more nuanced understanding of failure tolerance during raid sessions. Both the research and popular raiding literature and our officers all suggest that teams get into grooves, but that these grooves have limits, a phenomenon we refer to as the *flow state duration*. To get an empirical approximation of this duration, we combined two pieces of data. First is the highest number of consecutive failed attempts that have ever led to a success. The second is the longest number of minutes wiping in a boss fight (not including downtime between fights) that has led to a subsequent win. Each of these numbers suggests the outer limit a team can remain in a flow state on a single boss. Table 8 shows these numbers for our two teams.

	Titans	Judgment
Max number of unsuccessful attempts leading to eventual win (avg, all bosses)	4.6	6.6
Max amount of unsuccessful minutes leading to eventual win (avg all bosses)	31	41

Table 8. Flow state duration, by team

This data suggests that Judgment can stay in a flow state somewhat longer than Titans. Recalling that approximately 25% of all time is spent in actual boss fights, if we multiply Titans’ 31 minutes by 4 (to estimate the total time spent in such sequences), we see that Titans almost exactly honored the two-hour policy specified by their officer in our interview, as cited earlier.

Setting outer limits on consecutive failure is not merely a matter of managing emotions, but it has productivity benefits. Once a guild falls out of a flow state, it begins to *backslide*, that is, to show a decrease in performance after many failed attempts on a single boss in one session. The Judgment officer also reported observing backslides and noted that these were often caused by a repeat problem that does

not improve over multiple attempts. Given Titans’ apparent higher tolerance for failure, Judgment’s longer flow state might appear anomalous. One possible explanation for Judgment’s lengthier flow states could be a difference in attitude towards failure during wipes. In our interviews, we asked both guilds’ officers if they could recall sessions where they took on a boss many times without defeating it, and yet still felt that they had made progress. A Judgment officer replied that if a given boss has 3 phases, and over the course of the evening they survived into successive phases, or if they got boss health progressively lower, then that still felt like progress. A Titans officer, in contrast, replied, “Doesn’t happen very often ... [I] can’t think of any raids where the boss didn’t go down and it felt like progress was made.” It is possible that Judgment had longer flow states because their progress criteria were more granular than Titans’, though this is speculative.

Data at the raid session trajectory level exposes raid team behaviors that result from the leaders’ dispositions with regard to managing risk and reward during the session. Our findings are compatible with prior literature on raiding, but they also add considerable detail in revealing how raid team dispositions translate into action, from feel-good strategies to the length in minutes of raid team flow states.

The Individual Boss Fight Trajectory

Boss fights last a matter of minutes, yet decisions made in this period obviously have an enormous impact on the guild’s progress through raid zones. Much of the relevant data here pertains to low-level individual decision-making, such as should a healer cast a quick small heal or cast a slower more potent heal? Much of our data so far has suggested that raid teams generally have some room to make their own decisions and customize their gameplay experience to manage risk and reward in their own way. We wondered whether the boss fights themselves offered similar flexibility. We collected a number of descriptive statistics concerning specific boss fights, and they are presented in Table 9.

	Titans	Judgment
Num successful boss fights	92	105
Avg duration of successful boss fights (m:ss)	6:43	6:42
Standard deviation of successful fight durations	1:06	0:58
Num wipes	210	250
Avg duration of wipes (m:ss)	4:08	3:51
Standard deviation of wipe fight durations (m:ss)	2:00	2:16

Table 9. Durations of successful vs. failed boss fights, by team

Given the hundreds of boss fights between the two teams, spread across 12 different bosses and over three months, the average durations for both successes and failures is strikingly similar. The similarity of the two guilds’ boss fight statistics suggests that these fights are deterministically designed. We infer that Blizzard has approximated the *attention span* of collaborators engaged in an intense computer-mediated

collaborative activity to be around 7 minutes and designed their boss encounters to fit within this window.

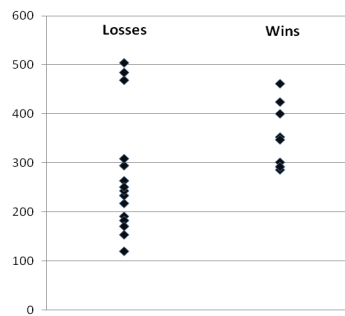


Figure 5. Scatterplot of loss vs. win durations in seconds, by Judgment on Hodir

To get a clearer picture of the timings of boss fights, we used scatterplots to show in seconds how long wins versus losses lasted per boss; Figure 5 shows a typical example. All the wins in this figure cluster around 390 seconds (6.5 minutes), while the losses cluster before or after this window. We infer that Blizzard designers use this window (400 seconds is a typical win duration for all the bosses) to help ensure that raid encounters are fun in at least three ways: 1) it introduces a sub-goal of surviving the fight long enough to get to the window, 2) it caps the fight length to prevent boredom and encourage failure, iteration, and learning, and 3) it discourages randomness in fight designs that could cause an otherwise well-executed attempt to suddenly end in failure. It should be noted that a few boss fights do not have a clear window of win. We do not believe that raid teams are explicitly aware of the window of win concept; however, our experiences suggest that some players may intuit it. It is not clear how raid teams might alter their management practices during a boss fight if they were explicitly aware of the window of win.

Data at the individual boss fight trajectory level suggests that raid teams have comparatively less ability to customize their experience or to differentiate themselves from one another during boss fights.

DISCUSSION

An ongoing question in our research was the extent to which raid teams can exercise agency, that is, make their own decisions in terms of how to raid, to manage their risks, and so forth. To the extent that this is possible, one expects to see some variation across teams. Our analyses of log data have shown many ways that raid teams, and in particular raid team leaders, can and do exercise such agency, with choices about which bosses to fight, how often and when in the sessions to insert feel-good wins, when to give up on a boss, etc. We have also found situations where raid teams have comparatively less room to differentiate themselves, primarily at the macro- and micro-levels. At the macro levels, overall failure rates, number of boss fights per hour, duration and number of raid sessions per week, and related statistics are quite similar: these seem to be defining qualities of a medium-core raiding guild. At the micro-

level, we have seen that individual boss fights appear to unfold in highly scripted ways; here the game designers have exercised their power to create very specific kinds of interactive experiences.

It is the meso-level where medium-core raid teams differentiate themselves, where Blizzard steps back and provides broader latitude (though within broad limits, as we saw with the farm boss data). At the meso level team leaders make crucial decisions about how to manage failure, when to play it safe and when to go for it. Part of the pleasure of progression raiding is that later bosses are harder than earlier ones, so the guild itself must improve its own resources in order to take on new bosses. A raid team has two fundamental types of resource that it can improve: gear and skill. The acquisition of *gear*—in-game weapons, armor, and accessories—offers characters increasing statistical advantages, such as more damaging hits or stronger armor. Teams acquire better gear primarily by repeatedly defeating already vanquished bosses to farm their loot and distribute them throughout the team. The other basic type of resource is *player skill*, which includes an in-depth understanding of the statistically optimal sequence of actions, situational awareness, appropriate customization of game client UIs, the ability to quickly form good judgments about when and how to adjust tactics, etc. For all medium-core raid teams, progression entails improving both in-game gear and player skill. However, individual teams decide how to prioritize progress in these areas. The outcome of these decisions often constitutes what we consider to be *styles* of raid progression: farm progression versus achievement styles.

A *farm progression style* is embodied in decisions that prioritize taking on relatively easy bosses to ensure that the team has widely benefitted from a given boss's loot. This strategy offers strong statistical benefits, because players become well geared. It also offers social/team benefits, because all players get more opportunities to win valuable loot, increasing the pool of qualified raiders. Because of the improved in-game character gear, human play skill also can improve at a more leisurely pace. Using this style implies that achievements have lower priority, because much of the time spent raiding is taking on previously vanquished bosses. New boss progression may entail taking on only one new boss at a time. Goal setting in this style is arguably more methodical and conservative with immediate and achievable goals set before long term and higher risk goals.

An *achievement progression style* inverts the priorities. This style seeks to take down new bosses as soon as possible, even if earlier bosses have not yet been farmed. Character gear in this approach can start to languish, placing much higher demands on player skill, both individually and as a team; alternatively, gear can become stratified, as only core raiders ever receive the comparatively infrequent loot drops, creating gear inequities within guild, a situation described in [23]. Guilds acting in this style may get further into a raid instance quite quickly, but the gains in new loot are minimal because fewer easy bosses are killed. Maintain-

ing gains may also be an issue: revisiting previously vanquished bosses is by no means an automatic win for these guilds, because they may not get practice on these bosses each week. Goal setting in this style reflects a higher tolerance for risk in favor of the higher potential returns of achieving more difficult goals.

Each of these styles represents possible ends of an emerging leadership/management spectrum, and raid teams have many options in between. And as we have already suggested, Judgment and Titans exemplify two different strategies reflecting fundamentally different predispositions about how to maximize the medium-core raiding experience for their teams: whereas Judgment leads toward a farm progression style, Titans leads toward an achievement style. These stylistic choices are evident in Judgment's prevalence for feel-good wins, higher use of farm bosses, and more systematic and predictable progress through the zone; and Titans' prevalence for achievement-before-all attitude, exemplified by their battles against Yogg, their lower reliance on farm status bosses, their comparatively chaotic kill order, and their patience for sessions without wins. From the perspective of socio-technical system design, both styles are acceptable, since both can be successful on their own terms. Indeed, the later history of the two guilds offers limited confirmation of this: more than two years and several raid zones, an expansion pack, and WoW competitor releases later, the core players of Titans and Judgment continue to raid together.

CONCLUSION

Progression raiding is a large-scale, complex social activity that places heavy demands on participants and evidently offers great rewards as well. Spanning hundreds of hours over months or years, a raid team's progression is slow and difficult. Previous research has provided a rich and compelling picture of raiding as a socio-technical experience. The present research has fleshed out this picture with new evidence of what medium-core raiding teams actually *do*, and how these behaviors relate to raid leadership decisions and constraints imposed by WoW's designers.

In particular, we identified a number of metrics, derived from system logs, that shed light on how raid team leaders are—or are not—able to customize their own play. We also saw that the behavioral decisions the metrics exposed, when viewed across different temporal trajectories, can be seen as constituting raiding styles, including farming versus achievement styles. This work sheds light on leadership aspects of virtual team collaboration in fast-paced, failure-prone socio-technical systems across different temporal trajectories.

REFERENCES

1. Andrews, S. *The Guild Leader's Handbook: Strategies and Guidance from a Battle-Scarred Veteran*. No Starch Press, San Francisco, 2010.
2. Bardzell, S. Enchanted Artifacts: Social Productivity and Identity in Virtual Material Ecologies. *Artifact*, 2 (2). 123-136.

3. Bardzell, S., Bardzell, J., Pace, T. and Reed, K. Blissfully productive: grouping and cooperation in world of warcraft instance runs *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM, San Diego, CA, USA, 2008.
4. Boellstorff, T. *Coming of Age in Second Life: An Anthropologist Explores the Virtually Human*. Princeton University Press, Princeton, 2008.
5. Brown, B. and Barkhuus, L. Leisure and CSCW: Introduction to Special Edition. *Computer Supported Cooperative Work*, 16 (1-2). 1-10.
6. Brown, B. and Bell, M. CSCW at play: 'There' as a collaborative virtual environment *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, ACM, Chicago, Illinois, USA, 2004.
7. Cefkin, M., Stucky, S. and Ark, W. The grand diversion: Play, work, and virtual worlds. *Artifact*, 2 (2). 108-115.
8. Chen, M. Communication, coordination, and camaraderie in World of Warcraft. *Games and Culture*, 4. 47-73.
9. Crabtree, A., Benford, S., Capra, M., Flintham, M., Drozd, A., Tandavanitj, N., Adams, M. and Farr, J.R. The Cooperative Work of Gaming: Orchestrating a Mobile SMS Game. *Computer Supported Cooperative Work*, 16 (1-2). 167-198.
10. Dibbell, J. "A Rape in Cyberspace." *The Village Voice*. 21 Dec 1993
11. Ducheneaut, N., Yee, N., Nickell, E., & Moore, R.J. The life and death of online gaming communities: a look at guilds in world of warcraft. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07)*. ACM, New York, 2007, 839-848.
12. Ducheneaut, N. and Moore, R.J. The social side of gaming: a study of interaction patterns in a massively multiplayer online game. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, ACM, Chicago, Illinois, USA, 2004.
13. Ducheneaut, N., Moore, R.J. and Nickell, E. Virtual "Third Places": A Case Study of Sociability in Massively Multiplayer Games. *Computer Supported Cooperative Work*, 16 (1-2). 129-166.
14. Ducheneaut, N., Yee, N., Nickell, E. and Moore, R.J. "Alone together?": exploring the social dynamics of massively multiplayer online games *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM, Montreal, Quebec, Canada, 2006.
15. Ellis, J.B., Luther, K., Bessiere, K. and Kellogg, W.A. Games for virtual team building *Proceedings of the 7th ACM conference on Designing interactive systems*, ACM, Cape Town, South Africa, 2008.
16. Golub, A. Being in the World (of Warcraft): Raiding, Realism, and Knowledge Production in a Massively Multiplayer Online Game. *Anthropological Quarterly*, 83 (1). 17-45.
17. GuildOx: WoW Guild Progress. <http://www.guildox.com/go/g.asp>.
18. Hansen, S., Berente, N., Pike, J. and Bateman, P.J. Productivity and Play in Organizations: Executive perspectives on the real-world organizational value of immersive virtual environments. *Artifact*, 2 (2). 69-81.
19. Kahai, S., Carroll, E., and Jestice, R. Team Collaboration in Virtual Worlds. *DATABASE* 38, 4 (2007), 61-68.
20. Malaby, T.M. These great urbanist games: New Babylon and Second Life. *Artifact*, 2 (2). 116-122.
21. Manninen, T., Virtual team interactions in networked multimedia games. in *In Proceedings of the 4th Annual International Workshop on Presence (PRESENCE 2001)*, (Philadelphia, PA, 2001).
22. Moore, R.J., Gathman, E.C.H., Ducheneaut, N. and Nickell, E. Coordinating joint activity in avatar-mediated interaction *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, San Jose, California, USA, 2007.
23. Nardi, B. *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft*. University of Michigan Press, Ann Arbor, 2010.
24. Nardi, B. and Harris, J. Strangers and friends: collaborative play in world of warcraft *Proc. of CSCW'10*, ACM, Banff, Alberta, Canada, 2006.
25. Pearce, C. *Communities of Play: Emergent Cultures in Multiplayer Games and Virtual Worlds*. MIT Press, Cambridge, MA, 2009.
26. Racy. WowLogParser, 2010.
27. Reddy, M.C., Dourish, P. and Pratt, W. Temporality in Medical Work: Time also Matters. *Computer Supported Cooperative Work*, 15 (1). 29-53.
28. Rezvani, J. *Guild Leadership: Lessons from the Virtual World*. Lulu, np, 2008.
29. Sweetser, P. and Wyeth, P. GameFlow: a model for evaluating player enjoyment in games. *Computer in Entertainment*, 3 (3). 3-3.
30. Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N. and Nickell, E. From Tree House to Barracks: The Social Life of Guilds in World of Warcraft *Games and Culture*, 1 (4). 338-361.
31. WoWProgress: WoW Rankings and History. <http://www.wowprogress.com/>.
32. Yee, N. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence*, 15 (3). 309-329.
33. Yee, N. The labor of fun: How video games blur the boundaries of work and play. *Games and Culture*, 1. 68-71.
34. Yee, N., Ducheneaut, N., Nelson, L., and Likarish, P. Introverted elves & conscientious gnomes: The expression of personality in World of Warcraft. *Proc. of CHI2011*, 753-762.